Dener Eduardo Bortolini

Detecção de Incêndios Ambientais Utilizando Persistência Espaço-temporal, Segmentação por Cor e Sub-amostragem de Vídeo

Belo Horizonte

2013

Dener Eduardo Bortolini

Detecção de Incêndios Ambientais Utilizando Persistência Espaço-temporal, Segmentação por Cor e Sub-amostragem de Vídeo

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Inteligência Computacional

Orientador: Hermes Aguiar Magalhães
Co-orientador: Hani Camille Yehia

Belo Horizonte
2013

Bortolini, Dener Eduardo.
B739d Detecção de Incêndios Ambientais Utilizando Persistência Espaço-temporal, Segmentação por Cor e Sub-amostragem de Vídeo
Dener Eduardo Bortolini. - 2013
xiv, 111 f., enc. : il.

Orientador: Hermes Aguiar Magalhães.
Co-orientador: Hani Camille Yehia.

Dissertação (mestrado) - Universidade Federal de Minas Gerais, Escola de Engenharia.

Bibliografia: f.107-111.

1. Engenharia elétrica Teses. 2. Incêndios Teses. 3. Monitoramento ambiental Teses. I. Magalhães, Hermes Aguiar. II. Yehia, Hani Camille. III. Universidade Federal de Minas Gerais, Escola de Engenharia. IV. Título.

CDU: 621.3(043)

Comissão Julgadora:

Dr. Carlos A. M. do Nascimento

Dr. Douglas A. G. Vieira

Prof. Dr. Orlando A. Gomes

Prof. Dr. Hermes Aguiar Magalhães
Orientador

Prof. Dr. Hani Camille Yehia
Co-orientador

Dedicatória

"Dedico este trabalho à minha família e amigos que me ajudaram nessa árdua jornada."
Dener Eduardo Bortolini

Agradecimentos

Agradeço ao meu orientador Hermes Aguiar Magalhães, ao meu co-orientador Hani Camille Yehia, aos meus amigos e tutores Orlando A. Gomes, Edel F. B. Stanciolli e Cioli C. Stanciolli. Bem como todos os outros amigos que fizeram parte da jornada de criação desse trabalho.

Especialmente agradeço à minha mãe Gessi B. Bortolini, meu pai Antônio B. Sguizzato e meu irmão Charles R. Bortolini por me apoiarem em tudo que faço, não sendo esse trabalho uma exceção.

Agradeço também a ANEEL e CEMIG pelo apoio financeiro no desenvolvimento e produção deste trabalho.

Resumo

Este trabalho propõe um método/algoritmo de monitoração ambiental no intuito de detectar focos de incêndios em áreas ambientais por meio de imagens de vídeo, utilizando segmentação dos pixels por cor no espaço de cores YCbCr. Características obtidas por outro sistema de cores (YUV) foram usadas aqui para gerar um novo classificador e melhorar o desempenho da classificação por cor por meio de um comitê de classificadores. Para tratar melhor os problemas encontrados em ambientes externos e extrair características relevantes ao problema, foi incorporado o método de subtração de fundo por médias adaptativas, utilizando cores ao invés de escala de cinza. Além disso, um novo modelo baseado em colônia de formigas foi proposto de forma a avaliar as características espaciais e temporais do fogo com o intuito de filtrar objetos com cor de fogo e reduzir a taxa de falsos positivos causados por estes objetos. Foi incorporado um mecanismo de sub-amostragem espacial em diversas resoluções, sendo realizada uma análise do tempo empregado no processamento dos vídeos em cada resolução para avaliar a possível utilização destas resoluções em tempo real. Experimentos foram realizados para comparação com métodos relevantes na literatura, onde foi demonstrado que o método proposto conseguiu reduzir o número de falsos positivos mantendo uma classificação superior ou bem próxima dos resultados obtidos por outros métodos, superando estes no quesito de tempo de processamento em resoluções superiores as usadas na literatura. Foi mostrado também que o método pode ser usado fora do contexto proposto com relativa facilidade, podendo ser utilizado em um ambiente industrial sem muitas restrições.

Palavras-chave: monitoração ambiental, detecção de incêndios, imagens de vídeo

Abstract

This work proposes a method/algorithm of environmental monitoring in order to detect outbreaks of fires in environmental areas through video images, using pixel segmentation by color in the YCbCr color system. Characteristics obtained by another color system (YUV) were used here to generate a new classifier and improve the color classification performance through a committe of classifiers. To better address the problems found in outdoor environments and to extract features relevant to the problem, a background subtraction method called running average was incorporated, using colors instead of gray scale. In addition, a new model based on ant colonies was proposed to assess the spatial and temporal features of fire in order to filter objects with the color of fire, reducing the rate of false positives caused by these objects. A subsampling mechanism for different resolutions was implemented and an analysis of the time spent in processing the videos on each resolution was done to evaluate the possible use of these resolutions in real time. Experiments were carried out and results compared to relevants methods in the literature, where it is pointed out that the proposed method was able to reduce the number of false positives while maintaining a top ranking or being close to the results obtained by other methods, as well as surpassing those in the category of processing time for higher resolutions. Furthermore, it was shown that the method can be used outside of the proposed context with relative ease, like an industrial environment, for example.

Keywords: environmental monitoring, fire detection, video images

Sumário

Lista de Abreviaturas ix
Lista de Figuras xi
Lista de Tabelas xiv

1 Introdução **1**
1.1 Contextualização 1
1.2 Definição do problema 5
1.3 Objetivos e contribuições deste trabalho 6
1.4 Organização do trabalho 7

2 Revisão Bibliográfica **8**
2.1 Características do fogo 8
2.2 Fundamentos da Imagem Digital 12
2.2.1 Percepção das cores e espaços de cores 13
2.3 Avaliação de características: Humanos x Máquinas 23
2.4 Detectores de focos de incêndio 26

3 Desenvolvimento **35**
3.1 Base de dados e medidas de desempenho 36
3.1.1 Base de dados 36
3.1.2 Medidas de desempenho 41
3.2 Segmentação dos objetos em movimento utilizando subtração de fundo 44
3.3 Classificação dos pixels pela cor 52
3.4 Eliminação de falsos positivos através da persistência temporal e espacial dos objetos 65
3.4.1 Agrupamento dos pixels com cor de fogo 66
3.4.2 Características espaciais e temporais 70
3.5 Técnicas para redução de custo computacional 76
3.5.1 Subamostragem de pixels 76
3.5.2 Ajuste automático do limiar de subtração de fundo 83

4 Resultados **87**
4.1 Resultados do experimento 1 88
4.2 Resultados do experimento 2 96
4.3 Resultados do experimento 3 100

5 Conclusão **104**

Lista de siglas e abreviaturas

4CIF 4 X CIF

ANEEL Agência Nacional de Energia Elétrica

ARMA Modelo de filtro autoregressivo e de médias adpatativas (*Autoregression and Moving Average model*)

AR Parte do ARMA ou modelo de filtro autoregressivo (*Autoregression model*)

AT Variável de atualização referente a etapa de persistência espaço-temporal

BBD *Bounding Box Disorder*

CAVIAR *Context Aware Vision using Image-based Active Recognition*

CCD *Charge-coupled device*

CEMIG Companhia Energética de Minas Gerais

CG Variável referente a etapa de persistência espaço-temporal que possui o valor do centro de gravidade de um objeto capturado no vídeo

CIE *International Comission on Illumination* ou *Comission Internationale de L'Eclairage*

CIF *Common Intermediate Format*

CLKP Çelik Puro

CLKS Çelik com subtração de fundo

CMYK *Cyan, Magenta, Yellow and Key (Black)*

Curva ROC *ROC Curve - Receiver Operating Characteristic Curve*

CVPR Lab Laboratório de Visão Computacional e Reconhecimento de Padrões da Universidade de Keimyung da Coréia

DFT Transformada discreta de Fourier*Discrete Fourier Transform*

EEENNS *Equal-average Equal-variance Equal-norm Nearest Neighbor Search*

EX Variável de existência referente a etapa de persistência espaço-temporal

FCR *Fire Color Ratio*

$\mathbf{F}_{D383}$ Regra geral de classificação dos pixels no método D383

FFT Transformada rápida de Fourier*Fast Fourier Transform*

FGALSSVM *Fuzzy Genectic Algorithm Least Square Support Vector Machine*

FNN Redes neurais Fuzzy (*Fuzzy neural network*)

Fnyq Abreviatura de frequência de Nyquist

Full HD *Full High Definition*

ImaSta Conjunto de imagens estáticas com imagens que possuem fogo e que não possuem fogo

IP *internet protocol*

HD *High Definition*

HMM *Hidden Markov Models*

Lab Espaço de cores onde L é a luminância, ”a” representa a crominância azul e ”b” a crominância vermelha

Luv Espaço de cores onde L é a luminância, ”u” a crominância vermelha e ”v” representa a crominância azul

MA Parte do ARMA ou modelo de filtro de médias adaptativas (*Moving Average model*)

NTSC *National Television System Comittee*

NWCG *National Wildfire Coordinating Group*

OPENCV biblioteca de visão computacional de livre acesso

PAL *Phase Alternating Line*

POD *Principal Orientation Disorder*

QQVGA *Quarter QVGA*

QVGA *Quarter Video Graphics Array*

RGB espaço de cores representado vermelho, verde e azul (*Red, Green, Blue*)

ROI *Region of interest*

$\mathbf{S}_T$ Limiar de sensibilidade de detecção de fogo do método Toreyin et al. (2006)

SVGA *Super Video Graphics Array*

TOF *Time-of-flight*

TOR Método de Toreyin et al. (2006)

T_e Limiar de tempo mínimo de existência para que grupo seja classificado como fogo

T_G Limiar de tamanho mínimo de um grupo

T_{dG} Limiar de distância mínima entre grupos

T_{mG} Limiar de movimento máximo de um grupo

T_s Limiar da subtração de fundo

T_{SV} Limiar da segmentação usando o canal V

VGA *Video Graphics Array*

VD Vídeo do conjunto VídeoDiv

VDC Vídeo do conjunto VídeoDivCtrl

VDT Vídeo do conjunto VídeoDis

VideoDis Conjunto de vídeos de experimentos de distância

VideoDiv Conjunto de vídeos de experimentos em diversos tipos de contexto

VideoDivCtrl Conjunto de vídeos de experimentos em diversos tipos de contexto sem fogo

VideoRes Conjunto de vídeos com experimentos em diversas resoluções

VR Vídeo do conjunto VídeoRes

YCbCr Espaço de cores onde Y é a luminância, Cb a crominância azul e Cr a crominância vermelha

YUV Espaço de cores onde Y é a luminância, U a crominância azul e V a crominância vermelha

WWF *World Wide Fund of Nature*

Lista de Figuras

1.1 Distribuição das linhas de transmissão de energia elétrica no território nacional e em Minas Gerais . . . 3

2.1 Triângulo de fogo . . . 8
2.2 Imagens de incêndios de copa e superfície . . . 12
2.3 Matriz de sensores CCD de uma câmera fotográfica e captura da imagem 12
2.4 Comprimento de onda referente à faixa visível do espectro eletromagnético 14
2.5 Cubo do espaço de cores RGB . . . 15
2.6 Imagem no espaço de cores RGB . . . 16
2.7 Imagem no espaço de cores YUV . . . 17
2.8 Imagem no espaço de cores YCbCr . . . 18
2.9 Funções de equiparação das cores CIE RGB e CIE XYZ . . . 19
2.10 Transformação do espaço de cores RGB para o espaço de cor CIE XYZ . 20
2.11 Transformação do espaço de cores RGB para o espaço de cor CIE Lab . . 22
2.12 Transformação do espaço de cores RGB para o espaço de cor CIE Luv . . 22
2.13 Objetos de ilusão de ótica: Cubo de Necker, Vaso de Rubin e Triângulo de Kanizsa . . . 23
2.14 Contraste simultâneo e Bandas de Mach . . . 25

3.1 Base de dados de imagens estáticas . . . 37
3.2 Conjunto de vídeos VídeoRes . . . 38
3.3 Conjunto de vídeos VídeoDiv . . . 39
3.4 Conjunto de vídeos VídeoDivCtrl . . . 41
3.5 Conjunto de vídeos VídeoDis . . . 41
3.6 Esquema do filtro $ARMA_{(1,1)}$ para a adaptação da imagem de fundo . . 46
3.7 Resultados de segmentação obtidos para diferentes valores de α . . . 49
3.8 Segmentação dos objetos em movimento . . . 50
3.9 Subtração de fundo usando escala de cinza e o esquema cores RGB . . . 51
3.10 Diferença entre as figuras 3.9E e 3.9F . . . 52
3.11 Curva ROC para a base de imagens estáticas . . . 55
3.12 Delimitação do espaço de cor para classificação como fogo usando k-médias . . . 57
3.13 Segmentação com k-médias . . . 58
3.14 Segmentação usando Çelik e Demirel (2009) . . . 59
3.15 Diferença de contraste entre os canais de diferentes espaços de cores . . 62
3.16 Funcionamento do classificador/segmentador F_{D383} . . . 64
3.17 Geração de um conjunto pelo algoritmo de floresta de conjuntos disjuntos 69

3.18 Agrupamento de diferentes focos de incêndio 69
3.19 Experimento de Goss et al. (1989) para colônias de formigas 72
3.20 Resultado final obtido usando a persistência espacial e temporal 75
3.21 Efeito *moiré* e *jaggies* causados por subamostragem 78
3.22 Quadros dos vídeos VR 1 e VR 8 79
3.23 Imagens do espectro de Fourier e os gráficos de atenuação do sinal dos quadros dos vídeos VR 1 e VR 8 na resolução de 1920 x 1080 pixels . . . 81
3.24 Imagens do espectro de Fourier e os gráficos de atenuação do sinal dos quadros dos vídeos VR 1 e VR 8 na resolução de 640 x 480 pixels 82

4.1 Razão de verdadeiros positivos no experimento 1 90
4.2 Razão de falsos positivos no experimento 1 90
4.3 Acurácia dos métodos no experimento 1 91
4.4 Medida F dos métodos no experimento 1 91
4.5 Tempo médio de processamento de vídeo dos métodos no experimento 1 92
4.6 Pixels utilizados nas técnicas de sub-amostragem 95
4.7 Número de verdadeiros positivos no experimento 2 97
4.8 Número de falsos positivos no experimento 2 98
4.9 Acurácia dos métodos no experimento 2 98
4.10 Medida F dos métodos no experimento 2 99
4.11 Bordas capturadas de uma chama saturada em diferentes distâncias . . 101
4.12 Métodos capturando a reflexão da chama no concreto 102

Lista de Tabelas

2.1 Tabela de relação entre cor da chama e temperatura. 9

3.1 Tabela resumo das bases de dados. 36
3.2 Descrição dos vídeos que compõem o conjunto VídeoRes. 38
3.3 Descrição dos vídeos que compõem o conjunto VídeoDiv. 39
3.4 Descrição dos vídeos que compõem o conjunto VídeoDivCtrl. 40
3.5 Descrição dos vídeos que compõem o conjunto VídeoDis. 41
3.6 Matriz de confusão para uma tarefa de classificação com duas classes. . 42
3.7 Resultados obtidos para os canais de crominância da cor vermelha dos diferentes espaços de cores 61
3.8 Tabela de subamostragem de pixels. 77
3.9 Tabela resumo das regras utilizadas em cada etapa 85

4.1 Tabela de valores para os parâmetros referentes a cada método. 89
4.2 Tabela de resultados do teste piloto para a técnica de sub amostragem. . 95
4.3 Tabela de resultados com a razão de verdadeiros positivos obtida para cada vídeo do experimento de distância. 101

Capítulo 1

Introdução

1.1 Contextualização

Nos últimos anos, vemos que a consciência ambiental tem se tornado cada vez mais importante na sociedade contemporânea. Estudos realizados sobre os impactos causados pelo modo de vida moderno indicam que providências devem ser tomadas ou a própria sociedade irá destruir o planeta que chama de lar. Lavorato (2002) fala da escassez de recursos naturais e o crescimento desordenado da população mundial que gera grandes impactos ambientais, necessitando o uso de um planejamento ambiental tão estratégico quanto os planejamentos urbanos.

Em 2002, a *World Wide Fund of Nature* (WWF), uma organização de conservação ambiental independente e uma das maiores no mundo, publicou em WWF (2002) que a humanidade usava 20% a mais de recursos naturais do que a Terra conseguia repor, avançando sobre os estoques naturais do planeta e comprometendo não só a sobrevivência da geração atual como também das gerações futuras. Mesmo com medidas sendo tomadas, em um novo relatório em WWF (2012), esses números aumentaram para 50% mais recursos que a Terra pode prover, e a menos que medidas drásticas sejam tomadas, no ano de 2030 dois planetas não serão suficientes para suprir a demanda.

Este problema é ainda mais grave no Brasil, onde grandes áreas florestais são substituídas por pasto ou campos de plantio, sendo geralmente utilizadas as queimadas para preparar a área de forma eficiente e barata. Motta (2008) fala em seu trabalho sobre a preocupação mundial com os incêndios ambientais, que não só trazem prejuízos financeiros, mas também prejuízos imensuráveis à biodiversidade e aos ecossistemas. O autor também menciona o significativo aumento do uso do fogo como uma ferramenta de preparação do solo para fins rurais. Este tipo de ação pode levar a um incêndio de grandes proporções, afetando áreas não só visadas para o plantio, mas também áreas importantes de fauna e flora, como florestas tropicais e equatoriais que possuem menor probabilidade de ocorrência de incêndios devido à densidade florestal e umidade presente nos materiais combustíveis, fato este mencionado por Crutzen e Goldammer (1993).

Um método de prevenção deste tipo de problema ambiental é a monitoração des-

tas áreas onde a fauna e a flora ainda não foram substituídas pelas áreas urbanas ou rurais em constante crescimento, prevenindo danos primários, neste caso incêndios e desmatamentos desordenados causados pelo aumento exponencial da demanda por recursos da sociedade e consequentemente, prevenindo danos secundários que incluem alterações ambientais, como por exemplo, o efeito estufa que causa o derretimento de calotas polares e alterações climáticas que levam a secas prolongadas e a um risco maior de incêndios.

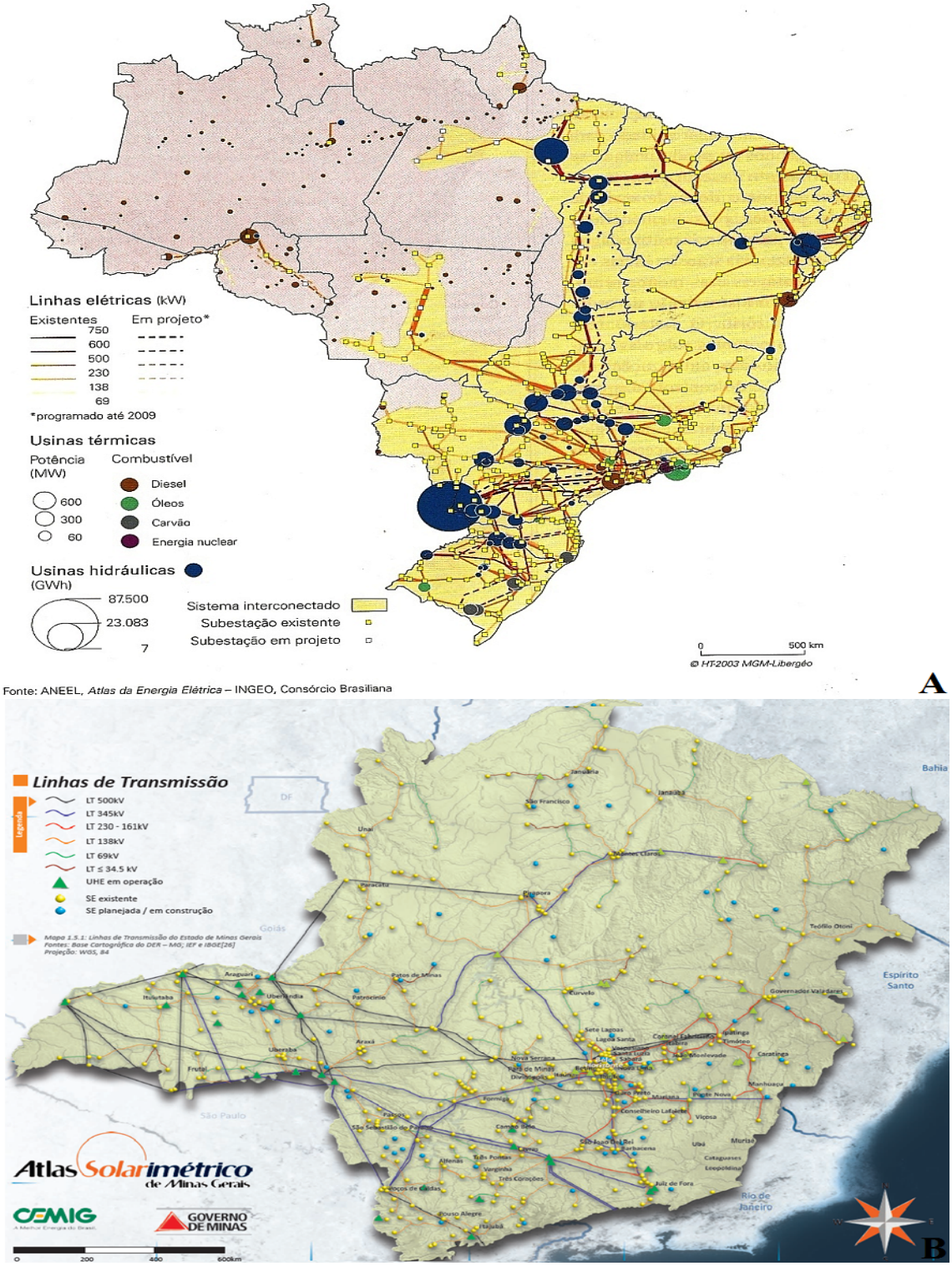
Uma solução de monitoração de incêndios é o uso de detectores térmicos e/ou fumaça e partículas, amplamente utilizado em complexos industriais e urbanos (Joglar et al. (2005); Yuen e Chow (2005)). Entretanto este tipo de detector tem uma área de alcance limitado e pode facilmente falhar em ambientes externos onde as condições ambientais mudam constantemente e as áreas a serem monitoradas possuem grande extensão.

Outra tecnologia muito usada atualmente na sociedade contemporânea é a captação de imagens para monitoração de lugares públicos ou privados, visando garantir a segurança seja das pessoas, bens valiosos ou informação, conforme Boult et al. (2001) explicam.

Esta ferramenta versátil pode ser facilmente usada para a monitoração de vastas extensões de áreas florestais, detectando incêndios e outros eventos depredativos como desmatamentos sem a necessidade de uma pessoa no local fazendo a vigilância ou a utilização de vários equipamentos e sensores, economizando recursos e sendo menos suscetível a falhas, seja por má fé de algumas pessoas ou mau funcionamento dos dispositivos.

Um problema que surge com esta abordagem é: onde colocar as câmeras? Uma resposta seria o uso da grade de distribuição de energia elétrica, composta de inúmeras torres que compõem a linha de transmissão de alta tensão. Estas linhas de transmissão transpõem grandes extensões do país, através de reservas florestais e locais que geralmente são de difícil acesso para a monitoração ambiental. Este é um recurso muito importante, já que parte dos custos de instalação de uma câmera está ligado a sua estrutura de sustentação, no caso devendo estar em um lugar alto para conseguir uma boa cobertura da área. Portanto, a utilização destas torres permite reduzir significativamente o trabalho e recursos utilizados para um projeto de grande escala, uma vez que é possível com uma adaptação relativamente fácil instalar as câmeras nessas torres. Outra vantagem indireta, é que estas câmeras ainda vão permitir monitorar as linhas de transmissão, obtendo dados importantes destas linhas sob diversas situações. A figura 1.1A mostra uma imagem do território brasileiro e a abrangência do sistema de distribuição de energia elétrica em sua extensão e a figura 1.1B mostra as linhas de transmissão no estado de Minas Gerais. Podemos ver em ambas as figuras o potencial do uso das torres de transmissão como recurso para a monitoração ambiental.

É importante salientar que a utilização deste sistema de monitoração por imagens de vídeo só é possível graças à evolução em diversas ciências e tecnologias nos últimos anos, já que há 10 anos seria praticamente impossível conceber tal idéia devido às



Fonte: ANEEL, *Atlas da Energia Elétrica* – INGEO, Consórcio Brasiliana

Figura 1.1: Distribuição das linhas de transmissão de energia elétrica no território nacional e Minas Gerais. (A) Área de abrangência do sistema interligado de transmissão de energia elétrica no território nacional FONTE: ANEEL (2003); (B) Linhas de transmissão no estado de Minas Gerais. FONTE: CEMIG (2012)

limitações de conexões de rede e equipamentos. Entre os equipamentos que merecem destaque estão câmeras que possuem uma qualidade de imagem elevada atualmente e que também funcionam por *internet protocol* (IP) permitindo uma alta conectividade, bem como a velocidade de processamento dos computadores atuais, permitindo o uso de algoritmos mais robustos e complexos para realizar a tarefa de monitoração ambiental.

Com esta evolução conjunta dos diversos dispositivos e tecnologias, houve um aumento significativo de dados obtidos através destes sistemas de monitoração e vigilância baseado em imagens, gerando a necessidade de usar a sistemas computacionais (Ko et al. (2010); Marbach et al. (2006); Schultze et al. (2006)), fazendo com que estes sistemas de monitoração sejam utilizados de forma mais eficiente, permitindo que processos de tomada de decisão sejam inseridos no processo e ajudando na captação de eventos relevantes ao contexto de monitoração, que no caso deste trabalho é a detecção de focos de incêndios em reservas florestais. No entanto, esta tarefa de decisão requer diversas etapas que podem ser sintetizadas em: envio e recepção das imagens captadas, segmentação, extração de características relevantes ao contexto ambiental monitorado, classificação ou tomada de decisão a partir dos dados, bem como o armazenamento dos eventos ocorridos para estudos posteriores.

Este tipo abordagem acima em que uma tomada de decisão está embutida no processo geralmente é definido como parte da área de visão computacional, conforme explicado por Jahne e Haussecker (2000) e Gonzalez e Woods (2010), no qual esta área representa um ramo da área de conhecimento de inteligência artificial. Ainda segundo Gonzalez e Woods (2010), a visão computacional e as áreas do conhecimento de processamento de imagens e análise de imagens compartilham limites de difícil separação, em que uma área pode utilizar métodos e técnicas que sejam comuns a outra área. Para resolver esse impasse, os autores separam os processos computacionais aplicados em uma imagem em 3 níveis distintos:

- Nível baixo - Inclui operações primitivas como ajuste de contraste, realce da imagem, redução de ruídos, etc.;
- Nível médio - Envolve tarefas de segmentação de objetos, descrição dos objetos e classificação dos mesmos. Geralmente este processo é caracterizado pela sua entrada ser uma imagem e as saídas serem atributos extraídos destas imagens, como por exemplo, bordas, contornos e identidade dos objetos;
- Nível alto - Este processo envolve dar sentido a um conjunto de objetos reconhecidos, realizando funções cognitivas associadas à visão humana.

Não é difícil associar o problema proposto de detecção de incêndios por vídeo ao nível médio e alto de processamento de imagens citados acima, nos quais segundo os autores, representam respectivamente as áreas de análise de imagens e visão computacional, indicando o relacionamento intrínseco entre as áreas do conhecimento.

Entretanto, utilizar processos de visão computacional em ambientes externos não controlados possui diversos desafios, entre eles a constante variação das condições ambientais ao longo do dia, problema este citado por Ruas (2009), assim como objetos com

cor de fogo que dificultam o processo de detecção, neste caso de contexto florestal, galhos, flores ou folhas sofrendo processo de abscisão foliar devido a condições adversas do meio ambiente (BrasilEscola (2013); Sóbiologia (2013)), fazendo com que estas assumam uma coloração amarelo alaranjado.

Este trabalho é fruto do projeto D383 realizado em parceria com a Companhia Energética de Minas Gerais (CEMIG), que propõe a utilização da redes de transmissão de alta tensão para a monitoração ambiental com enfoque nos focos de incêndios, que causam grandes danos ambientais e geralmente afetam também as linhas de transmissão. O projeto D383 tem como objetivo criar uma solução completa de monitoração por vídeo automatizada pelo uso de modelos computacionais. O projeto pretende contar também com o apoio da comunidade via *web* no site da própria CEMIG, onde os resultados do processamento e os próprios vídeos serão disponibilizados. Nosso trabalho aqui apresentado, portanto, contou com o apoio financeiro da ANEEL e CEMIG, e está restrito apenas à criação da solução computacional para a detecção de focos de incêndios automática, no qual o algoritmo proposto herda como nome as siglas do projeto com o sufixo de sua funcionalidade que é detecção de fogo "F", sendo referido durante todo o trabalho como método D383F.

1.2 Definição do problema

A monitoração em ambientes externos está sujeita às condições ambientais e variações ao longo do tempo, gerando vários problemas para o processamento e análise das imagens, já que estas condições não podem ser controladas conforme Ruas (2009) explica.

Um dos maiores problemas a ser levado em conta é causado pela iluminação natural, que pode alterar significativamente a cena, como o caso da passagem de uma nuvem gerando sombras na área sendo monitorada, ou ainda o anoitecer ou amanhecer do dia.

Outro problema com características semelhantes aos mostrados acima é causado pelo dispositivo de captura de vídeo, isto é a câmera, como Ruas (2009) explica. Os sensores das câmeras digitais mapeiam uma grandeza contínua em valores discretos, no caso a quantidade de luz refletida ou emitida pelos objetos. Este mapeamento acarreta em uma perda de precisão durante a quantização para o armazenamento digital, limitando a faixa de valores possíveis para representar as imagens. Um exemplo deste problema acontece quando uma cena apresenta brilhos maiores ou menores que os limites do sensor (faixa dinâmica), nos quais estes valores são mapeados para os limites da faixa dinâmica do sensor. A maior parte das câmeras atualmente possui mecanismos de ajuste automático de ganho, exposição e outros parâmetros para tratar estas situações, simplificando o uso do equipamento e diminuindo a intervenção humana para o funcionamento correto do mesmo. Entretanto, esses ajustes automáticos geram problemas similares aos causados pela alteração de iluminação no ambiente.

Estes tipos de situações podem fazer com que o algoritmo trate cada pequena variação na cena como uma nova cena, levando a erros sistemáticos, sendo necessário o uso de

técnicas que possam se adaptar a estas mudanças rapidamente e diminuir a taxa de erros do algoritmo. Dentro do contexto florestal, existe ainda o problema causado pela queda natural de folhas pelo processo de abscisão foliar, nas quais folhas adquirem uma cor amarelo alaranjado e podem ser facilmente confundidas pela cor com fogo, bem como suas flores e galhos que podem possuir uma coloração de fogo. Além disso, quando esses objetos são sujeitos à força dos ventos, adquirem movimento e podem facilmente assumir padrões semelhantes ao do fogo, representando outro desafio a ser resolvido.

Outro fator importante a ser levado em conta é o tempo gasto para o processamento das imagens obtidas, já que o algoritmo irá realizar uma tarefa crítica a partir de imagens de vídeo adquiridas continuamente, exigindo que o processamento seja rápido o suficiente para a monitoração em tempo real e diminuindo os prejuízos causados pelo fogo da melhor forma possível. Com a evolução tecnológica das câmeras, estes dispositivos são capazes de enviar imagens em resoluções cada vez mais altas, possuindo maior qualidade e quantidade de informações para serem processadas. Esse é outro desafio a ser superado, já que se espera que o algoritmo seja capaz de processar imagens capturadas por qualquer câmera, independente de sua época, tecnologia ou resolução, armazenando estas imagens e gerando uma base de dados que possa ser utilizada posteriormente em outros estudos ambientais.

1.3 Objetivos e contribuições deste trabalho

O objetivo geral dessa dissertação de mestrado é realizar a detecção de focos de incêndios em ambientes florestais por meio de imagens de vídeo, utilizando técnicas de visão computacional citados pela comunidade científica e especificamente trabalhando com esquemas de cores conhecidos de forma a extrair características que permitam a detecção do objeto de interesse em tempo real.

Como objetivos específicos deste trabalho são propostos:

- Reduzir o máximo possível a taxa de falsos positivos sem comprometer o desempenho na detecção dos focos de incêndios através de técnicas derivadas de outras áreas do conhecimento como computação natural e classificação de padrões;
- Utilizar imagens em resoluções mais altas que as utilizadas na literatura (320 x 240 pixels) mantendo a detecção em tempo real, de tal forma que essas imagens possam ser utilizadas posteriormente para outras pesquisas ambientais.

Ao perseguir esses objetivos, as seguintes contribuições foram alcançadas e serão mostradas nesse trabalho:

- A segmentação dos pixels em movimento utilizando cores ao invés da escala de cinza amplamente utilizada na literatura;
- Utilização de um comitê de classificadores, em que o método proposto por Çelik e Demirel (2009) foi alterado para a inclusão de outro classificador e ambos trabalham juntos na classificação dos pixels segundo sua cor;

- Uma técnica de persistência espaço-temporal utilizando uma idéia parcialmente baseada em colônias de formigas em que são avaliadas as características de movimento aleatório das chamas em torno do material combustível e o tempo de duração do evento de combustão;
- Sub-amostragem de imagens de vídeo em resoluções superiores a 352 x 288 pixels para que a detecção em tempo real seja possível.

Além disso, testes foram realizados fora do escopo florestal para mostrar a eficiência do método D383F na detecção de chamas em geral incluindo ambientes urbanos, que geralmente possuem chamas saturadas geradas a partir de materiais combustíveis derivados de petróleo, que liberam mais energia durante a combustão.

1.4 Organização do trabalho

No capítulo 2 é feita uma revisão bibliográfica sobre tema de estudo, em que algumas técnicas utilizadas para resolver o problema são detalhadas. Em seguida, o capítulo 3 descreve a metodologia proposta para o desenvolvimento do algoritmo. O capítulo 4 mostra os resultados obtidos para diferentes testes em que algumas funcionalidades do algoritmo são demonstradas. Finalmente, o capítulo 5 relata as conclusões obtidas no trabalho.

Capítulo 2

Revisão Bibliográfica

2.1 Características do fogo

Para gerar o algoritmo de detecção de focos de incêndio, primeiramente um estudo sobre o comportamento do fogo deve ser realizado, identificando e enunciando características que sejam viáveis para extração e processamento.

Segundo National Wildfire Coordinating Group (NWCG) (2009), o fogo é definido como uma rápida oxidação de um material em um processo químico exotérmico liberando calor, luz e vários produtos resultantes da reação química, ou seja, uma reação de combustão que produz luz conforme descrito em Beutling (2009). O fogo se inicia quando um material combustível combinado a uma quantidade suficiente de comburente é exposto a uma fonte calor. Estes três elementos formam o triângulo de fogo, ilustrado na figura 2.1.



Figura 2.1: Triângulo de fogo, onde os três fatores: calor, combustível e oxigênio se unem para gerar o fogo. FONTE: Beutling (2009)

Após iniciado, o fogo produz calor suficiente para sustentar a reação de combustão, que continuará até que um dos elementos do triângulo seja eliminado (Batista (1995)).

A chama é a parte visível do fogo, que nada mais é que uma mistura de sólidos e gases incandescentes emitindo energia luminosa. Segundo Ito et al. (1992), a cor da

chama é determinada pela radiação da chama na região de comprimento de ondas visíveis do espectro eletromagnético. Existem dois tipos de eventos que dão cor à chama: a emissão espectral causada pela excitação de átomos, em que elétrons saltam de um estado de mais alta energia para outro de menor energia, liberando um fóton no processo e a emissão causada pela radiação termal nas partículas sólidas ou fuligem geradas na combustão.

Ainda segundo Ito et al. (1992), essas emissões variam dependendo de vários fatores, em que o material combustível e sua mistura com o oxidante determinam a qualidade da combustão e conseqüentemente a quantidade de fuligem e temperatura da chama. Isto leva à separação da combustão em dois tipos distintos: completa e incompleta. A combustão completa acontece quando todo o combustível é queimado, este tipo de combustão só ocorre em situações controladas como, por exemplo, em maçaricos e fogões domésticos, gerando uma chama de cor azul referente à emissão espectral pela excitação dos átomos de água e gás carbônico.

Em incêndios florestais e combustíveis queimando ao ar livre geralmente irá ocorrer a combustão incompleta do material combustível, produzindo uma grande quantidade de partículas ou fuligem. Segundo Eisberg e Resnick (1999), estas partículas tem o comportamento de um corpo negro, que é definido como todo corpo que emite um espectro de radiação universal que depende apenas de sua temperatura, não de sua composição. Este tipo de corpo absorve toda a radiação que incide sobre eles.

Portanto, no nosso caso, a cor da chama é determinada principalmente pela radiação termal emitida por essas partículas ou fuligem, em que a cor, variando do amarelo claro ou branco até o vermelho, indicará a temperatura da chama (Stirling Consolidated Boiler Company (1905)), temperatura que será praticamente a mesma para qualquer tipo de combustível sob condições similares. A tabela 2.1 mostra a relação entre a cor da chama e a temperatura da mesma.

Cor da chama		**Temperatura (∘ C)**
Vermelho	Visível	525
	Escuro	700
	Cereja escuro	800
	Cereja	900
	Cereja Puro	1000
Laranja	Escuro	1100
	puro	1200
Branco	Esbranquiçado (amarelo claro)	1300
	Brilhante	1400
	Puro	1500

Tabela 2.1: Tabela de relação entre cor da chama e temperatura.

De acordo com a *Natural Resources Canada*, um incêndio florestal médio pode possuir chamas com temperaturas de 800 ºC, e em condições extremas pode chegar a 1200

ºC, possuindo valores que corroboram com a tabela acima em que incêndios florestais geralmente possuem chamas amarelo-alaranjadas. Esta é a característica principal das chamas em incêndios florestais a ser levada em conta pelos classificadores de chamas pela cor.

Convém mencionar que esses valores variam substancialmente em função de diversos fatores, sendo os principais: **material combustível, clima e topografia**.

O material combustível florestal, segundo Batista (1995) é qualquer material orgânico, vivo ou morto, no solo ou acima deste, suscetível de participação no processo de combustão.

No trabalho de Beutling (2009), ele resume várias características importantes de incêndios florestais levantadas por diversos autores, entre elas está o fato de os combustíveis florestais serem o fator principal para estudo do início e propagação do fogo, em que as particularidades deste fator permitem estimar o comportamento de um incêndio. Estes combustíveis são divididos em diversas classes sob diferentes contextos:

- Aéreo e superficial;
- Quantidade total e disponível;
- Classificação segundo sua periculosidade.

Os combustíveis aéreos se encontram acima de 1,80 metros de altura, se referindo a galhos e copas das árvores, e os superficiais abaixo desta altura, contendo todo o resto como folhas secas, gramíneas, troncos, etc.

A divisão por quantidade total e disponível está ligada ao fato que nem todo o combustível pode ser consumido pelo fogo devido ao alto teor de umidade. Estes valores de combustível total e disponível variam principalmente em função de fatores climáticos como época do ano, tempo atmosférico, etc.

E por último, os combustíveis florestais possuem as seguintes divisões segundo sua periculosidade:

- perigosos: Todos os materiais secos com diâmetro $\leq$ 1 cm, constituindo de pequenos galhos, folhas e gramíneas. Estes materiais perdem umidade rapidamente e possuem uma temperatura de ignição menor, propiciando o início do fogo e propagação. Constitui a matéria principal de consumo pelo fogo;
- semi-perigosos: materiais secos com diâmetro >1 cm. Constituindo-se de galhos, troncos, húmus e turfa. São materiais lenhosos ou em decomposição que queimam lentamente, mas desenvolvem um calor intenso, mantendo-se em combustão latente e gerando riscos de reiniciar incêndios;
- verde: se refere à vegetação viva, podem ser considerados como não inflamáveis devido ao alto teor de umidade, no entanto, o calor liberado pela combustão de outros materiais pode secar estes materiais e torná-los igualmente perigosos.

O clima pode interferir na temperatura e propagação das chamas com maior ou menor relevância em um incêndio florestal, dependendo de como a insolação, umidade relativa do ar, temperatura do ar, vento e precipitação se relacionam com o material combustível, principalmente em épocas de seca, em que a falta de água faz com que o processo de abscisão foliar se torne mais intenso e uma maior quantidade de material combustível perigoso se acumule sobre a superfície do solo.

Ainda, segundo Beutling (2009), a topografia também é um fator importante já que a velocidade de propagação de um incêndio é diretamente proporcional à inclinação topográfica, podendo esta influir sobre o comportamento do fogo de maneira semelhante ao vento.

Estes três fatores citados acima (material combustível, clima e topografia) levam a três diferentes tipos de incêndios florestais:

- incêndios subterrâneos: este tipo de incêndio se propaga através das camadas de húmus e turfa sobre o solo mineral e abaixo do piso da floresta. Estes incêndios se propagam lentamente sem a presença de chamas e com pouca fumaça, sendo de difícil detecção. Este incêndio tem efeitos bastante negativos na vegetação, uma vez que este causa a morte das raízes e compromete a fertilidade no solo. Embora menos comum, este tipo de incêndio no Brasil ocorre em regiões de cerrado;

- incêndios superficiais: este incêndio se propaga sobre o piso da floresta, queimando o material combustível superficial. Em períodos de seca, se propaga muito rapidamente, gerando uma grande quantidade de chamas e calor. Este é o tipo de incêndio mais comum e é a forma pela qual começam todos os incêndios florestais.

- incêndios de copa: este tipo se caracteriza pela propagação do fogo através das copas das árvores de maneira independente dos incêndios superficiais. Este tipo de incêndio é comum em florestas de coníferas e árvores que apresentam folhagem inflamável. Incêndios deste tipo são de difícil combate, liberam uma grande quantidade de calor, ocasionando conseqüentemente uma grande mortandade na população vegetal.

Dentro do que foi relatado, em que a cor da chama será a característica mais importante a ser avaliada, é de se esperar chamas caracterizadas por cores variando do amarelo-esbranquiçado ao laranja para os incêndios de copas mais extremos, e chamas amarelas alaranjadas para incêndios superficiais e de copas comuns, conforme ilustra a figura 2.2.

Após o breve detalhamento das características e comportamento do fogo florestal, podemos passar a próxima etapa do estudo que é o entendimento básico da ferramenta utilizada para capturar e extrair as informações relevantes para o processo de detecção: a imagem digital.



Figura 2.2: Imagens de incêndios de copa e superfície. (A) Incêndio de copa com chamas variando da cor branca ao laranja; (B) Incêndio de superfície com chamas amarelas alaranjadas. FONTE: Summit Counts (2013)

2.2 Fundamentos da Imagem Digital

Uma imagem digital pode ser definida como uma matriz bidimensional retangular em que cada elemento desta matriz, que são chamados de pixels, representa uma intensidade e cor. As câmeras digitais geralmente utilizam uma matriz de sensores *Charge-coupled device* (CCD), nos quais a resposta de cada sensor é proporcional à integral de energia luminosa projetado sobre o sensor (Gonzalez e Woods (2010)). A imagem é definida nesta matriz de maneira similar a um mosaico e este é o tipo de arranjo predominante nas câmeras digitais. A figura 2.3 ilustra como uma imagem é representada em um CCD de uma câmera.

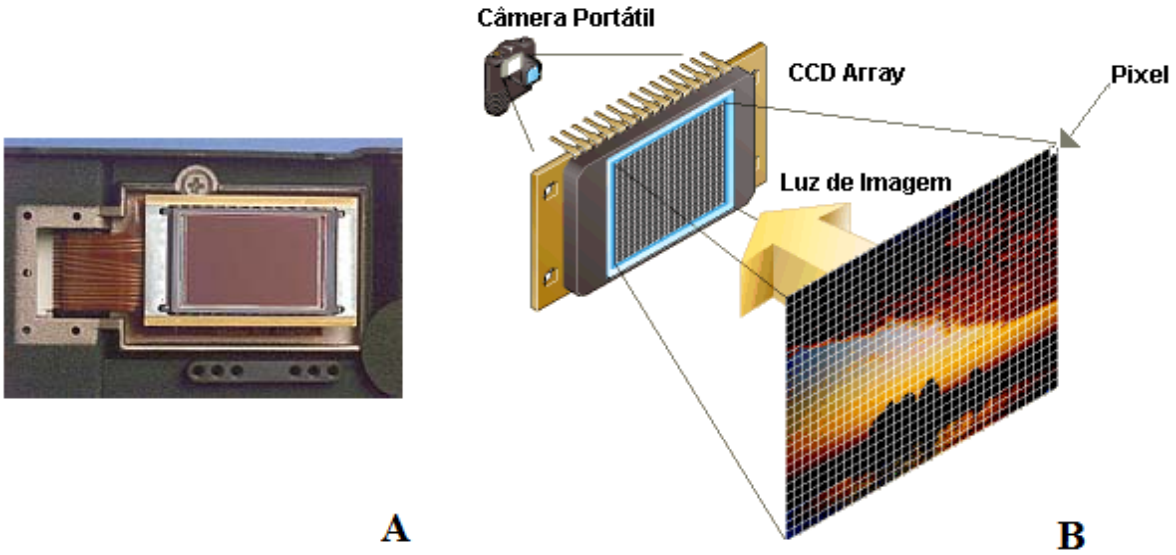


Figura 2.3: Matriz de sensores CCD de uma câmera fotográfica e captura da imagem. (A) Sensores CCD de uma câmera fotográfica; (B) Matriz de captura da imagem representada no CCD. FONTE: Universidade de São Paulo - USP (2013)

A resolução de pixels máxima de uma imagem capturada por uma câmera, observada certas dependências de compatibilidade com a lente, corresponde ao número de pixels que formam a matriz CCD, portanto, quanto mais pixels ou sensores o CCD possui, maior a resolução da imagem digital obtida, conseqüentemente, maior será o consumo de memória para armazenar e tempo de processamento para gerar e trabalhar com esta imagem.

Para a criação da imagem digital é necessário definir que tipo de representação ela irá usar, isto é, escala de cinza ou cores, nas quais diversas codificações podem ser utilizadas, como por exemplo, número de tonalidades de cinza, ou sistemas de cores diferentes, cada um com sua codificação específica.

2.2.1 Percepção das cores e espaços de cores

A cor pode ser definida como uma sensação subjetiva vivenciada por um observador através dos mecanismos de recepção sensíveis à luz presentes nos olhos (Chamberlin e Chamberlin (1980)).

Os seres humanos possuem dois tipos de receptores em seus olhos (especificamente na retina) que são responsáveis pela visão das cores e luminância: os cones que respondem pela percepção das cores e os bastonetes respondem pela percepção de iluminação. Os cones detectam três cores principais, também chamadas de cores primárias: vermelho, verde e azul. Os bastonetes, mais sensíveis, determinam os níveis de luminosidade ou tons de cinza.

Segundo Gonzalez e Woods (2010), a percepção e interpretação das cores no cérebro humano é tido como um fenômeno fisiopsicológico que não é totalmente compreendido, no qual a natureza das cores é expressa através de resultados experimentais e teóricos. Como a sensação de cor depende do observador, diferentes pessoas tem diferentes julgamentos sobre uma cor, sendo normal haver um certo desentendimento em cores intermediárias, como o laranja e amarelo.

Ainda em Gonzalez e Woods (2010), os autores ainda explicam como as cores percebidas por seres humanos e animais são determinadas pela natureza da luz refletida pelo objeto. A luz visível é composta por uma banda de frequências relativamente estreita no espectro de energia eletromagnética. A figura 2.4 ilustra a faixa de luz visível no espectro eletromagnético.

A cor de um objeto, assumindo que este não seja emissivo como chamas ou lâmpadas, portanto, nada mais é que a refletância de um objeto em função do comprimento de onda do espectro eletromagnético, assim um objeto será verde se este refletir a luz com comprimento de onda no intervalo de 500 a 570 nm, absorvendo o restante da energia de outros comprimentos de onda.

Nas câmeras, o processo de captura e representação de uma imagem procura fidelizar ao máximo a cena observada, no entanto, devido ao processo de quantificação de

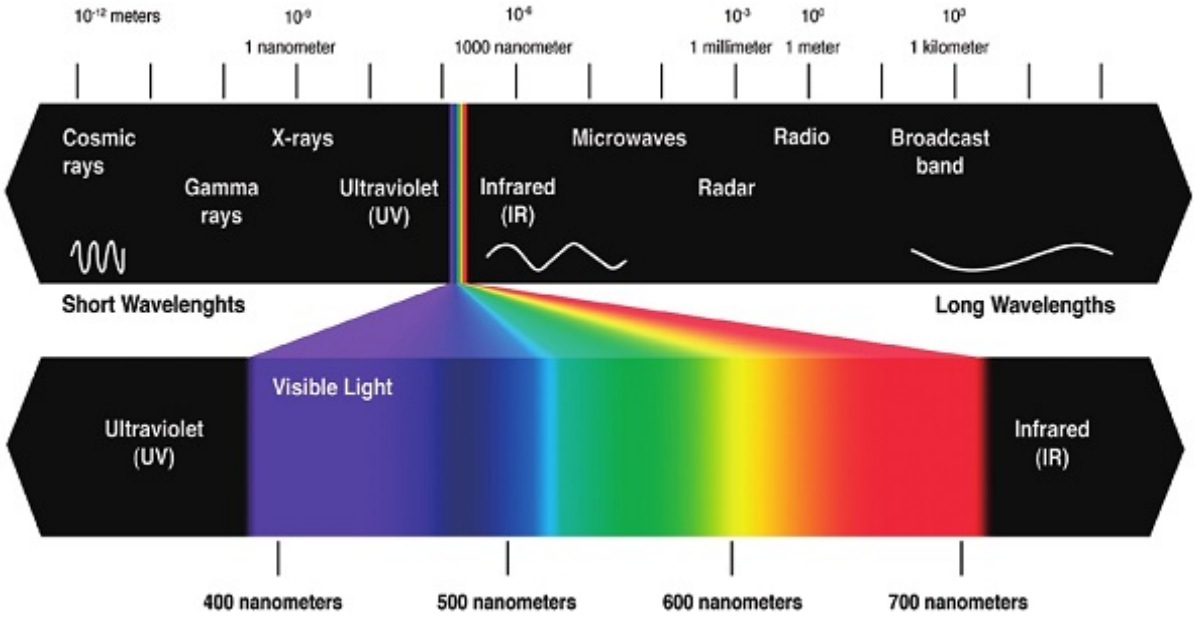


Figura 2.4: Comprimento de onda referente à faixa visível do espectro eletromagnético.

grandezas contínuas para valores discretos, pode ocorrer uma perda significativa de atributos na imagem.

Para tentar remediar alguns problemas, diversos algoritmos são usados para fazer ajustes automáticos dependendo de algumas características presentes na cena, levando ao problema de alteração constante nos vídeos capturados citado por Ruas (2009), bem como a uma captura das chamas com um nível menor de detalhamento como, por exemplo, uma chama sob determinadas condições pode ser capturada por uma câmera próximo do valor de saturação, isto é, quase branco, quando a cor das chamas deveria estar variando de amarelo claro a amarelo (limitação da faixa dinâmica da câmera). É evidente que as câmeras são bem mais limitadas na captura de uma cena do que o olho humano e este é um problema que os algoritmos devem na medida do possível tratar utilizando técnicas robustas ou espaços de cores que propiciem a obtenção das características relevantes ao processo de classificação.

De forma a tratar alguns destes problemas e facilitar a especificação das cores em alguma forma padronizada, que seja amplamente aceita, existem diversos espaços de cores. É graças a essa característica que existe relativa diversidade de métodos propostos na literatura com o intuito de detectar focos de incêndio, onde cada autor explora diferentes características de um determinado espaço de cores para executar a classificação dos pixels segundo sua cor.

A maior parte dos espaços de cores utilizada atualmente é orientada a *hardware* (como monitores ou televisores e impressoras) ou ainda aplicações que envolvem manipulação de cores (como a criação de imagens coloridas para uma animação). As aplicações que envolvem processamento de imagens, no qual será exibido em algum dispositivo de

exibição de imagens emissivo (eletrônicos), utilizam o modelo de composição de cores por adição, em que a soma de cores primárias (vermelho, verde e azul) produz cores intermediárias (ciano, amarelo e magenta). Aplicações que envolvem materiais refletivos, como papel, banners e outros, utilizam o modelo subtrativo em que a subtração de cores intermediárias produz as cores primárias, não sendo este modelo de interesse para o presente trabalho.

Dentro deste aspecto, alguns destes espaços de cores que farão parte do escopo do trabalho serão explicados a seguir.

Espaço de cores RGB

A primeira representação básica de uma imagem digital em cores é o espaço de cores representado vermelho, verde e azul (*Red*, *Green*, *Blue*) (RGB). Uma imagem nesse espaço de cores, bem como em qualquer outro espaço de cor tem no mínimo 3 matrizes para sua representação (com exceção do espaço de cores CMYK, que é usado em impressoras). Cada matriz ou canal de cor possui em cada pixel um valor em intensidade da cor primária que ela representa e a composição dos diferentes canais que formam o espaço de cores indicará qual cor será visualizada no pixel. Assim, este modelo se baseia em um sistema de coordenadas cartesianas, formando um cubo, conforme mostrado na figura 2.5.

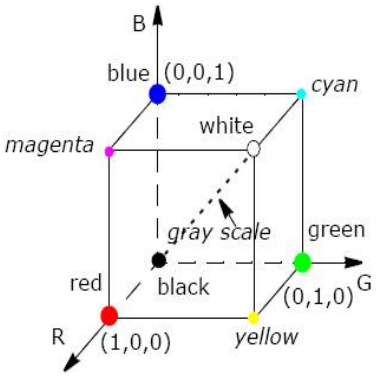


Figura 2.5: Cubo do espaço de cores RGB. O valor 1 indica saturação no canal, onde o valor correspondente em 8 bits é 255. FONTE: Intel (2013)

A representação deste sistema geralmente usa 8 bits por pixel para cada canal de cor, ou seja, cada pixel é representado por 24 bits, podendo este sistema gerar $(2^8)^3$ ou 16.777.216 cores diferentes. Este sistema é o mais utilizado pelas câmeras em geral, em que estas entregam as imagens neste esquema de cores, sendo o preferido por muitos autores (Chen et al. (2004); Çelik et al. (2007); Ko et al. (2010); Toreyin et al. (2006); Wang et al. (2010); Wen et al. (2010); Zhou et al. (2010)) por evitar um custo computacional extra ao realizar uma transformação do espaço de cores. A figura 2.6 mostra em escala de cinza a relevância de cada componente do espaço de cores RGB na composição de uma imagem exemplo com fogo.

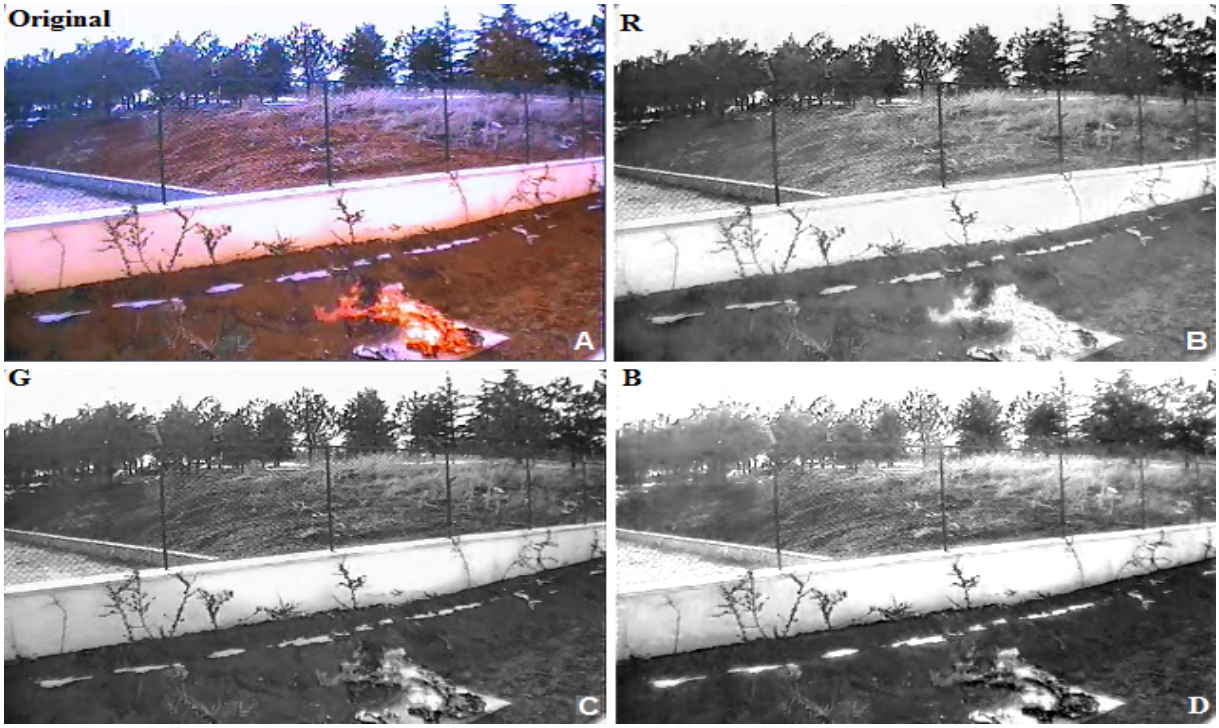


Figura 2.6: Imagem no espaço de cores RGB. (A) Imagem original; (B) Imagem de intensidade referente ao componente "R"; (C) Imagem de intensidade referente ao componente "G"; (D) Imagem de intensidade referente ao componente "B". A chama apresenta diferentes valores de intensidade em cada componente. Imagem original retirada de Toreyin et al. (2006)

Espaço de cores YUV

O espaço de cores YUV é usado pelos sistemas de codificação colorida *Phase Alternating Line* (PAL) e *National Television System Comittee* (NTSC). Este espaço de cores foi desenvolvido para ser utilizado tanto em aparelhos de televisão preto e brancos quanto coloridos. A informação da escala de cinza ou brilho está presente no componente "Y", ou luminância, e as informações de cromaticidade ou cor nos componentes "U" e "V".

Segundo Equasys (2013), este espaço de cores tira vantagem de características baseadas na percepção visual humana, uma vez que os olhos são muito mais sensíveis a informação de brilho do que a crominância, permitindo uma redução das informações de cromaticidade sem perda significativa da qualidade. Este também é um espaço de cores que alinha seus eixos com relação ao brilho, conforme Bradski (2008) relata em seu livro.

A transformação do espaço de cores RGB para o YUV pode ser feita segundo (Intel (2013)):

$$\begin{aligned} Y &= 0.299 * R + 0.587 * G + 0.114 * B; \\ U &= -0.147R - 0.289G + 0.436B = (B - Y) * 0.492; \\ V &= 0.615R - 0.515G - 0.100B = (R - Y) * 0.877; \end{aligned} \tag{2.1}$$

A figura 2.7 mostra a representação de cada componente na mesma imagem com fogo utilizada anteriormente. Após a transformação da imagem do RGB para o YUV, percebe-se que a região da chama apresenta valores de intensidade baixos no componente "U" e valores altos no componente "V".

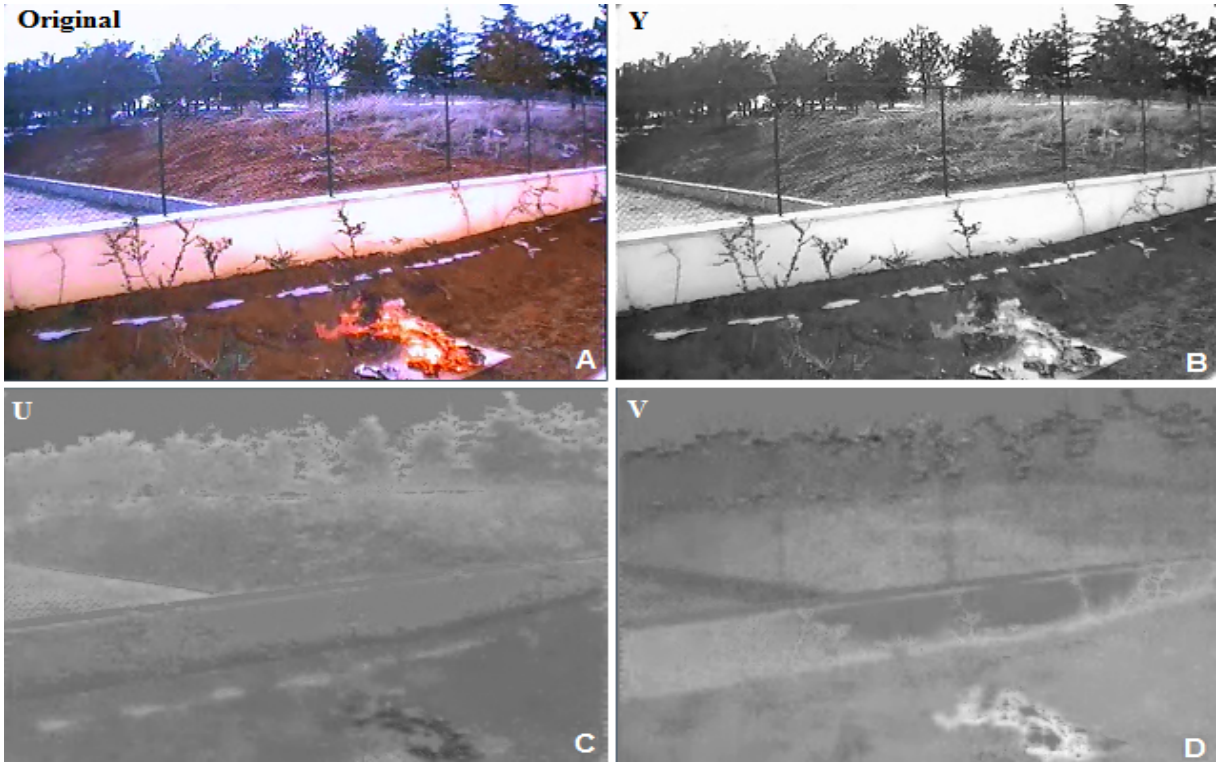


Figura 2.7: Imagem no espaço de cores YUV. (A) Imagem original em RGB; (B) Imagem de intensidade referente ao componente "Y"; (C) Imagem de intensidade referente ao componente "U"; (D) Imagem de intensidade referente ao componente "V". A chama apresenta valores de intensidade baixos no componente "U" e valores altos no componente "V". Imagem original retirada de Toreyin et al. (2006)

YCbCr

Segundo Intel (2013), o espaço de cores YCbCr é uma versão dimensionada e deslocada do espaço YUV, possuindo bastante similaridade com o YUV com um componente de luminância ("Y") e dois de cromaticidade ("Cb" e "Cr", crominância azul e vermelha respectivamente). Em contraste ao YUV, este espaço de cores é usado no contexto de imagens digitais e processamento de vídeos, especialmente para imagens JPG e codificação de vídeos MPG (Equasys (2013)). A transformação do espaço RGB para o espaço YCbCr é feita segundo a equação (OpenCV (2013)):

$$Y = 0.299 * R + 0.587 * G + 0.114 * B;$$

$$Cr = (R - Y) * 0.712 + delta;$$

$$Cb = (B - Y) * 0.564 + delta;$$

onde

$$delta = \begin{cases} 128 & \text{para imagens de 8 bits} \\ 32768 & \text{para imagens de 16 bits} \\ 0.5 & \text{para imagens de ponto flutuante} \end{cases} \quad (2.2)$$

A figura 2.8 mostra a transformação de uma imagem do espaço de cor RGB para os diferentes componentes do espaço de cor YCbCr, podendo ser observado que a chama apresenta características de intensidade bem semelhantes aos relatados no espaço YUV, com valores baixos no componente "Cb" (cromaticidade azul) e valores altos no componente "Cr" (cromaticidade vermelha).

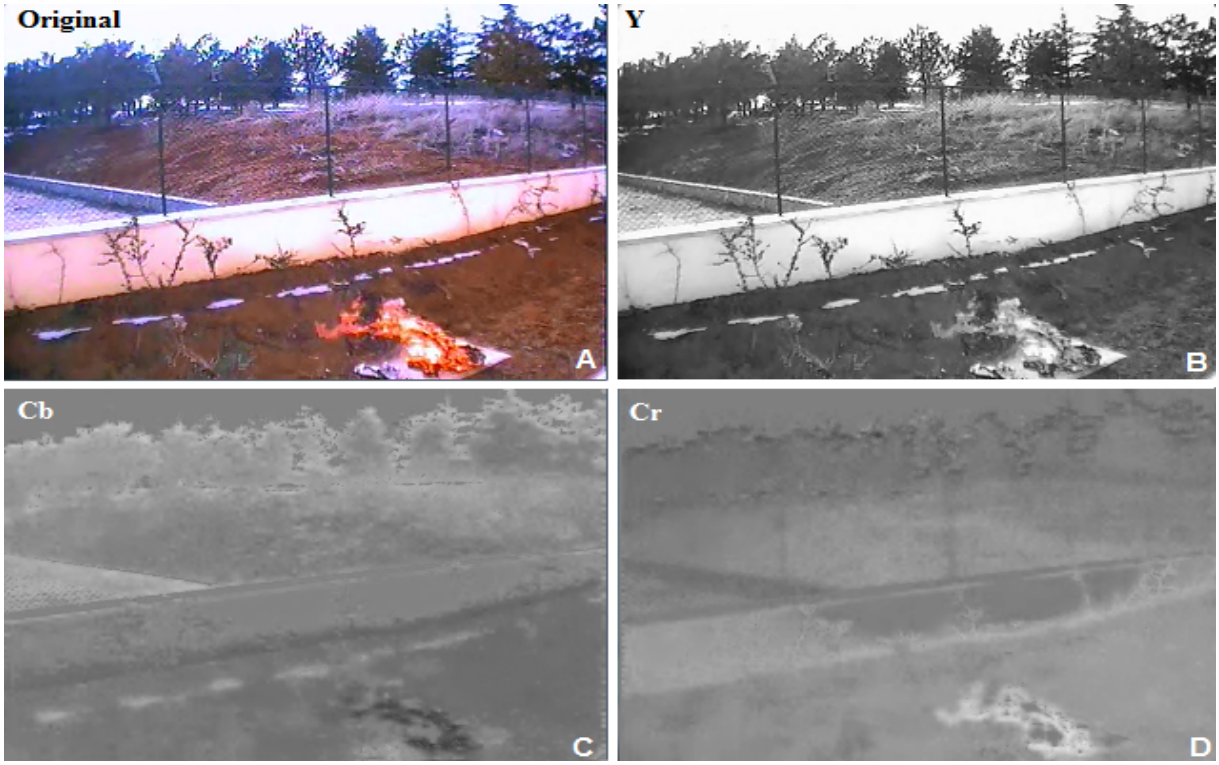


Figura 2.8: Imagem no espaço de cores YCbCr. (A) Imagem original em RGB; (B) Imagem de intensidade referente ao componente "Y"; (C) Imagem de intensidade referente ao componente "Cb"; (D) Imagem de intensidade referente ao componente "Cr". Semelhante ao espaço YUV, o fogo apresenta valores de intensidade baixos no componente "Cb" e valores altos no "Cr". Imagem original retirada de Toreyin et al. (2006)

Espaço de cores CIE XYZ

A *International Comission on Illumination* ou *Comission Internationale de L'Eclairage* (CIE) em francês, é referência mundial em luz, iluminação e espaços de cores. Graças à suas pesquisas, diversos espaços de cores foram criados para resolver vários problemas.

Um desses problemas é relatado em Lima (2010), no qual remete para a limitação da representação das cores pelo espaço RGB. Para que todas as cores perceptíveis pelo ser humano possam ser codificadas em RGB, pelo menos um componente do espaço

de cor deve assumir valores negativos, o que não é possível de ser reproduzido em dispositivos emissores, impedindo que estes sejam capazes de representar toda a gama de cores do espectro eletromagnético visível.

O espaço de cores CIE XYZ procura resolver esse problema representando todas as cores visíveis pelo ser humano apenas com valores positivos. Nesse espaço, o componente "Y" representa a informação de luminância, "Z" representa o estímulo da luz azul nos cones responsáveis pela percepção do azul e "X" uma mistura (combinação linear) das curvas de respostas dos cones escolhidas para serem não negativas. Dessa forma, dado um valor "Y" de luminância, o plano XZ contém todas as cromaticidades possíveis com esse valor.
A figura 2.9 exemplifica o que foi dito acima, em que o gráfico CIE RGB mostra que determinadas cores do espectro demandam valores negativos de cores primárias (*tristimulus*), no caso a cor vermelha, para serem representadas. No mesmo gráfico ainda é possível ver a quantidade necessária de vermelho e verde para que o amarelo seja representado e o comprimento de onda de cada cor primária pura neste espaço. Já o gráfico CIE XYZ demonstra como as componentes XYZ representam as cores no espectro eletromagnético apenas com valores positivos.

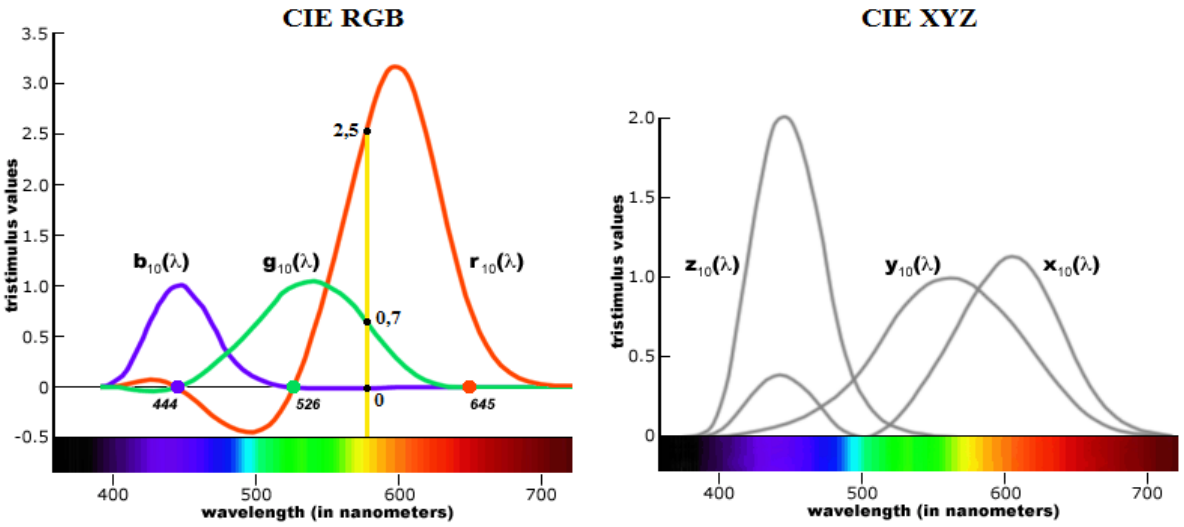


Figura 2.9: Funções de equiparação das cores CIE RGB e CIE XYZ. De forma a não usar valores negativos presentes no CIE RGB, em 1931 a CIE definiu o novo padrão XYZ derivado das cores RGB para representar o espectro da cor. FONTE: Lima (2010)

A transformação do espaço RGB para XYZ pode ser feita segundo:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0.4124 & 0.3576 & 0.1805 \\ 0.2126 & 0.7152 & 0.0722 \\ 0.0193 & 0.1192 & 0.9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{2.3}$$

A figura 2.10 ilustra a representação de cada componente XYZ após a transformação da imagem RGB. Na figura é possível observar que a chama não apresenta características de intensidade que permitam distinguir a chama do outros pontos da imagem em

nenhum dos componentes, mas este espaço de cores tem grande relevância para o trabalho, uma vez que este é usado como base para a transformação de dois outros espaços de cores que serão apresentados a seguir.



Figura 2.10: Transformação do espaço de cores RGB para o espaço de cor CIE XYZ. (A) Imagem original no espaço RGB; (B) Imagem de intensidade referente ao componente "X"; (C) Imagem de intensidade referente ao componente "Y"; (D) Imagem de intensidade referente ao componente "Z". Nesse espaço de cor não existe uma caracterização específica das chamas em relação a outros objetos da cena. Imagem original retirada de Toreyin et al. (2006)

Espaços de cores CIE Lab* e Luv

O CIE Lab e o CIE Luv são ambos derivados do espaço de cores CIE XYZ, a idéia por trás desses dois espaços de cores é criar um espaço que seja perceptivelmente uniforme, isto é, uma escala uniforme onde dois pontos no plano de cores sejam equivalentes a diferença visual entre as cores.

Entre seus atributos, o mais importante é sua independência de dispositivos, o que significa que as cores são definidas independentes de sua natureza de criação ou dispositivo que estão sendo usados para a representação.

Para o CIE Lab, a componente "L" é um equivalente próximo da percepção humana da luminosidade variando entre 0 para preto e 100 para o branco e os componentes "a" e "b" variam de -128 a 127, para o quais os valores negativos representam o verde e azul respectivamente para cada canal e valores positivos representam o vermelho e

o amarelo. O CIE Lab usa transformações não lineares do espaço XYZ, logo o espaço pode ser calculado fazendo primeiramente a transformação referente à equação 2.3 seguido por:

$$X = X/X_n, \text{ onde } X_n = 0.950456$$

$$Z = Z/Z_n, \text{ onde } Z_n = 1.088754$$

$$L = \begin{cases} 116 \times \mathrm{Y}^{1/3} - 16 & \text{para Y} > 0.008856 \\ 903.3 \times Y & \text{para Y} \leq 0.008856 \end{cases}$$

$$a = 500(f(\mathrm{X}) - f(\mathrm{Y})) + 128$$

$$b = 200(f(\mathrm{Y}) - f(\mathrm{Z})) + 128$$

onde

$$f(t) = \begin{cases} t^{1/3} & \text{para Y} > 0.008856 \\ 7.787t + 16/116 & \text{para Y} \leq 0.008856 \end{cases}$$

e

$$delta = \begin{cases} 128 & \text{para imagens de 8 bits} \\ 0 & \text{para imagens de ponto flutuante} \end{cases} \tag{2.4}$$

O CIE Luv é basicamente usado em aplicações de computação gráficas para a reprodução de efeitos luminosos. Seu componente "L" é análogo ao componente "L" do Lab, porém seus componentes "u" e "v" variam os seus valores entre -134 a 120 e -140 a 122 respectivamente. O componente "u" representa uma faixa de transição da cor verde até o vermelho, enquanto o componente "v" representa o magenta para o azul. Assim como o CIE Lab, o CIE Luv usa transformações não lineares do espaço XYZ, sendo o cálculo da transformação realizada sobre o espaço de cores XYZ (equação 2.3) segundo:

$$L = \begin{cases} 116 \times \mathrm{Y}^{1/3} - 16 & \text{para Y} > 0.008856 \\ 903.3 \times \mathrm{Y} & \text{para Y} \leq 0.008856 \end{cases} \tag{2.5}$$

$$u' = 4 \times \mathrm{X}/(\mathrm{X} + 15 \times \mathrm{Y} + 3\mathrm{Z})$$

$$v' = 9 \times \mathrm{Y}/(\mathrm{X} + 15 \times \mathrm{Y} + 3\mathrm{Z})$$

$$u = 13 \times L(u' - un) \text{ onde } un = 0.19793943$$

$$v = 13 \times L(v' - vn) \text{ onde } un = 0.46831096$$

As figuras 2.11 e 2.12 mostram as transformações de uma imagem para os espaços de cores CIE Lab e CIE Luv respectivamente. Na figura do CIE Lab, a região das chamas se destacam levemente do restante da cena no componente "a", responsável pela representação da cromaticidade vermelha. No caso do CIE Luv, o componente "u" mostra claramente valores de intensidade altos na região das chamas em relação ao restante da cena.

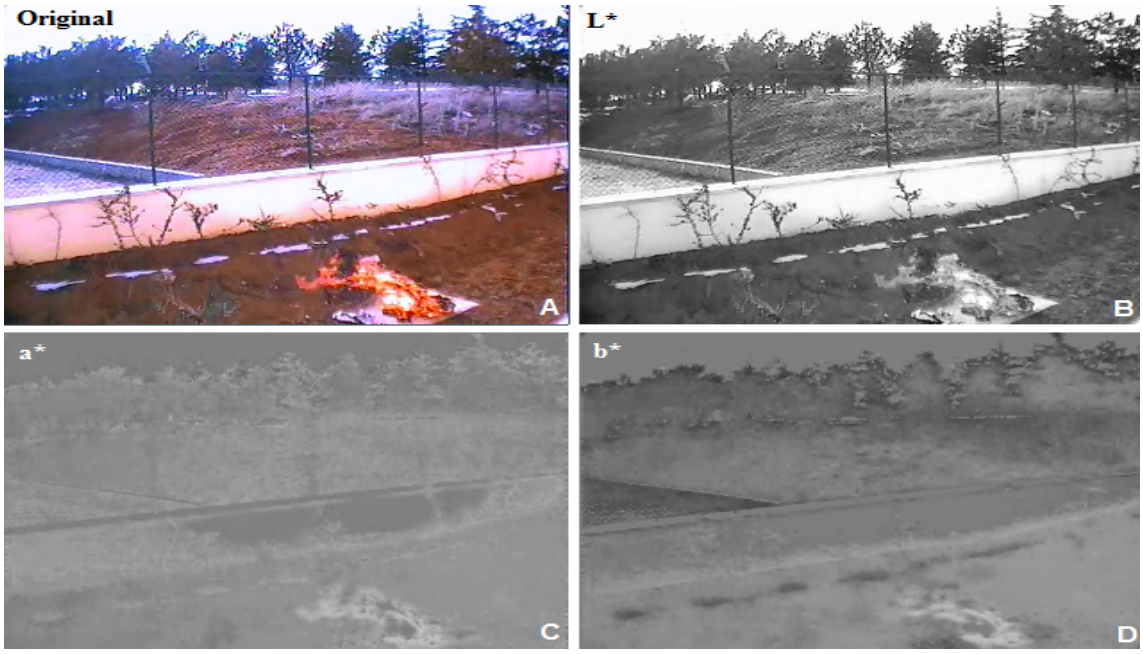


Figura 2.11: Transformação do espaço de cores RGB para o espaço de cor CIE Lab. (A) Imagem original no espaço RGB; (B) Imagem de intensidade referente ao componente "L"; (C) Imagem de intensidade referente ao componente "a"; (D) Imagem de intensidade referente ao componente "b". Chama possui valores de intensidade levemente mais altos que o restante da cena no componente "a". Imagem original retirada de Toreyin et al. (2006)

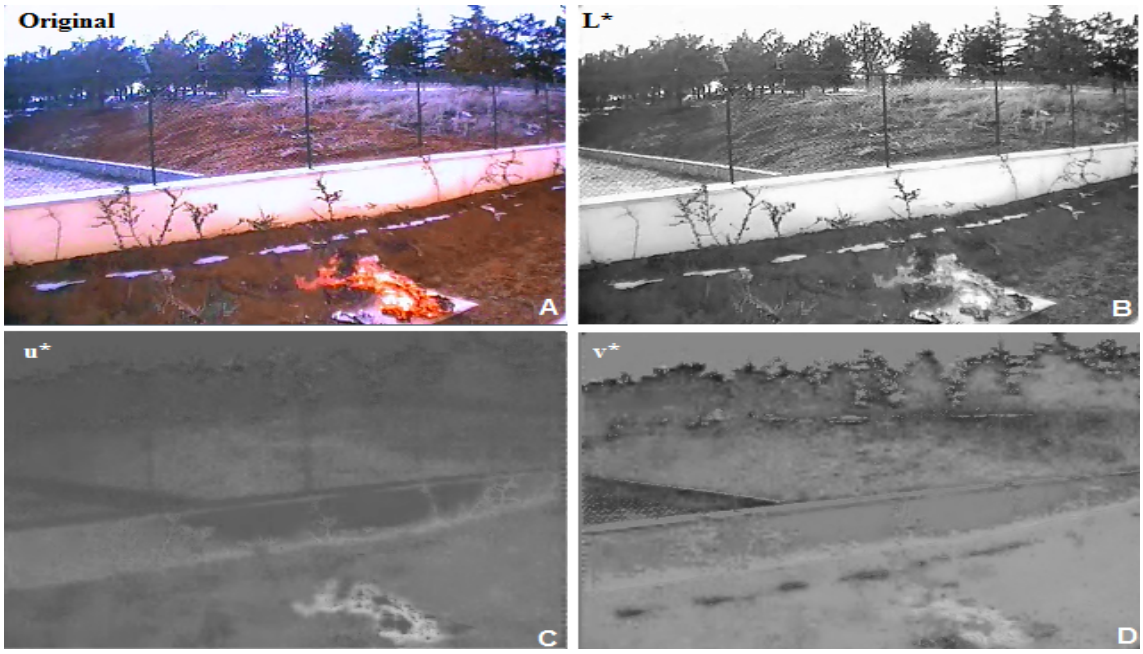


Figura 2.12: Transformação do espaço de cores RGB para o espaço de cor CIE Luv. (A) Imagem original no espaço RGB; (B) Imagem de intensidade referente ao componente "L"; (C) Imagem de intensidade referente ao componente "u"; (D) Imagem de intensidade referente ao componente "v". Chama possui valores altos de intensidade no componente "u". Imagem original retirada de Toreyin et al. (2006)

2.3 Avaliação de características: Humanos x Máquinas

Antes de passar adiante no estudo de técnicas citadas na literatura para a classificação e detecção de focos de incêndios, é necessário avaliar algumas características humanas e computacionais de forma a levantar algum conhecimento geral sobre as limitações encontradas ao se utilizar visão computacional para resolver problemas de classificação de padrões, simulando o ser humano através de técnicas matemáticas e algoritmos.

É sabido que seres humanos e animais são extremamente eficientes no processamento de informação visual, levando várias características em consideração para avaliar e interagir com o ambiente ao seu redor. Neste quesito, são levantadas várias questões filosóficas sobre o funcionamento do cérebro na interpretação destas informações visuais, uma vez que o sistema visual não é completamente compreendido e existem diversas subjetividades que fazem com que a percepção visual varie de uma pessoa para outra.

A teoria de Gestalt ou psicologia da forma é uma das mais citadas e conhecidas para discutir e avaliar o funcionamento do sistema visual (bem como outros sistemas sensoriais), referindo-se a um processo de dar forma ou configurar o que é colocado diante dos olhos. Para exemplificar o que foi dito, avalie a figura 2.13.

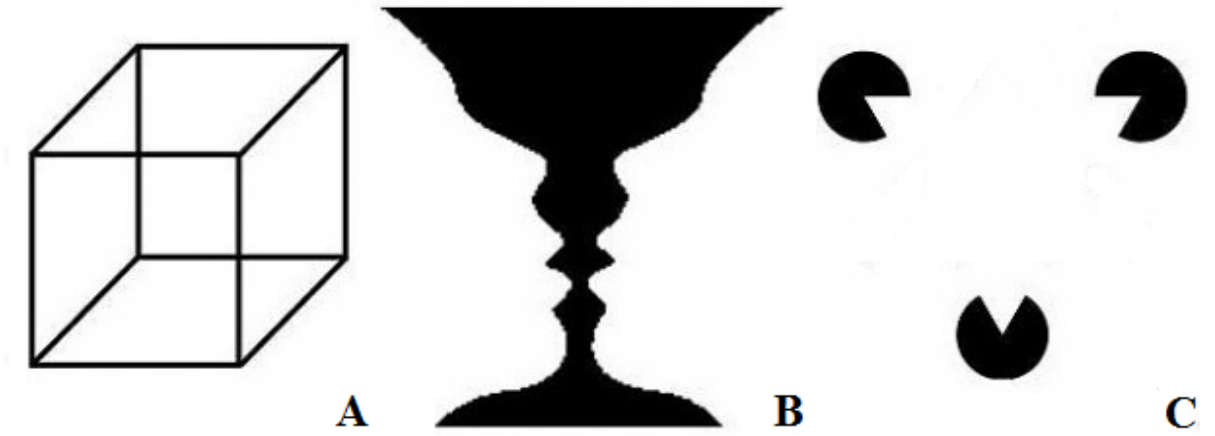


Figura 2.13: Objetos de ilusão de ótica: Cubo de Necker, Vaso de Rubin e Triângulo de Kanisza. (A) Cubo de Necker: permite uma dupla visualização como um cubo orientado em profundidade da esquerda para direita ou da direita para esquerda; (B) Vaso de Rubin: silhueta de um vaso ou a silhueta de duas faces humanas; (C) Triângulo de Kanisza modificado: é possível visualizar um triângulo mesmo que este não esteja representado na figura.

Na figura 2.13 para o cubo de Necker é possível visualizar dois cubos distintos, no entanto, um ser humano não consegue visualizar os dois ao mesmo tempo. De maneira semelhante, o vaso de Rubin permite visualizar uma silhueta de um vaso facilmente, mas enxergar as silhuetas das duas faces pode necessitar de um pequeno estímulo, ainda assim, não sendo possível visualizar as duas formas simultaneamente. O triângulo de Kanisza talvez seja uma das figuras mais importantes dentro do Gestaltismo, pois é partir dela que se prova a afirmativa feita por Kurt Koffka (um dos pioneiros na área): "O todo é maior que a soma de suas partes". Nesta figura não existe qualquer menção

a um triângulo, no entanto, todos o vêem como se fosse um objeto sólido.

Continuando a avaliar esta figura, levando em consideração apenas um círculo e ignorando os outros, uma pessoa pode ainda abstrair uma pizza (caso esteja com fome) ou ainda se ela pertence a geração X (primeira geração a crescer com os *video games*, que consiste das pessoas nascidas entre 1960 e meados de 1980) ou posterior, esta pessoa vai dizer que o círculo é o personagem do jogo Pac-Man. Não obstante, não demora mais do que 5 segundos após esta informação ser inteligível para que uma pessoa desta geração veja o triângulo de uma forma ainda mais abstrata baseada em seus conhecimentos: São 3 Pac-Man tentando engolir um triângulo, ou ainda Pac-Man e sua família (Ms. Pac-Man e Pac-Man Jr.) jantando!

Voltando para a figura do vaso de Rubin, não visualizamos apenas a silhueta, mas também imaginamos inconscientemente sua profundidade dado suas proporções. Outro exemplo clássico está no exemplo das letras r, o, s, a, em que estas letras não são apenas uma palavra solta em nossas mentes, mas evocam o formato, cheiro e simbolismo desta em função de nossa experiência vivida. Ainda, de maneira semelhante somos perfeitamente capazes de entender a seguinte frase: "V0c3 c0n536u3 l3r 1ss0".

Ao avaliar estas situações vemos o quão complexo é o sistema visual e a dificuldade de se "materializar" isso em fórmulas e algoritmos. Para um computador analisar apenas o cubo de Necker, isto representa um problema complexo, pois como poderia ser tomada a decisão de decidir por um cubo ou outro visualizado na cena, uma vez que um ser humano não é capaz?

Devido a esta complexidade, programas que utilizam visão computacional se baseiam em modelos "primitivos", que podem ser matematicamente viáveis e plausíveis de serem executados por uma máquina. A união desses modelos primitivos permite avaliar e decidir situações complexas dentro da limitação de processamento atual das máquinas.

Um exemplo clássico de inteligência artificial que demonstra essa idéia está nas redes neurais, onde um perceptron simples, que representa analogamente a um neurônio, consegue classificar uma base de dados através de uma simples reta no espaço, dividindo estes dados em dois grupos distintos. Ao se criar uma rede com vários perceptrons simples com múltiplas camadas, estes são capazes de classificar base de dados de difícil separação, criando subespaços de formas complexas e amorfas que separam um grupo de outros. Detalhes sobre a técnica podem ser vistas em Braga (2007). Embora seja fácil para os seres humanos observar e separar rapidamente objetos no espaço, para uma máquina este geralmente é um processo lento e extenuante, pois são necessários diversos cálculos para que este resultado possa ser alcançado. Dentro dessa limitação de desempenho deve-se atentar para o contexto proposto em que o algoritmo a ser gerado deve trabalhar.

No caso deste trabalho, é importante usar características que sejam de fácil extração e que permitam um cálculo rápido o suficiente para que o algoritmo possa executar sua tarefa utilizando imagens em alta resolução. Esse tipo de abordagem acaba por trazer

outros problemas que devem ser levados em consideração no trabalho: o tempo gasto para processar informações extras em resoluções mais altas e o ganho obtido por estas tanto para máquina quanto para o ser humano.

A priori se sabe que quanto maior a resolução mais fácil é para o ser humano avaliar o conteúdo da imagem e tomar decisões. As informações visuais obtidas durante e após um incêndio podem levar a estimativas do comportamento do fogo que permitem inferir diversas características do ambiente sendo monitorado, um exemplo disso está na altura de crestamento ou estado de carbonização dos troncos das árvores. Conforme mostrado em Gonzalez e Woods (2010), essa facilidade de extrair informações de imagens por seres humanos se deve a fenômenos visuais como, por exemplo, as bandas de Mach que ilustram como o sistema visual humano realça as bordas dos objetos e o contraste simultâneo que se refere à maneira como duas cores diferentes se afetam diretamente. Esses fenômenos são de difícil replicação em máquinas, pois envolvem certas subjetividades relacionadas ao funcionamento do cérebro humano.

A figura 2.14 ilustra os fenômenos descritos, sendo importante ressaltar que a resolução da imagem pode afetar o efeito destes fenômenos na visão humana.

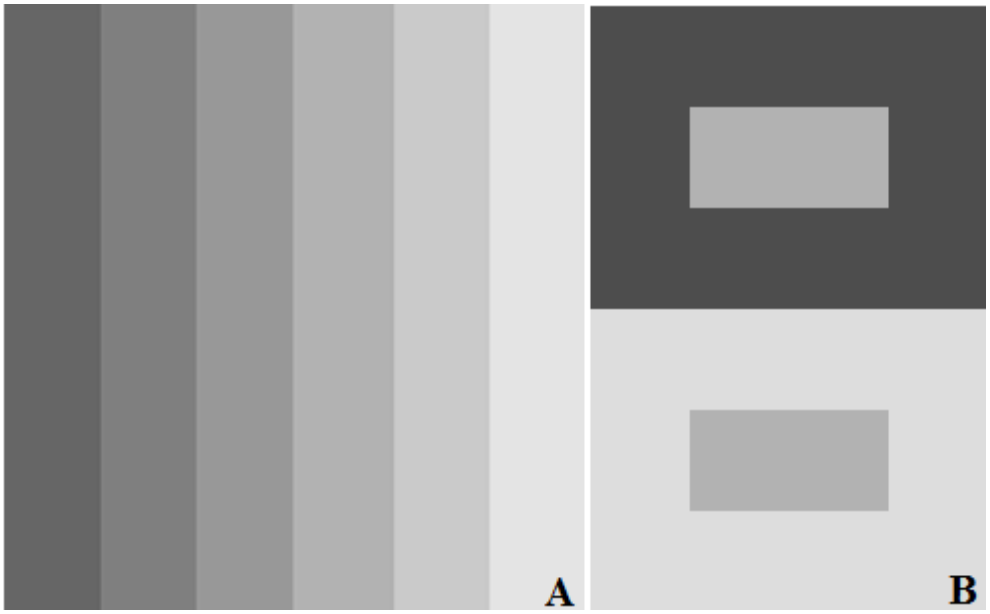


Figura 2.14: Contraste simultâneo e Bandas de Mach. (A) Bandas de Mach, as linhas possuem um valor constante, mas o brilho se altera visivelmente próximo as bordas, efeito esse gerado pelo nosso sistema visual; (B) Contraste simultâneo, os retângulos do meio possuem a mesma intensidade em ambas figuras, entretanto o inferior é visto mais escuro. FONTE: Gonzalez e Woods (2010)

Entretanto, uma resolução mais alta do ponto de vista computacional pode não significar um aumento significativo do desempenho do algoritmo, já que este faz uso apenas das características "primitivas" para avaliar o meio, classificando e tomando decisões a partir dessas informações limitadas.

Ao avaliar o que foi descrito sobre as características do fogo anteriormente, pode-se separar duas características principais e "primitivas" em relação a este elemento: A sua cor e o seu movimento. A cor da chama geralmente irá variar do amarelo claro para laranja-avermelhado nas suas bordas, à medida que a chama perde calor e produz menos energia por estar mais longe da fonte de combustível. O movimento sempre será contínuo devido à ação das correntes de convecção geradas pelo calor extremo da reação, o que faz com que a chama geralmente assuma grandes proporções (enquanto houver uma quantidade de combustível suficiente) e formas aleatórias. Embora a chama se movimente muito, a distância percorrida pelo fogo de acordo com Batista (1995); Beutling (2009) e outros trabalhos é geralmente inferior a 10 quilômetros por hora, o que indica que a chama geralmente fica presa durante um tempo considerável a uma fonte de combustível, sendo esta a terceira característica "primitiva" a ser avaliada.

Estas características levantadas são suficientes para classificar uma chama de forma computacionalmente barata, conforme será mostrado na próxima seção com as metodologias propostas na literatura para tratar o problema de detecção de focos de incêndio usando imagens.

2.4 Detectores de focos de incêndio

Quando se pensa em detecção de incêndios imediatamente surge a idéia sobre detectores de fumaça ou térmicos. De fato estes tipos de detectores ganharam muita atenção e funcionam dentro das suas limitações de maneira eficiente, evitando grandes prejuízos dentro de áreas urbanas. Estes tipos de sensores podem até serem usados em ambientes externos como estacionamentos e áreas florestais, usando técnicas elaboradas em conjunto com algoritmos.

Dentre alguns trabalhos que usam estes tipos de sensores podemos destacar Yuen e Chow (2005) onde os autores usam a técnica probabilística de simulação de Monte Carlo para tratar incertezas decorrentes dos modelos padrões de distribuição destes sensores, em que uma abordagem determinística geralmente é usada levando a um posicionamento inadequado dos sensores e uma cobertura menos eficiente. Este modelo leva em conta uma importante característica de incêndios, que é a expansão do fogo. De maneira similar, Joglar et al. (2005) propõem outra abordagem probabilística para estes detectores de teto, explicando que a temperatura do detector não é linear ao longo do tempo em relação à temperatura dos gases em contato com ele. Este modelo proposto por estes autores usa as características de velocidade e temperatura dos gases gerados pelo fogo.

Em Singh e Singh (2012), os autores propõem um modelo interessante de grupos de sensores utilizando a idéia de que um deles é líder, sendo responsável por receber as informações e passar para um servidor. Essa técnica permite economizar energia já que apenas um sensor terá que transmitir para longas distâncias. Os dados coletados pelos grupos de sensores são enviados a um servidor que usará a técnica de *fuzzy* para uma tomada de decisão simples, em que a saída é discretizada em níveis de probabilidade

de um incêndio ocorrer.

Embora um sistema com esses tipos de sensores permita tratar o problema de detecção de incêndios em áreas externas, existem vários problemas ao tentar usar este tipo de técnica em campo, entre eles está a área a ser monitorada. Áreas florestais são extensas, de difícil acesso e geralmente com grande densidade de vegetação. Colocar os sensores em locais adequados nestas áreas seria difícil, já que esses detectores funcionam melhor quando instalados em alturas elevadas para captar os gases gerados pelo fogo. Outro problema está no número de sensores a serem instalados, que seria grande, tendo um custo alto por unidade de área monitorada, além de exigir uma manutenção periódica, o que seria impraticável na maior parte das áreas a serem monitoradas.

Devido a estes problemas e outros, outra abordagem deve ser utilizada, e conforme dito anteriormente, graças à evolução das câmeras e dos computadores na última década, o uso de imagens para detecção de eventos tem se tornado cada vez mais comum.

Nesse aspecto um grande número de autores tem usado algoritmos de visão computacional para monitorar todo tipo de evento, como visto em Boult et al. (2001) e Collins et al. (2000), incluindo detecção de incêndios.

Entre os autores pioneiros nesta área estão Cheng et al. (1999), em que os eles demonstram algumas dificuldades encontradas ao se avaliar uma chama em uma imagem de vídeo. Utilizando imagens em escala de cinza, os autores avaliam a luminância dos objetos e explicam que objetos refletores podem possuir luminância próxima a, por exemplo, a chama de uma vela, que é um objeto emissor. Para resolver este problema, os autores propõem o uso de um filtro na câmera que permite capturar o espectro infravermelho próximo e usa esta informação para separar as chamas de outros objetos. Ainda assim, segundo os autores, ainda é possível obter falsos positivos, sendo necessário usar outros critérios para a análise de uma chama, em que propõem usar o número de pixels para avaliar a área da chama ao longo do tempo.

Outro trabalho muito citado na literatura é Chen et al. (2004), aqui os autores tratam o uso das cores para a detecção das chamas de um incêndio, utilizando o espaço RGB e levando em conta as características enunciadas anteriormente de temperatura e cor do fogo, bem como o movimento oscilatório do fogo em função das correntes de convecção. Os autores neste trabalho definem uma regra básica que é usada ou citada por muitos outros autores posteriormente, que é a regra da saturação das chamas no canal vermelho, em que R>G>B. Esta regra consegue definir uma região de fogo com grande sucesso, no entanto, outros objetos podem possuir estas mesmas características levando a falsos positivos. Para resolver este problema, os autores estipulam dois parâmetros limiares para o canal R em que eles controlam a quantidade de saturação da cor vermelha para que um objeto seja considerado fogo, e ainda propõem outro parâmetro de limiar que é responsável por avaliar a diferença da quantidade de pixels com cor de fogo detectados entre dois frames, realizando uma persistência pelo movimento oscilatório da chama, similar ao proposto por Cheng et al. (1999).

Em Toreyin e Çetin (2005), os autores propõem o uso da subtração de fundo para detectar os pixels em movimento e após esta etapa selecionam os pixels com cor de fogo segundo as regras propostas por Chen et al. (2004). Os autores neste trabalho levam em conta a cintilância ou oscilação (*flickering*) do fogo que fica em torno dos 10 Hz. Devido à cintilância do fogo variar constantemente, a técnica de modelos ocultos de Markov (*Hidden Markov Models* (HMM)) é usada para a modelagem probabilística temporal e espacial dos pixels. Os pixels são analisados quanto a sua cor ao longo do tempo (durante 20 quadros), se existe uma alta probabilidade dos pixels pertencerem a uma chama (quando estes se alteram frequentemente), a análise espacial destes pixels é realizada. Nessa etapa, a área correspondente do fogo é avaliada segundo sua variação de cor, que deverá ser maior que outros objetos com cor de fogo, ou seja, a chama possuirá várias tonalidades de cor enquanto objetos sólidos serão praticamente da mesma cor.

Em Marbach et al. (2006), os autores utilizam o espaço de cores YUV e propõem o uso da variação temporal da intensidade dos pixels de fogo (canal "Y"), levando em consideração o movimento da chama e que esta geralmente possui o maior brilho na imagem, gerando uma máscara com possíveis áreas com fogo. Após esta etapa de seleção das áreas candidatas, os autores extraem características referentes à luminância e crominância (canais "U" e "V") que se combinam para determinar a presença ou não de fogo. Embora relatem resultados excelentes, não são dados muitos detalhes sobre os testes realizados, onde o método em si é complexo e dependente de diversos parâmetros a serem controlados.

Schultze et al. (2006) trabalham nas características de cintilância e movimento do fogo, em que afirmam que estas são as informações mais importantes que permitem distinguir o fogo de outros objetos. Os autores também avaliam as informações contidas no som para verificar a frequência da cintilância. Segundo eles, o som do fogo apresenta características semelhantes às encontradas nas imagens, em que o som revela a mesma frequência de cintilância. Os autores fazem a análise do movimento usando imagens gravadas em câmeras de alta velocidade, e ao avaliar os dados obtidos eles mostram o mesmo padrão de cintilância encontradas no som. Mas os autores criticam que apenas o uso destas características não é suficiente para tratar o problema de detecção das chamas, já que em ambientes externos diversos objetos podem simular as condições tratadas.

Toreyin et al. (2006) é outro trabalho muito citado na literatura. De maneira similar ao seu trabalho anterior usando HMM, ele agora usa a transformada de wavelet para realizar a análise espacial e temporal no domínio da frequência. O algoritmo proposto tem a seguinte estrutura:

- Detecção de movimento por subtração de fundo, em que o modelo de médias adaptativas (Donohoe et al. (1988)) é usado com alterações feitas por Collins et al. (2000);
- Detecção dos pixels com cor de fogo usando os canais RGB, porém eles usam uma abordagem estatística onde imagens de amostras com fogo foram segmentadas

manualmente e a partir destes pixels amostrados um modelo de mistura de gaussianas é utilizado para gerar esferas de contenção dos pixels no espaço RGB. Assim caso um pixel esteja dentro de uma das esferas ele possui cor de fogo;

- Análise temporal por wavelet, esta etapa detecta a cintilância do fogo usando as informações contidas no canal vermelho da máscara gerada pelas duas etapas anteriores. Esta análise é feita usando um histórico de vários quadros do vídeo. Estas informações são passadas para o banco de filtros de dois estágios do wavelet. Quando um pixel apresenta alta frequência, o sinal de saída apresenta valores diferentes de zero, no entanto quando um pixel se encontra estacionário ao longo do tempo, os sinais serão zero ou muito próximo a zero, equivalente a um filtro passa-altas;

- Análise espacial por wavelet, esta etapa irá analisar os pixels da máscara gerada pelas duas primeiras etapas quanto à variação de cores nos pixels dentro de uma determinada área. Objetos sólidos terão pouca variação espacial de cor e chamas terão um valor significante.

Um pixel é dito como fogo quando todas as etapas retornam verdadeiro, isto é, o pixel tem movimento, tem cor de fogo, se o pixel varia em alta frequência durante um tempo observado e se este pertence a uma área em que as cores variam significativamente entre os pixels.

Este trabalho possui resultados excelentes, sendo citado e usado para comparações ao longo dos anos pelos seus resultados consistentes. Os autores também utilizam o mesmo em uma solução comercial que está disponível para a comunidade testar e avaliar sua robustez em diferentes situações. Em Toreyin et al. (2007) os mesmos autores geram um algoritmo que usa imagens capturadas em infravermelho (térmica), alertando para os problemas encontrados quando se usa imagens capturadas no espectro da luz visível, como folhas avermelhadas ao vento ou luzes de alerta como usados em carros de bombeiros. Os autores utilizam um híbrido das técnicas usadas nos trabalhos anteriores, onde o HMM em conjunto com wavelet é usado para a análise temporal da cintilância e o wavelet para a análise espacial dos contornos dos objetos.

A hipótese do uso de uma câmera térmica é pertinente ao trabalho, entretanto é necessário frisar que essas câmeras são extremamente caras e de baixa resolução. Dentro do contexto de preservação ambiental, as imagens capturadas por este tipo de câmera têm pouca relevância na extração de características referentes ao meio ambiente, pois mostra em intensidade a temperatura dos objetos capturados, eliminando todas as outras características como textura, cor, etc., que podem ser úteis aos profissionais de monitoração ambiental.

Çelik et al. (2007) usa um modelo de 3 etapas de maneira análoga a alguns trabalhos citados acima, em que o movimento é tratado pela subtração de fundo, a classificação dos pixels segundo regras de cores baseadas no espaço RGB e variações temporais levando o movimento em consideração. O método de subtração de fundo usada pelos autores é a mistura de gaussianas (Stauffer e Grimson (1999)), que também se adapta às mudanças de cenário. As regras para classificação por cores criadas pelos autores

foram extraídas após uma análise estatística de várias imagens de amostras com fogo. Suas regras são análogas às citadas por Chen et al. (2004), sendo adicionadas novas regras que procuram separar a luminância da crominância dos componentes RGB para evitar falsos positivos causados pela mudança constante de iluminação em ambientes externos. De forma a evitar falsos positivos causados por ruído, os autores eliminam qualquer grupo de pixels que seja menor que 5 pixels. As variações temporais são feitas analisando a média dos grupos ao longo dos quadros e observando se os retângulos criados em volta dos grupos aumentam ou diminuem, avaliando a característica do fogo de movimento contínuo.

No trabalho de Çelik e Demirel (2009), os autores tratam o problema de um ponto específico, trabalhando apenas no classificador de pixels pela cor de fogo. Segundo eles a parte crítica da classificação do fogo está nesta etapa, em que um classificador com um bom desempenho irá ditar a resultados obtidos pelo algoritmo. O trabalho consiste primariamente em utilizar um espaço de cor que permita a separação da luminância da crominância, sendo usado o espaço de cores YCbCr para este fim. Os autores deixam claro que a informação mais importante da chama não vem da luminância, mas da crominância, permitindo uma classificação mais robusta. Esta afirmação corrobora com o que foi descrito em Cheng et al. (1999). Novas regras baseadas nas regras geradas por Chen et al. (2004) e Çelik et al. (2007) (R>G>B) são geradas para este espaço de cor, onde "Y" e "Cr" devem ser maiores que "Cb" para chamas. Outras regras baseadas na característica do fogo de objeto emissor também são formuladas, em que o fogo sempre terá um brilho maior que a maioria dos objetos, assim $Y > Y_{medio}$, $Cr > Cr_{medio}$ e $Cb < Cb_{medio}$.

Em outra análise das imagens no espaço de cores YCbCr, os autores percebem que as chamas apresentam uma intensidade maior no componente Cr (quase saturado) do que no componente Cb (próximo ao preto), e propõe uma regra de avaliar os pixels segundo essa diferença, controlado através de um parâmetro de limiar τ ($|Cb - Cr| \geq \tau$). Esse parâmetro é escolhido através da técnica da curva ROC (*ROC Curve - Receiver Operating Characteristic Curve* (Curva ROC)), em que um conjunto de imagens de teste é utilizado e os resultados obtidos pelo classificador neste conjunto em função da variação deste parâmetro indicam qual valor possui o melhor custo-benefício entre taxa de acertos e taxa de erros.

Ainda os autores geram um modelo estatístico no qual os pixels de fogo segmentados manualmente de imagens de amostras são utilizados para gerar um espaço de amostras no espaço dos componentes "Cb" e "Cr" (histograma). Com este espaço, os autores usaram a técnica de mínimos quadrados para estimar polinômios de contenção dos pixels de amostras, desta forma qualquer pixel que estiver dentro desta área delimitada pelos polinômios será considerado um pixel de fogo.

Finalmente, os autores classificam como pixel de fogo aquele que obedecer todas as regras acima. Os resultados obtidos no trabalho mostraram que utilizar um espaço de cores que permite a separação da luminância da crominância reduz significativamente o número de falsos positivos gerados pelo classificador de fogo por cor, caindo quase 50% em relação aos modelos baseados no espaço RGB.

No trabalho Çelik (2010), o autor usa outro espaço de cor para a tarefa de classificação, o CIE Lab. Neste trabalho, o autor trata o movimento com a subtração de fundo e analisa temporalmente as características de movimento do fogo, de maneira semelhante ao seu trabalho de 2007. As regras geradas para este espaço de cores seguem idéias similares ao proposto em seus trabalhos anteriores, em que o fogo apresenta um brilho maior que a maioria dos objetos. Também é feita uma análise estatística em imagens de amostras com fogo segmentadas manualmente, analisando os histogramas em espaços formados pelos diferentes pares de componentes do CIE Lab para indicar probabilisticamente a chance de um pixel pertencer ou não a uma chama, informação esta controlada por um parâmetro de limiar α. Este parâmetro é escolhido de maneira análoga ao parâmetro τ em seu trabalho anterior com o uso da técnica Curva ROC. Um pixel será classificado com cor de fogo quando obedecer todas as regras formuladas. Para a análise de movimento, o crescimento dos pixels em um grupo ou objeto é analisado, incrementando uma variável cada vez que esse número cresce. Usando esta variável, ele define uma regra matemática simples que permite avaliar o movimento e expansão do fogo na cena ao longo do tempo através do uso outro parâmetro de limiar D(t), determinado empiricamente.

Angayarkkani e Radhakrishnam (2010) trabalham com imagens obtidas de satélites em que a transformação para o espaço de cores XYZ é realizada, após esta conversão é feita uma segmentação das áreas com chamas usando a técnica de difusão anisotrópica. Esta técnica permite melhorar a detecção das bordas, suavizando o interior dos objetos, eliminando detalhes espúrios e erradicando ruídos. Infelizmente, esta é uma técnica cara computacionalmente que é usada em situações que a detecção em tempo real não se faz necessária. Os dados obtidos na segmentação são passados para uma rede neural de base radial treinada que irá classificar os pixels segundo sua cor no espaço XYZ e indicando se há fogo ou não na imagem. Imagens de satélite ainda tem as desvantagens de não possuírem uma visualização contínua, podendo ter seu campo de visão obstruído por nuvens e um certo tempo necessário para a disponibilização destas para o uso, impedindo o uso desse tipo de imagem para detecção em tempo real.

Wen et al. (2010) usa um modelo de quantização vetorial (VQ), gerando um *codebook* (livro de códigos ou informações) de cores de fogo e outro de controle (sem cor de fogo). As regras utilizadas para determinar a cor se baseiam no espaço de cores RGB (R>G>B e R>130), sendo dividido em 3 *ranks* ou níveis diferentes de severidade segundo os valores de intensidade de cada canal ou, no caso da composição dos 3 canais, as cores dos pixels. Esta informação é usada por um algoritmo de análise de procura de vizinho mais próximo através de médias, variância e normas (*Equal-average Equal-variance Equal-norm Nearest Neighbor Search* (EEENNS)). Após este algoritmo classificar os pixels segundo sua cor e severidade a detecção por movimento é feita usando o algoritmo de médias adaptativas. Dividindo toda a cena em blocos de 4x4 pixels, os autores avaliam cada um destes blocos ao longo do tempo segundo a quantidade de pixels classificados por grau de severidade, indicando que há fogo quando estes blocos sofrem mudanças constantes em seus graus de severidade ou não há fogo quando não há alterações significantes.

Yuan (2010) trata casos específicos em ambientes internos ou urbanos, onde a detecção de chamas ativa um sistema supressor de incêndios. A técnica proposta avalia primeiramente o movimento na cena através da técnica de subtração de fundo, em seguida a detecção por cor é feita usando a intensidade de cada componente do espaço RGB através de modelos gaussianos, assim quando um fogo se inicia estes valores obtidos pelos modelos variam significativamente, gerando o alarme. Os autores também avaliam a área do fogo, onde se esta área estiver expandindo há grandes chances de incêndio estar ocorrendo. Outro fator importante neste trabalho é a avaliação crítica da técnica proposta pelo próprio autor em que este analisa o uso dela em ambientes externos, em que, neste caso, devido a objetos em constante movimento seria necessário usar vários modelos gaussianos por pixel, o que tornaria o modelo computacionalmente mais caro.

Em Zhou et al. (2010), os autores propõem outra técnica de 3 etapas para resolver o problema de detecção de fogo utilizando o espaço RGB. A primeira etapa consiste em classificar os pixels segundo sua cor, comparando cada pixel a um parâmetro de limiar por componente do espaço (Rt, Gt e Bt), definidos experimentalmente pelos autores. Os pixels são contabilizados e comparados ao número de pixels contabilizados no quadro anterior (se houver) e a outro parâmetro de limiar de quantidade mínima de pixels. Caso os valores obtidos estejam crescendo ao longo do tempo, outras etapas serão realizadas nos quadros subsequentes.

A segunda etapa seleciona a área de fogo em cada quadro. Um pixel é classificado como fogo se este obedece as regras propostas por Chen et al. (2004). A terceira etapa consiste em usar a máscara gerada pelos pixels classificados como fogo para eliminar grupos pequenos e falhas nas chamas (buracos) através das técnicas de erosão e dilatação, técnicas comuns em algoritmos de visão computacional que utilizam pixels vizinhos para aumentar ou diminuir a área de um objeto (detalhes podem ser vistos em Bradski (2008)). Após realizar estes dois processos, a extração de bordas é feita usando a técnica de extração de bordas de Canny. Estas bordas são então utilizadas para calcular a área de expansão da chama, o perímetro de expansão e a permanência destas bordas em uma área. Estes três valores são comparados a outros 3 parâmetros de limiares que vão determinar ou não a presença da chama naquele vídeo.

Wang et al. (2010) trabalham com o mesmo esquema de 3 etapas, em que se detecta o movimento na cena, detecta os pixels com cor de fogo e remove regiões sem fogo das regiões candidatas a fogo. O autor usa o algoritmo de subtração de fundo por médias adaptativas para detectar o movimento e a regras propostas por Chen et al. (2004) para a classificação dos pixels segundo sua cor. Estes pixels são então analisados segundo sua cintilância usando HMM e transformada de wavelet para analisar as características espaciais e temporais, semelhante ao proposto em Toreyin e Çetin (2005). Uma pequena alteração de movimentar o ponto de observação dinamicamente (que foca a borda da chama) foi feita para compensar as variações de tamanho do fogo em ambientes externos. Outra metodologia chamada mapa de luminância é incluída nesta etapa. Ela trata a característica do fogo de possuir mais brilho que os demais objetos da cena e é feita usando várias técnicas matemáticas, avaliando a variação da luminância ao longo do

tempo na cena. Os autores não levam em conta o tempo de processamento gasto nesse trabalho.

Em Ko et al. (2010), os autores também usam 3 etapas para a detecção de fogo. A primeira etapa consiste na detecção de movimento usando a subtração de fundo por médias adaptativas, em que o parâmetro limiar é calculado dinamicamente. Usando o espaço RGB normalizado, modelos probabilísticos são criados na segunda etapa, usando distribuição de Gaussianas e os pixels são classificados segundo sua cor. A terceira etapa é feita usando redes Bayesianas em conjunto com wavelet para avaliar as características espaciais e temporais.

Hou et al. (2011) propõem um novo algoritmo para detecção, sendo usadas redes neurais Fuzzy (FNN) e máquinas de vetores de suporte com Fuzzy (*Fuzzy Genectic Algorithm Least Square Support Vector Machine* (FGALSSVM)) nos pixels detectados com movimento, em que é usado um algoritmo de subtração de fundo modificado chamado NAUTEA, que consiste em subtrair a imagem de fundo duas vezes para um melhor desempenho. Um fato importante é que o algoritmo trata os objetos detectados como: sem fogo, suspeito e fogo. Esse algoritmo apresenta um tempo extenso de treinamento para que uma chama possa ser detectada, embora apresente bons resultados.

Verstockt et al. (2012) faz uso de novos dispositivos alternativos aos sensores infravermelhos. Essas câmeras recebem o nome de *time-of-flight* (TOF) e são capazes de captar informações 3D, com funcionamento análogo ao de um radar, em que emissores infravermelhos enviam sinais modulados e uma lente capta a luz refletida pelos objetos, gerando um mapa de profundidade (distância) e amplitude (refletância). Estas câmeras ainda possuem um uso limitado devido a sua baixa resolução (176 x 144 pixels) e alcance reduzido (<10 metros). No caso da detecção de fogo em áreas externas apenas a imagem de amplitude pode ser usada, mas essa informação pura não é suficiente, sendo necessário o uso de imagens normais no espaço RGB. Estas imagens em RGB são segmentadas pelo movimento usando a técnica de subtração de fundo de médias adaptativas proposta por Collins et al. (2000) e 3 características são extraídas destes pixels: Desordem da caixa limítrofe (*Bounding Box Disorder* (BBD)), desordem de orientação principal (*Principal Orientation Disorder* (POD)) e razão da cor de fogo (*Fire Color Ratio* (FCR)). O BBD se refere ao máximo e mínimo local dos pixels referente a área de contenção do objeto, o POD avalia o ângulo do objeto em relação ao eixo x (o autor deixa a entender que este valor é transformado em unidades de pixel) e o FCR a razão entre o total de pixels do objeto e os pixels com cor de fogo segundo as regras de Chen et al. (2004). A média dos valores obtidos pelas 3 características é calculada e comparada a um limiar que irá determinar se há fogo ou não no quadro. A imagem de amplitude é usada para distinguir objetos de movimento rápido na cena (bordas das chamas), utilizando 3 quadros para realizar esta tarefa através de operações simples de soma e subtrações pixel a pixel. Como segundo método de avaliação de movimento, a transformada de wavelet é usada aqui também, onde o fogo apresentará pixels saturados nas imagens de características resultantes da transformada. Um pixel é dito pertencente a uma chama quando este apresenta um valor de amplitude maior que 0, valores saturados nas imagens resultantes do wavelet e uma média das características

obtidas no espaço RGB superior ao limiar informado pelo usuário.

Conforme enunciado anteriormente no trabalho, a maior parte dos trabalhos consiste primordialmente das características de cor e movimento da chama para efetuar a detecção, variando apenas a metodologias aplicadas para extrair e analisar essas características. Geralmente, os passos abordados fazem uso de alguma metodologia matemática complexa, que tem um alto custo computacional quando se leva em conta o aumento exponencial do número de pixels à medida que a resolução aumenta. Isso leva a acreditar que por esse motivo os autores apenas mostram resultados em resoluções de 320 x 240 pixels ou inferiores. Como o objetivo aqui é gerar um algoritmo que seja capaz de trabalhar em diferentes resoluções com preferência pelas mais altas por motivos citados anteriormente, as características e metodologias usadas deverão ser as mais simples e eficientes possíveis para permitir alcançar este objetivo, necessitando a criação e uso de algumas técnicas que permitam esse funcionamento.

Síntese do capítulo

Nesse capítulo foram mostradas informações básicas sobre o fogo e suas características mais relevantes, bem como os fatores de origem e influência destas informações sobre as características. Um breve relato foi feito sobre como é realizada a captura das imagens pelas câmeras e como se dá a representação dessas imagens em diversos espaços de cores, gerados a partir de informações e características derivadas do funcionamento conhecido da visão humana. Um levantamento sobre as limitações de avaliação de características das máquinas em relação ao ser humano foi feito, visto que o processo da visão humana ainda não é totalmente compreendido e que muitas vezes a chave para reproduzir um sistema biológico está na simplicidade das informações analisadas. Finalmente, trabalhos realizados na literatura mostram que as características mais simples geram os melhores resultados, mesmo que embora alguns autores utilizem técnicas computacionalmente caras ao preço de algum ganho de desempenho pertinente à visualização do contexto abordado. Graças a essas diferenças de entendimento e situações abordadas, diferentes espaços de cores são usados para extrair características referentes à cor de diversas maneiras, bem como diversas idéias são geradas para resolver o problema de maneira satisfatória. Algumas destas idéias serão aproveitadas ao longo do próximo capítulo, que irá descrever as metodologias e idéias abordadas para o desenvolvimento do método D383F.

Capítulo 3

Desenvolvimento

O trabalho aqui proposto tem como objetivo a criação de um algoritmo que seja capaz de efetuar a detecção de focos de incêndio automaticamente, dentro do contexto de ser usado para processar imagens capturadas por dispositivos montados em torres que compõe a rede de transmissão de energia elétrica. Essa rede transpõe grandes distâncias do território brasileiro, incluindo áreas florestais que são em sua maioria de difícil acesso para a monitoração por pessoas. Dentro desse contexto, vários problemas devem ser levados em consideração, como mudanças constantes de luminosidade, oscilação da torre de transmissão, cores dos objetos sendo capturados na cena, alterações causadas por ventos e outros eventos atmosféricos.

O método D383F proposto neste trabalho consiste de 3 diferentes etapas: a segmentação pelo movimento, classificação dos pixels segundo sua cor e a persistência temporal e espacial dos objetos segmentados.

Na primeira etapa é realizada a separação dos objetos com relação ao movimento, sendo observados alguns critérios quanto à qualidade da imagem e distância de cobertura do campo de visão do dispositivo de captura. Esta etapa é de grande importância para a classificação adequada pelas próximas etapas, já que ela seleciona a partir de seus parâmetros os pixels com movimento relevantes ao processo de classificação, diminuindo o custo computacional exigido pela tarefa de classificação dos pixels. Ressaltando que esta etapa é feita utilizando os canais de cores do RGB, buscando maximizar a captura das chamas. Posteriormente a sub-amostragem de pixels é anexada a esta etapa, permitindo que o algoritmo possa trabalhar em tempo real com imagens em resoluções mais altas que 352x288 pixels.

A segunda etapa consiste na classificação dos pixels cor de fogo, que é responsável por eliminar os objetos que se movimentam pela cena, mas que não possuam cor de fogo. Nesta etapa é utilizado o método proposto por Çelik e Demirel (2009) com alteração, em que sua quinta regra foi eliminada e um classificador baseado em outro espaço de cores foi criado para gerar um comitê de classificadores, aumentando a eficiência desta etapa. É importante enfatizar que após estas duas etapas citadas, objetos com cor de fogo em movimento ainda são capturados e registrados como fogo, sendo necessária outra etapa para tratar estes tipos de objetos.

A persistência temporal e espacial é a terceira etapa, que utiliza uma idéia parcialmente baseada na técnica de colônia de formigas referente ao depósito e evaporação de feromônios liberados pelas formigas. O objetivo desta etapa é avaliar os objetos capturados com relação a duas características físicas importantes do fogo: movimento aleatório em torno do combustível e tempo de duração do mesmo. Assim, uma grande parte dos falsos positivos pode ser eliminada dentro do contexto de utilização proposto para o algoritmo, no qual incêndios florestais geralmente apresentam chamas de grande porte e estas duram longos períodos de tempo em relação a outros objetos que geram falsos positivos.

3.1 Base de dados e medidas de desempenho

3.1.1 Base de dados

Antes de definir as etapas do método D383F, é necessária a criação de uma base de dados para validar e testar as idéias propostas para resolver o problema de classificação de focos de incêndio. Não há uma base de dados padrão para estes tipos de testes, conforme Ko et al. (2010) relatam em seu trabalho, mas alguns autores disponibilizam os vídeos usados em seus trabalhos para a comunidade. Entre esses está Toreyin, que disponibiliza no site de seu sistema VisiFire, o grupo de Ko pertencente a Laboratório de Visão Computacional e Reconhecimento de Padrões da Universidade de Keimyung da Coréia (CVPR Lab) e o grupo *Context Aware Vision using Image-based Active Recognition* (CAVIAR) da Universidade de Edimburgo na Inglaterra.

Para este trabalho, bases de dados de imagens estáticas e de vídeos foram criadas e denominadas com o propósito de avaliar os objetivos propostos e o funcionamento do algoritmo da melhor forma possível. A tabela 3.1 mostra resumidamente as bases geradas e seu objetivo ao longo do trabalho.

Base de dados	**Objetivo**
ImaSta	Conjunto de imagens estáticas usadas para diversos testes.
VideoRes	Conjunto de vídeos para avaliar o desempenho de detecção de fogo em diversas resoluções em âmbito florestal.
VideoDiv	Conjunto de vídeos com fogo para avaliar a detecção em ambientes diversos.
VideoDivCtrl	Conjunto de vídeos sem fogo para avaliar a detecção em ambientes diversos.
VideoDis	Conjunto de vídeos para avaliar a detecção com chama saturada e teste de distância máxima de detecção.

Tabela 3.1: Tabela resumo das bases de dados.

Para a base de imagens estáticas, ImaSta, 100 imagens de tamanhos variados foram recolhidas na internet, das quais 50 delas apresentam fogo e as outras 50 não possuem fogo, porém, possuem objetos com cor de fogo para dificultar a classificação pelos métodos, bem como apresentam condições variadas de iluminação, simulando

condições reais de captura de imagens.

Este tipo de amostragem (amostras positivas e amostras de controle) permite uma avaliação mais eficiente do experimento já que não é necessária a classificação manual dos objetos detectados na cena a posteriori para definir os verdadeiros positivos e falsos positivos capturados pelos métodos na mesma cena como alguns autores na literatura fazem. Mesmo que um falso positivo ajude na detecção de uma chama nas imagens com fogo, o método provavelmente irá gerar falsos positivos nas imagens de controle, comprometendo seu desempenho geral. A figura 3.1 mostra algumas das imagens utilizadas.



Figura 3.1: Base de dados de imagens estáticas: (A-E) Imagens de controle com objetos capturados na cena que possuem cor de fogo; (F-J) Diferentes imagens contendo fogo

Para a base de dados de vídeos, 4 diferentes conjuntos são usados conforme mostrado na tabela. O primeiro conjunto, VideoRes, é formado por 10 vídeos colhidos da internet com 10 segundos de duração a uma taxa de 10 quadros por segundo (100 frames por vídeo), dos quais 5 deles possuem fogo em todos os seus quadros e os outros 5 não possuem fogo, mas possuem objetos com cor de fogo para dificultar e testar o desempenho dos métodos, de maneira análoga à base de imagens estáticas. Esses vídeos também possuem diferentes distâncias de cobertura na cena e diferentes condições de iluminação.

Os vídeos da base VideoRes foram obtidos na resolução de 1920 x 1080 pixels e foram convertidos (*downsampling*) para 7 outras resoluções menores de forma a avaliar o desempenho dos métodos e o tempo gasto de processamento sob diferentes resoluções. As resoluções utilizadas nesses 10 vídeos são as seguintes: 1920 x 1080 *Full High Definition* (Full HD), 1280 x 720 *High Definition* (HD), 800 x 600 *Super Video Graphics Array* (SVGA), 704 x 576 4 X CIF (4CIF), 640 x 480 *Video Graphics Array* (VGA), 352 x 288 *Common Intermediate Format* (CIF), 320 x 240 *Quarter Video Graphics Array* (QVGA), 160 x 120 *Quarter QVGA* (QQVGA).

É importante salientar que este conjunto foi cuidadosamente montado para testar diferentes situações a serem tratadas pelos métodos em condições reais de uso no meio externo, como por exemplo: dia e noite, objetos próximos e distantes, movimento de objetos com cor de fogo de origem natural ou criada pelos humanos e incidência de

raios solares sobre objetos. A tabela 3.2 exibe as descrições dos vídeos (VR) integrantes do conjunto VídeoRes e a figura 3.2 mostra as capturas de tela ou *screenshots*.

Vídeo	Descrição do Vídeo
VR 1	Incêndio florestal com chamas laranja e fumaça
VR 2	Incêndio florestal com chamas laranja e fumaça com *zoom*
VR 3	Incêndio em uma montanha (longa distância)
VR 4	Incêndio florestal à noite (longa distância)
VR 5	Incêndio florestal à noite
VR 6	Controle, incidência de luz solar na cena (grama amarelo fogo)
VR 7	Controle, árvores com folhas cor laranja ao vento (distante)
VR 8	Controle, galhos balançando em frente ao pôr do sol (perto)
VR 9	Controle, leito de um rio com diversos objetos de cor vermelha na cena
VR 10	Controle, galho de cor vermelho escuro balançando ao vento (perto)

Tabela 3.2: Descrição dos vídeos que compõem o conjunto VídeoRes.



Figura 3.2: Conjunto de vídeos VídeoRes: (VR1-VR5) Vídeos com incêndios variando em distância e condições de iluminação, inclusive à noite; (VR6-VR10) Vídeos com objetos capturados na cena que possuem cor de fogo e com movimento, como folhas, galhos, etc.

A base VideoDiv foi gerada para realizar testes comparativos entre método de Toreyin et al. (2006) e o método D383F, e consiste de 10 vídeos com fogo retirados da base disponibilizada por Toreyin, sendo todos em ambiente externo com resolução de 320 x 240 pixels a uma taxa de 10 quadros por segundo com tempo de duração variada, possuindo diferentes tipos de fogo e situações como: queima de diferentes tipos de materiais combustíveis, queimadas controladas e incêndios.

Outros 3 vídeos com incêndios florestais foram adicionados à base VideoDiv, sendo um deles de dia com apenas um foco de incêndio à vista e os outros dois com incêndios à noite para adicionar mais diversidade e diferentes condições de iluminação. Ainda, foram adicionados mais dois vídeos obtidos da internet referentes a um incêndio no estádio da cidade de Bradford na Inglaterra, que aconteceu no dia 11 de maio de 1985. Esses vídeos foram gerados pelas emissoras de televisão e complementam a base de

vídeos para avaliar os métodos com relação à capacidade dos mesmos de serem utilizados fora do escopo de incêndios florestais, mas ainda no escopo de ambientes externos. Esses 5 vídeos extras seguem o mesmo padrão de resolução e taxa de quadros por segundo dos vídeos de Toreyin. No total, a base VideoDiv possui 15 vídeos e 4330 quadros de vídeo com fogo. A tabela 3.3 mostra as descrições dos vídeos (VD) e a figura 3.3 mostra capturas de tela dos mesmos.

Vídeo	Descrição do Vídeo
VD 1	Fogo em uma caixa
VD 2	Madeira queimando
VD 3	Fogo com uma pessoa andando em volta (longa distância)
VD 4	Bombeiros gerando uma queimada
VD 5	Árvore queimando de forma controlada
VD 6	Incêndio florestal (perto)
VD 7	Incêndio florestal (meia distância)
VD 8	Incêndio florestal (distante)
VD 9	Incêndio florestal (longa distância)
VD 10	Incêndio florestal (longa distância)
VD 11	Incêndio florestal com apenas um foco visível
VD 12	Incêndio florestal a noite (meia distância)
VD 13	Incêndio florestal a noite (longa distância)
VD 14	Incêndio no estádio de Bradford (visão lateral)
VD 15	Incêndio no estádio de Bradford(visão frontal)

Tabela 3.3: Descrição dos vídeos que compõem o conjunto VídeoDiv.



Figura 3.3: Conjunto VídeoDiv: (VD1-VD10) Vídeos disponibilizados por Toreyin; (VD11-VD13) Vídeos com incêndios complementares variando a iluminação; (VD14-VD15) Vídeos do incêndio no estádio de Bradford em 1985.

De forma a manter o segundo experimento homogêneo e aumentar a quantidade de situações a serem avaliadas pelos métodos, outro conjunto chamado VídeoDivCtrl foi criado com 15 vídeos de controle secundários, sendo 4 deles parte da base do grupo CAVIAR, em que pessoas com roupas de diferentes cores transitam em um saguão com incidência de luz solar gerando desafios para os métodos utilizados. Os 11 vídeos restantes correspondem a vídeos de sistemas de vigilância que foram obtidos da internet, abrangendo diversos tipos de cenas como: estacionamentos, saída de um supermercado com diversos objetos e carros com cor de fogo, um complexo industrial, entre outros. Esses vídeos não possuem fogo em toda sua extensão, estão na resolução de 320 x 240 pixels a 10 quadros por segundo, com um total de 4550 quadros de duração. A descrição e a captura de tela destes vídeos (VDC) podem ser vistos respectivamente na tabela 3.4 e na figura 3.4.

Vídeo	Descrição do Vídeo
VDC 1	Vídeo de monitoração de um saguão com luz solar (CAVIAR)
VDC 2	Vídeo de monitoração de um saguão com luz solar (CAVIAR)
VDC 3	Vídeo de monitoração de um saguão com luz solar (CAVIAR)
VDC 4	Vídeo de monitoração de um saguão com luz solar (CAVIAR)
VDC 5	Vídeo de monitoração do leito de um rio
VDC 6	Vídeo de monitoração de um parque
VDC 7	Vídeo de monitoração de uma rodovia
VDC 8	Vídeo de monitoração de um complexo industrial
VDC 9	Vídeo de monitoração de uma cidade
VDC 10	Vídeo de monitoração de um estacionamento (sombreado)
VDC 11	Vídeo de monitoração de um estacionamento (luz do sol)
VDC 12	Vídeo de monitoração de uma saída de supermercado (pessoas)
VDC 13	Vídeo de monitoração de uma saída de supermercado (veículo)
VDC 14	Vídeo de monitoração de uma saída de supermercado (pessoas e veículo)
VDC 15	Vídeo de monitoração de um condomínio

Tabela 3.4: Descrição dos vídeos que compõem o conjunto VídeoDivCtrl.

Para um último experimento com o método D383F foram escolhidos 3 vídeos de um experimento controlado da base disponibilizada pelo grupo CVPR Lab utilizando a queima de querosene em diferentes distâncias de 10, 20 e 30 metros, no qual este tipo de chama é capturado pela câmera perto do limite de saturação (cor branca), permitindo avaliar o desempenho dos métodos com relação ao tipo de chama sendo capturada e a sensibilidade à distância. Esses vídeos possuem fogo em todos os seus quadros e estão na resolução de 320 x 240 pixels com uma taxa de 24 quadros por segundo com tempo de duração variada, somando um total de 15263 quadros. A tabela 3.5 e a figura 3.5 mostram as descrições e imagens destes vídeos (VDT) que compõem o conjunto VídeoDis.



Figura 3.4: Conjunto VídeoDivCtrl: (VDC1-VDC4) Vídeos disponibilizados pelo grupo CAVIAR; (VDC5-VDC15) Vídeos de diferentes sistemas de vigilância obtidos da internet com objetos capturados na cena que possuem cor de fogo e com movimento.

Vídeo	Descrição do Vídeo
VDT 1	Chama a 10 metros
VDT 2	Chama a 20 metros
VDT 3	Chama a 30 metros

Tabela 3.5: Descrição dos vídeos que compõem o conjunto VídeoDis.



Figura 3.5: Conjunto VídeoDis com chamas geradas a partir da queima de querosene que é capturada pela câmera quase na cor branca (saturação): (VDT1-VDT3) Experimentos realizados a 10, 20 e 30m respectivamente.

3.1.2 Medidas de desempenho

Uma vez detalhada as bases de dados a serem utilizadas ao longo do trabalho, é necessário também fazer uma breve explicação das medidas de desempenho que são utilizadas em conjunto com essas bases de dados ao longo do trabalho.

Primeiro é preciso entender o funcionamento da matriz de confusão de classificação

de duas classes definida na tabela 3.6, que permite a utilização de diversas medidas de desempenho a partir de seus elementos, que são explicados a seguir.

		Classe verdadeira	
		Positivo	Negativo
Classe predita	Positivo	Verdadeiros Positivos (TP)	Falsos Positivos (FP)
	Negativo	Falsos Negativos (FN)	Verdadeiros Negativos (TN)

Tabela 3.6: Matriz de confusão para uma tarefa de classificação com duas classes.

Um verdadeiro positivo ou TP se refere a um item predito pelo classificador como pertencente a classe positiva quando este realmente pertence a classe positiva. Um falso negativo ou FN ocorre quando o item predito como pertencente da classe negativa pertence à classe positiva. De maneira análoga, o mesmo ocorre para o verdadeiro negativo ou TN, em que o classificador prediz a classe do item como negativa de forma correta e o falso positivo ou FP quando o classificador prediz o item como positivo de forma incorreta. A partir desses tipos de classificações é possível utilizar várias medidas de desempenho com o intuito de avaliar os classificadores em diversos aspectos, como veremos adiante.

Para avaliar o desempenho de um classificador é comum o uso na literatura da acurácia, definida como:

$$ACURÁCIA = \frac{TP + TN}{TP + FN + TN + FP} \tag{3.1}$$

No entanto, devido ao contexto desse trabalho de uso do classificador para detectar focos de incêndio em áreas florestais, é de se esperar que a base de dados de imagens geradas seja extremamente desbalanceada com o tempo (já que imagens com fogo serão raramente capturadas quando comparadas à taxa de imagens sem fogo) e, nesse sentido uma consideração deve ser feita sobre o uso da acurácia: a acurácia por si só muitas vezes pode não ser uma medida de desempenho ideal, devido à existência de bases desbalanceadas, pois um classificador avaliado apenas pela acurácia pode ser dito como bom ignorando a classe minoritária e classificando todas as amostras como pertencentes à classe majoritária.

Em (Kubat e Matwin (1997)), os autores citam um exemplo em que a acurácia falha em um experimento com dados de vazamento de óleo, indicando que o classificador está com um bom desempenho. No entanto, este classificador está simplesmente deixando de classificar 80% dos vazamentos de óleo, sendo necessário o uso de outra métrica para auxiliar no diagnóstico de avaliação de desempenho do classificador. Esse exemplo é bem parecido com o contexto de focos de incêndio no sentido que um vazamento de óleo possui uma amostragem muito baixa em relação ao conjunto de amostras completo.

Para realizar uma medição mais correta, é usada neste trabalho a medida F (*F-measure*), em que se deve tomar cuidado para não confundir com distribuição F, já que a medida

F possui o "F" no nome devido a uma confusão acidental com outra função F estudada pelo autor David D. Lewis, conforme Sasaki (2007) explica. A medida F é composta por dois componentes: precisão e revocação ou sensibilidade. Estas medidas são definidas por:

$$PRECISÃO = \frac{TP}{TP + FP} \tag{3.2}$$

$$REVOCAÇÃO \text{ ou } SENSIBILIDADE = \frac{TP}{TP + FN} \tag{3.3}$$

$$MEDIDA\ F = 2 * \frac{PRECISÃO * REVOCAÇÃO}{PRECISÃO + REVOCAÇÃO} \tag{3.4}$$

A precisão e a revocação são medidas usadas em reconhecimento de padrões e mineração de dados (*data mining*) em bases de dados que em sua maioria são desbalanceadas. A precisão representa a fração de instâncias relevantes ou, exemplificando, quantos quadros do vídeo classificados como possuindo fogo realmente possuem fogo. A revocação é a fração relevante dos resultados obtidos que nos diz quantos quadros de fogo totais existentes no vídeo foram classificados corretamente, sendo também chamada de razão de verdadeiros positivos. É fácil perceber pelos exemplos que a precisão é uma medida de exatidão ou qualidade, enquanto a revocação é uma medida de plenitude ou quantidade (Harleam (2012)).

As medidas precisão e revocação são usadas na literatura quando o desempenho da classe minoritária (classe positiva) deve ser levado em conta. No caso do trabalho aqui proposto a classe positiva fogo sempre será minoritária em uma situação de funcionamento real, sendo necessária a avaliação da qualidade e quantidade no aspecto da classificação, já que é importante saber quantos dos verdadeiros positivos deixaram de ser classificados como tal (falsos negativos) e quantos resultados negativos foram classificados como positivos (falsos positivos). Logo, em uma base de dados ampla, formada ao longo do tempo, a avaliação dos resultados obtidos pelo classificador poderá ser falha se apenas a acurácia for usada, na qual um número muito grande de verdadeiros negativos estará presente na base.

A medida F é definida como a média harmônica da precisão e revocação, no qual um valor alto de F indica valores altos para ambas a precisão e revocação (Phung et al. (2009); Sasaki (2007)). Esta medida, portanto, permite a comparação dos diferentes classificadores de acordo com sua qualidade e quantidade de informações obtidas a partir de uma base de dados para a classe minoritária. Esta medida, embora muito usada em áreas de mineração de dados, é pouco utilizada em reconhecimento de padrões e outras áreas do conhecimento devido à dificuldade de avaliar a importância específica de cada classe a ser classificada. Neste caso em particular, não há dúvida que a classe de imagens com fogo (verdadeiros positivos) tem a maior importância, já que o classificador deve sumariamente acionar um alarme de incêndio quando necessário, sendo os acertos da classe negativa (verdadeiros negativos) de menor importância.

Temos ainda outra medida de desempenho quantitativa com relação à classe nega-

tiva chamada especificidade. A especificidade ou razão de verdadeiros negativos é definida como:

$$ESPECIFICIDADE = \frac{TN}{TN + FP} \tag{3.5}$$

Essa medida é usada em conjunto com a revocação na técnica de curva ROC mostrada posteriormente no trabalho, mostrando a razão de falsos positivos ou alarmes falsos de um classificador segundo:

$$RAZÃO\ DE\ ALARMES\ FALSOS = 1 - ESPECIFICIDADE \tag{3.6}$$

Uma vez detalhadas as bases de dados e as medidas de desempenho a serem usadas ao longo do trabalho, pode-se então definir as etapas correspondentes ao método D383F, e posteriormente, analisar os resultados obtidos para as diferentes bases de vídeos.

3.2 Segmentação dos objetos em movimento utilizando subtração de fundo

A primeira etapa a ser definida para o método D383F é a segmentação ou classificação dos pixels em movimento.

O objetivo geral do trabalho é a monitoração de uma área florestal a partir de um dispositivo de captura colocado em uma torre de transmissão de alta tensão da rede elétrica, portanto o ponto de vista da câmera é fixo. Neste contexto, a imagem capturada é relativamente estática, em que movimentos esporádicos e de magnitude significativa deverão ser capturados como objetos de interesse. Alguns problemas como a vibração por causa de ventos e mudanças constantes de iluminação devem ser levadas em conta nesta configuração proposta. Ruas (2009) cita em seu trabalho estes problemas ao monitorar o tráfego de uma rodovia a partir de uma câmera posicionada em um suporte (poste) ao lado da mesma.

É necessário então que o algoritmo escolhido para a segmentação seja capaz de lidar com estas pequenas vibrações e principalmente com mudanças de luminosidade características de longos períodos de monitoração. Além disso, o algoritmo deve ser rápido o suficiente para garantir a monitoração da área florestal em tempo real em diferentes resoluções.

Segundo Ferreira (2007), o uso da subtração de fundo (*background subtraction*) é bem comum para a segmentação em situações que a imagem é relativamente estática. Estes algoritmos consistem basicamente em selecionar ou estimar um quadro que será usado como fundo da imagem, o qual não deve conter objetos em movimento. Em seguida, é calculada a diferença pixel a pixel entre o quadro atual e a imagem de fundo. Caso o valor desta diferença seja menor que um limiar definido, este pixel pertence à imagem de fundo, caso contrário, pertence a um objeto detectado.

Sonaglio e Moecke (2010) falam sobre o comportamento dos algoritmos de subtração de fundo de um modo geral, sendo desejável que a imagem de fundo deva ser constantemente atualizada para compensar as variações de iluminação e alterações na cena produzidas pelo movimento de objetos na mesma, dando maior robustez ao sistema de vigilância.

Em Ko et al. (2010) e Wang et al. (2010), os autores afirmam que o uso de algoritmos de subtração de fundo é essencial para melhorar o desempenho dos métodos de detecção de fogo e reduzir o tempo de detecção dos algoritmos, já que a técnica possui um desempenho robusto e atende à necessidade de aplicações em tempo real.

Na literatura, diversos tipos de algoritmos de subtração de fundo são usados em conjunto com os métodos de detecção de fogo, nos quais cada autor escolhe o algoritmo que melhor atenda os seus objetivos. Para o método proposto neste trabalho, um modelo simples de segmentação será adotado, no qual objetos que não sejam fogo, mas que possuem movimento deverão ser tratados pelas etapas posteriores, ficando a cargo da subtração de fundo apenas definir o que está se movendo sem grandes restrições, eliminando grande parte dos pixels que se mantêm estáticos na cena. Toreyin et al. (2006) também adota a mesma estratégia em seu trabalho, em que o autor prioriza o tempo de processamento às custas da qualidade de segmentação, problema esse tratado também por sistemas de rastreamento de objetos em que a qualidade tem maior peso (Ruas (2009)).

Piccardi (2004) relata sobre os problemas citados anteriormente de alterações de luminosidade na cena e movimentos causados pela oscilação da câmera devido a ventos ou outros eventos em sua revisão dos algoritmos de subtração de fundo, e coloca o algoritmo de subtração de fundo estimado pelas médias adaptativas (*running average*) em destaque como uma técnica de alta velocidade e baixo consumo de memória, superando todas as outras nestes quesitos, assim como Rauta e Fernandes (2012) relatam da simplicidade e robustez da mesma para problemas de alterações de iluminação e alterações no cenário. Com base nestes estudos, esta técnica apresenta características ideais de simplicidade, robustez e velocidade, sendo escolhida para ser usada na etapa de segmentação por movimento do método D383F. Este algoritmo foi proposto originalmente por Donohoe et al. (1988) para resolver o problema dos métodos que usavam uma imagem de fundo estática e não eram capazes de lidar satisfatoriamente com alterações constantes no ambiente da cena, sendo que o algoritmo é dividido em duas etapas segundo Rauta e Fernandes (2012): a adaptação da imagem de fundo e o cálculo do alvo através do limiar.

A etapa de adaptação da imagem de fundo é realizada através do uso de um filtro passa-baixas temporal auto-regressivo e de média móvel ou $\text{ARMA}_{(1,1)}$ (*Autoregression and Moving Average*), que segundo Donohoe et al. (1988) estima recursivamente a média da imagem de fundo e conseqüentemente elimina os objetos em movimento da mesma.

Segundo Institute for Statistics Education (2013), os filtros ARMA são usados na análise de séries temporais para descrever séries temporais estacionárias, ou seja, séries que

não alteram suas propriedades estatísticas (média e variância) ao longo do tempo. A expressão geral de um filtro ARMA *y(t)* é:

$$y(t) = \underbrace{\sum_{i=0}^{n} b(i) \times x(t-i)}_{MA} + \underbrace{\sum_{i=1}^{m} a(i) \times y(t-i)}_{AR} \tag{3.7}$$

onde *a(i)* é o coeficiente da parte AR do modelo, *b(i)* é o coeficiente da parte MA e *x(t)* os elementos de entrada ou ruído branco a serem passados pelos filtros.

Sabendo que o etapa de adaptação de fundo se baseia no modelo ARMA com ordem 1 para ambas as partes AR e MA, o cálculo da imagem de fundo usando todos os canais do espaço de cores RGB é dado por (Ferreira (2007); Rauta e Fernandes (2012)):

$$\forall (R,G,B)\ bg_n(x,y) = \begin{cases} img_{n-1}(x,y), & \text{se } n = 2; \\ \underbrace{\alpha \times img_{n-1}(x,y)}_{MA} + \underbrace{(1-\alpha) \times bg_{n-1}(x,y)}_{AR}, & \text{se } n > 2 \end{cases} \tag{3.8}$$

onde $bg_{n-1}(x,y)$ e $img_{n-1}(x,y)$ se referem, respectivamente, ao valor de intensidade do pixel na posição (x,y) da imagem de fundo calculada anteriormente e do quadro anterior da sequência de vídeo. Com $\alpha \in (0,1)$, assumindo um valor constante que determina o tempo necessário para que a influência de quadros antigos diminua na imagem de fundo (Bradski (2008)). No caso, os coeficientes do filtro (equação 3.7) são: *a(1)*= $1-\alpha$; *b(0)*=0, em que a parte referente ao primeiro termo do somatório da parte MA fica excluída da equação 3.8 e *b(1)*=α. A figura 3.6 mostra o esquema de funcionamento desse filtro.

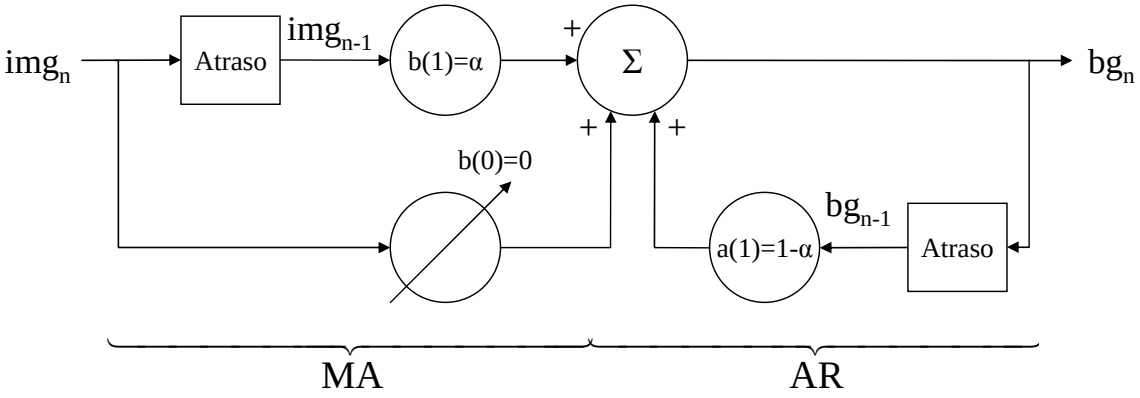


Figura 3.6: Esquema do filtro $ARMA_{(1,1)}$ para a adaptação da imagem de fundo, o parâmetro α irá controlar a taxa de atualização da imagem de fundo, podendo variar a importância de cada parte de modelo, atualizando a imagem de fundo lentamente quando próximo a 0 (ignorando a parte MA) e atualizando rapidamente com valores próximos a 1 (ignorando a parte AR do modelo).

Valores baixos ou muito próximos a 0 de α levam a uma atualização lenta da imagem de fundo, ignorando a parte MA do modelo e sendo incapaz de acompanhar as mudanças de cenário. Valores altos de α ou muito próximos a 1 levam a uma atualização rápida da imagem de fundo ignorando a parte AR, resultando em uma incorporação rápida das mudanças de cenário e uma segmentação imprecisa dos objetos. Esse problema é citado por Rauta e Fernandes (2012), em que os autores apontam a dificuldade na escolha do parâmetro α. Os autores Donohoe et al. (1988) mencionam que o valor de α deve ser escolhido experimentalmente com um tempo de atualização de muitos segundos ou até mesmo minutos dependendo do objeto de interesse a ser capturado, o que permite, portanto, que o algoritmo seja ajustado para diversas situações dependendo do contexto.

A segunda etapa, referente ao cálculo do alvo através do limiar, irá identificar um objeto que não pertence ao fundo através do uso de um parâmetro de limiar T_s. Para isso, realiza-se a subtração pixel a pixel do quadro atual da sequência de vídeo img_n da imagem de fundo bg_n, calculada pela equação 3.8:

$$sub_n(x, y) = |img_n(x, y) - bg_n(x, y)| \tag{3.9}$$

onde o resultado da operação contido em $sub(x, y)$ é convertido para a escala de cinza segundo a equação 3.10:

$$cinza(x, y) = 0.299 \times R(x, y) + 0.587 \times G(x, y) + 0.114 \times B(x, y) \tag{3.10}$$

e submetido à seguinte regra para gerar a segmentação dos pixels que representam os objetos em movimento na cena:

$$\forall x, y \in img\ F_0(x, y) = \begin{cases} 1, & \text{se } sub(x, y) > T_s; \\ 0, & \text{caso contrário} \end{cases} \tag{3.11}$$

onde $F_0(x, y)$ para x e y varrendo toda imagem é uma máscara do objeto ou alvo a ser identificado como um conjunto estático de pixels (parte do fundo) ou um conjunto de pixels em movimento, com $0 \leq T_s \leq 255$.

O parâmetro T_s permite o controle da informação de quantidade de movimento de acordo com valor de intensidade, podendo ser alterado de forma a evitar que ruídos e movimentos de baixa magnitude provenientes da captura gerem *clutter* (objetos que não são de interesse) para o restante do método avaliar, obtendo uma segmentação mais precisa.

Este parâmetro depende exclusivamente dos objetos sendo capturados e da resolução de captura da cena, já que uma resolução maior irá dar mais detalhamento às bordas de galhos, folhas e outros objetos que possam interferir na classificação correta, sendo necessário especificá-lo experimentalmente assim como o parâmetro α, conforme Rauta e Fernandes (2012) mostram em seu trabalho.

Para o correto funcionamento dessa técnica de subtração básica de fundo, os parâmetros

α e T_s precisam trabalhar de forma conjunta para que uma boa segmentação de movimento na cena possa ser feita. Após testes do algoritmo em diversos vídeos e os resultados avaliados para selecionar os valores que melhor generalizam as diversas situações encontradas em um ambiente externo, os valores escolhidos para os parâmetros α e T_s foram respectivamente α =0.0075 e $15 < T_s < 36$, sendo os valores de T_s proporcionais em magnitude a resolução da imagem sendo processada, que podem ser vistos no capítulo 4 na tabela 4.1.

Esse valor bem baixo de α torna a construção da máscara bastante sensível a movimentos, uma característica necessária para capturar o maior número de pixels possível do fogo para as etapas subseqüentes do método D383F. Com este valor, objetos de movimento rápido como pessoas e pequenas vibrações não são incorporadas de imediato à imagem de fundo. Caso o movimento seja constante dentro de uma área, esta área se torna "embaçada" ao longo do tempo na imagem de fundo. Logo, galhos, folhas e artefatos obtidos pela vibração da câmera com movimentos constantes acabam por causar esse efeito e podem ser eliminadas posteriormente pelo limiar da subtração de fundo T_s, uma vez que essa área embaçada gera uma atenuação dos valores de intensidade obtidos na etapa de subtração de fundo. No entanto, esse efeito é pequeno se a frequência de movimento na área for baixa. Nesse caso, o algoritmo continuará captando o movimento e incluindo estes clutters que serão tratados posteriormente no método.

Com relação ao fogo, a adaptação da imagem de fundo lenta acaba por ajudar na captura das bordas com seus movimentos aleatórios de expansão e contração, o que dificulta a inclusão destas áreas na imagem de fundo. Esta é uma área importante que define o movimento aleatório do fogo quando este se estabiliza ou é pequeno e não possui grandes diferenças de cores em suas chamas.

Observou-se neste estudo que valores menores que o estipulado de α não obtém resultados melhores, uma vez que pequenas alterações de luminosidade e ruídos passam a ser captados constantemente devido à taxa de atualização de fundo muito lenta, gerando neste caso falsos positivos em objetos estáticos com cor de fogo. Quanto às mudanças de iluminação na cena, esse valor de α ainda é capaz de assimilar em poucos segundos estas alterações, podendo esse tempo ser facilmente ajustado posteriormente para uma atualização mais rápida da imagem de fundo ao aumentar o valor aqui estipulado.

Definidas as duas etapas do algoritmo, pode-se então mostrar através de exemplos o funcionamento dos parâmetros α e T_s. A figura 3.7 mostra um exemplo do funcionamento de α ao longo da duração do vídeo VD 1 (1200 quadros), no qual cada linha representa a segmentação com diferentes valores de α, contendo as imagens originais, imagens de fundo e resultado da segmentação respectivamente obtidas em dois pontos distintos do vídeo VD1 (quadro 400 e 1200) com o valor de limiar $T_s = 40$.

No caso do vídeo VD 1, valores mais baixos de α que o estipulado levam à aquisição de falsos positivos nos objetos estáticos com cor de fogo devido à atualização lenta da imagem de fundo, como a parede e a grama na parte atrás da grade que foram capturados

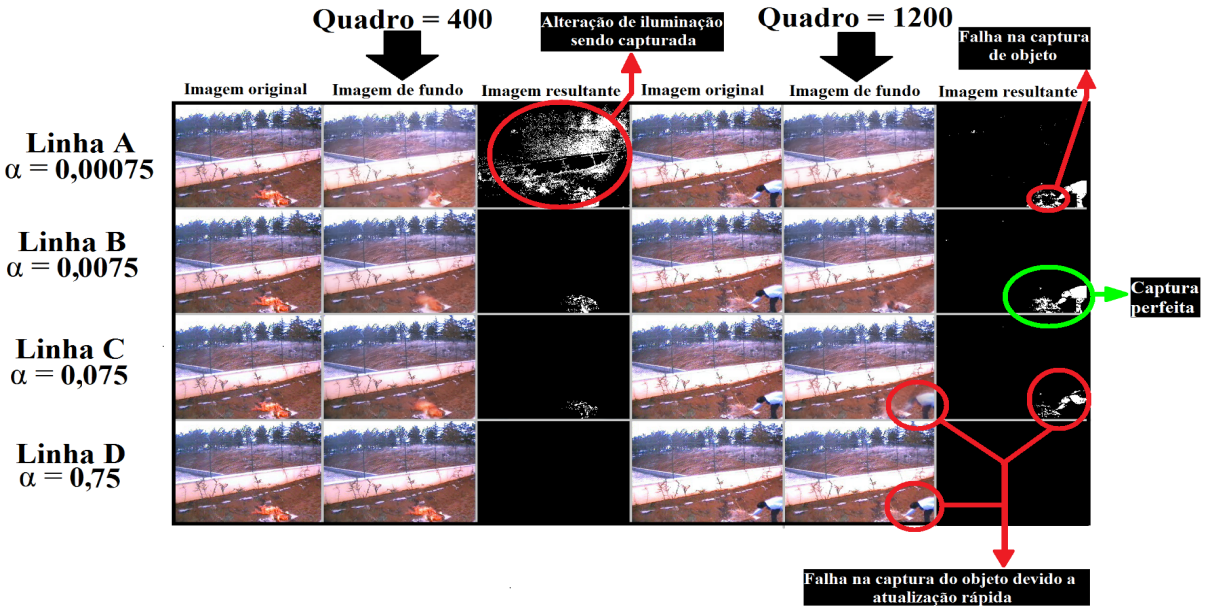


Figura 3.7: Resultados de segmentação obtidos para $T_s = 40$ e diferentes valores de α. Para cada linha, temos diferentes valores de α, com a imagem original, imagem de fundo e resultado da segmentação respectivamente em dois pontos distintos do vídeo VD1, quadro 400 e 1200. (Linha A) Resultados obtidos com α=0.00075: É possível ver problemas na segmentação nos dois pontos distintos do vídeo causados pela atualização lenta da imagem de fundo. No quadro 400, temos uma alteração de iluminação no vídeo ocorrida já à algum no tempo no vídeo sendo capturada, segmentando grande parte do objetos estáticos na cena. No quadro 1200, o combustível sendo colocado sobre a chama é confundido com chama, gerando uma segmentação incompleta do objeto; (Linha B) Resultados obtidos com α=0.0075: Vemos que o problema da alteração de iluminação foi resolvido, em que a imagem de fundo incorpora a mudança em poucos segundos, segmentado apenas os pixels referentes a chama do vídeo. No quadro 1200, com esse valor vemos uma segmentação quase perfeita da pessoa e do combustível; (Linha C e D) Resultados obtidos com α=0.075 e 0.75 respectivamente: Valores mais altos de α levam a uma degradação na segmentação devido a uma atualização muito rápida da imagem de fundo. Neste caso, em ambas as linhas na imagem de fundo do quadro 1200 vemos que a pessoa é rapidamente incorporada a imagem de fundo, gerando uma segmentação imprecisa.

no quadro 400 após uma alteração de iluminação acontecer durante o vídeo. E no quadro 1200, o combustível não é segmentado corretamente devido ao mesmo problema e por possuir cores semelhantes ao fogo. Com valores de α superiores a 0.0075, há uma degradação na segmentação dos objetos devido a uma atualização muito rápida da imagem de fundo, que pode ser facilmente observada no quadro 1200, em que a pessoa é rapidamente assimilada pela imagem de fundo e gera uma segmentação imprecisa da mesma.

A figura 3.8 mostra os resultados obtidos para diferentes valores de T_s com $\alpha = 0.0075$ no quadro de número 400 do vídeo VD 1, exemplificando a dificuldade de definir o parâmetro sem levar em conta os objetos capturados na cena.

Outra característica importante a ser discutida é que geralmente os métodos de subtração

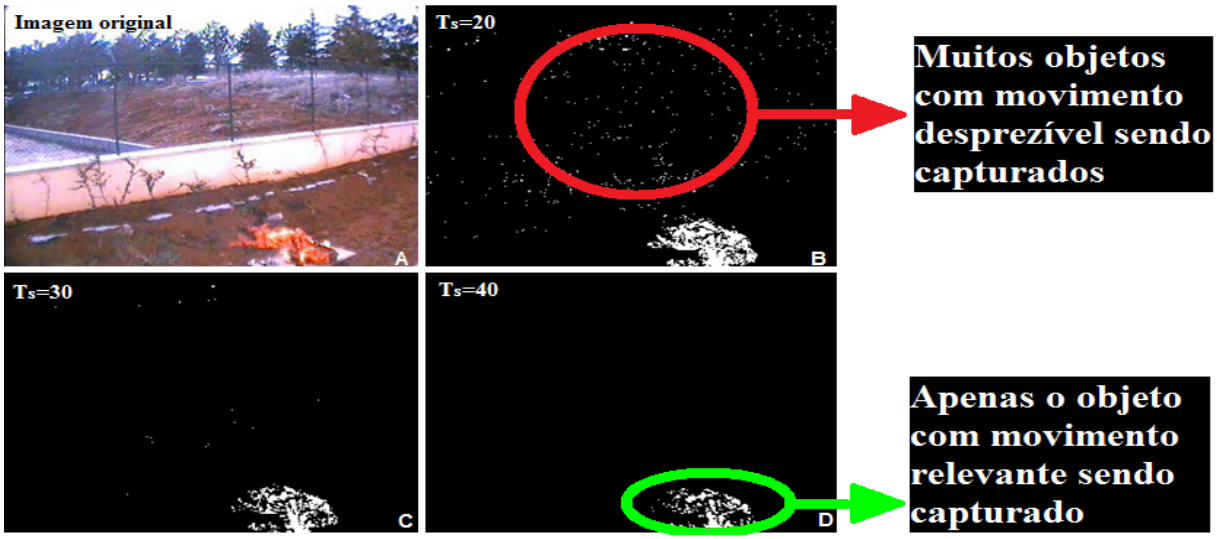


Figura 3.8: Segmentação dos objetos em movimento. (A) Imagem original; (B) Segmentação com $T_s = 20$; (C) Segmentação com $T_s = 30$; (D) Segmentação com $T_s = 40$; Percebe-se que valores baixos de T_s objetos com movimento desprezíveis são segmentados, mas com valores adequados apenas objetos com movimento relevante são capturados.

de fundo são usados na literatura com imagens em tons de cinza ao invés de imagens no espaço de cores RGB para ter um custo computacional mais baixo. Uma técnica para realizar a transformação de uma imagem RGB para uma imagem em tons de cinza é dada pela média ponderada para cada canal de cor, calculada segundo a equação 3.10.

No método aqui proposto as etapas referentes às equações 3.8 e 3.9 são feitas em cores, ou seja, 3 dimensões, isto se deve ao fato que o fogo está intrinsecamente ligado às suas cores, geralmente possuindo uma cor laranja-avermelhada em suas bordas. Como a escala de cinza é uma média ponderada é de se esperar que vários conjuntos de diferentes valores de intensidade para os diferentes canais levem a intensidades de cinza iguais. Um exemplo seria a própria fumaça criada pelo fogo, em que o tom de cinza pode facilmente assumir valores do fogo na escala de cinza e gerar problemas para a segmentação do fogo. Portanto, ao transformar um objeto para a escala de cinza informações são perdidas, neste caso, a média absorve esta importante informação do fogo contida no esquema de cores RGB.

No método D383F, a transformação para a escala de cinza só é feita para o último passo (equação 3.11), no qual a imagem colorida $sub(x, y)$ resultante da subtração de fundo (equação 3.9) é então transformada em escala de cinza (equação 3.10) para que a segmentação por movimento seja efetuada. Não foi encontrado na literatura nenhum estudo sobre as limitações impostas pela subtração de fundo em escala de cinza na detecção de objetos, e no caso das chamas foi mostrado que a utilização das cores é importante para a segmentação correta das mesmas na cena, sendo esta uma importante contribuição para futuros trabalhos na detecção de chamas em vídeo. A figura 3.9 ilustra o problema encontrado no vídeo VR1, em que o fogo transformado para escala de cinza tem praticamente a mesma intensidade da fumaça, sendo difícil a segmentação do mesmo.

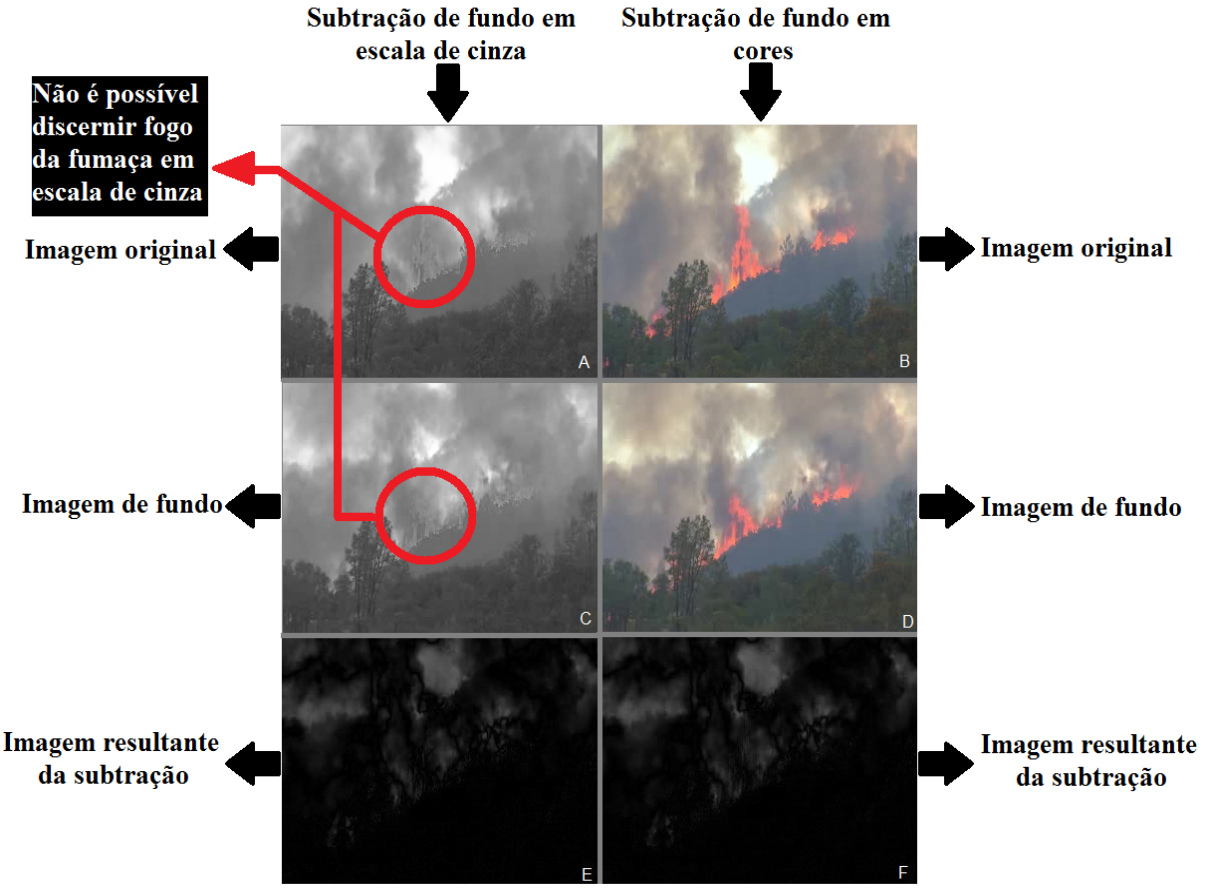


Figura 3.9: Subtração de fundo usando escala de cinza e o esquema cores RGB. (A) Imagem original do quadro 20 do vídeo VR1 convertida para escala de cinza, nota-se que é extremamente difícil ver as chamas nessa imagem; (B) Imagem original à cores do mesmo quadro mostrado em A; (C) Imagem de fundo em escala de cinza apresentando o mesmo problema da A; (D) Imagem de fundo colorida; (E) Imagem resultante da subtração em escala de cinza entre A e C; (F) Imagem resultante da subtração em cores entre B e D; Existe uma pequena diferença entre E e F exatamente na área que representa o fogo.

Para mostrar com mais clareza a perda de informação na subtração de fundo em escala de cinza em relação a subtração de fundo em cores, foi feita uma operação de subtração entre as figuras 3.9E e 3.9F. Da imagem resultante, foi gerada uma máscara binária à partir dos pixels com valores não nulos, no qual se observa que a informação faltante corresponde exatamente à área das chamas, conforme ilustra a figura 3.10.

O resultado obtido da segmentação por movimento irá definir quais pixels devem ser sujeitados ao classificador de pixels por cor, sendo esta a próxima etapa detalhada.



Figura 3.10: Subtração entre as figuras 3.9E (escala de cinza) e 3.9F (cores). (A) Máscara binária da subtração; (B) Quadro original do vídeo em cores. Percebe-se que a informação faltante corresponde exatamente a área das chamas.

3.3 Classificação dos pixels pela cor

É comum entre os trabalhos realizados na literatura para a classificação de fogo o uso do espaço de cores RGB (Chen et al. (2004); Çelik et al. (2007); Ko et al. (2010); Toreyin et al. (2006); Wang et al. (2010); Wen et al. (2010); Zhou et al. (2010)), já que os dispositivos de captura atuais na sua maioria retornam a imagem em função desse espaço de cores, sendo vantajoso a nível computacional manter o espaço de cores original da câmera. No entanto, o espaço RGB possui em todas as suas 3 componentes as informações de luminosidade e cromaticidade agregadas segundo Çelik e Demirel (2009) e Ruas (2009) explicam, sendo esta uma limitação para algumas aplicações.

Çelik e Demirel (2009) expõem de maneira clara as limitações geradas ao se utilizar o espaço RGB puro para a classificação de pixels cor de fogo, em que cada canal de cor do RGB é uma medida de intensidade que leva em conta a luminosidade da cena. Segundo estes autores uma mudança na iluminação da imagem capturada pode fazer com que a classificação baseada em regras no espaço RGB venha falhar, o que o leva a afirmar que a crominância é muito mais robusta para modelar a cor do fogo do que a intensidade. Para esta tarefa de separar a informação de luminância da crominância, ele propõe o uso do espaço de cor YCbCr, sendo Y a luminância, Cb a crominância azul e Cr a crominância vermelha. O espaço YCbCr é uma conversão linear do espaço RGB, logo a conversão tem baixo custo computacional, sendo uma característica importante a ser explorada.

Outro ponto extremamente importante foi que Çelik e Demirel (2009) obtiveram em seus testes de comparação com métodos baseados no espaço de cor RGB bons resultados com relação à taxa de falsos positivos capturados usando apenas regras baseadas no espaço de cor, sem o uso de outros algoritmos para auxiliar na tarefa de classificação dos pixels.

Levando em conta os pontos positivos citados acima, o método proposto por Çelik e Demirel (2009) foi adotado como método base para o sistema proposto de forma a executar a classificação por cores dos pixels.

Como primeiro passo do método é necessário transformar o espaço RGB para o espaço YCbCr segundo a equação 2.2 na seção 2.2.1. Também, na mesma seção, a transformação do espaço de cor RGB para os diferentes canais do espaço de cor YCbCr pode ser vista na figura 2.8.

Seguindo as regras propostas por Chen et al. (2004) e Çelik et al. (2007) para o espaço RGB, no qual a área correspondente de uma chama geralmente obedece a regra $R > G > B$ e $R > R_{medio}$, os autores formulam novas regras para o espaço de cor YCbCr:

$$F_1(x,y) = \begin{cases} 1, & \text{se } \mathrm{Y}(x,y) > Cb(x,y) \\ 0, & \text{caso contrário} \end{cases} \tag{3.12}$$

$$F_2(x,y) = \begin{cases} 1, & \text{se } Cr(x,y) > Cb(x,y) \\ 0, & \text{caso contrário} \end{cases} \tag{3.13}$$

Estas duas regras implicam que a luminância da chama deve ser maior que crominância do azul e a crominância do vermelho maior que a crominância do azul respectivamente, uma vez que em uma chama há a saturação do componente vermelho no espaço de cor RGB. Estas regras foram comprovadas por eles ao avaliarem as médias dos pixels referentes às regiões com fogo das imagens de sua base de dados.

Para a próxima regra, os autores levam em consideração que o fogo é geralmente a área com maior brilho da imagem, o que implica que os valores de cada pixel nos canais Y e Cr referente à área da chama terão valores superiores à média dos mesmos canais para toda a imagem, enquanto os valores do canal Cb nos pixels que compõem a chama terão valor menor que a média de toda a imagem:

$$Y_{medio} = \frac{1}{img_w \cdot img_h} \sum_{x=1}^{img_w} \sum_{y=1}^{img_h} \mathrm{Y}(x,y);$$

$$Cb_{medio} = \frac{1}{img_w \cdot img_h} \sum_{x=1}^{img_w} \sum_{y=1}^{img_h} Cb(x,y);$$

$$Cr_{medio} = \frac{1}{img_w \cdot img_h} \sum_{x=1}^{img_w} \sum_{y=1}^{img_h} Cr(x,y);$$

$$F_3(x,y) = \begin{cases} 1, & \text{se } Y(x,y) > Y_{medio} \text{ e } Cb(x,y) < Cb_{medio} \text{ e } Cr(x,y) > Cr_{medio} \\ 0, & \text{caso contrário} \end{cases} \tag{3.14}$$

onde img_w e img_h representam a largura e altura da imagem respectivamente.

Outra regra foi gerada pelos autores ao observar as diferenças entres os canais Cb e Cr quando uma chama está presente, no qual a chama no canal Cb é predominantemente preta e no canal Cr predominantemente branca, sendo possível visualizar esta característica de intensidade na figura 2.8. A regra gerada baseada nesta observação é:

$$F_4(x,y) = \begin{cases} 1, & \text{se } |Cb(x,y) - Cr(x,y)| > \tau \\ 0, & \text{caso contrário} \end{cases} \tag{3.15}$$

onde τ é uma constante definida usando a técnica de análise chamada Curva de Características de Operação do Receptor (*ROC Curve - Receiver Operating Characteristic Curve*) ou curva ROC.

A técnica da curva ROC consiste em demonstrar a relação antagônica entre a revocação (equação 3.3) ou razão de verdadeiros positivos e a razão de alarmes falsos (equação 3.6). A curva ROC foi desenvolvida no campo das comunicações para demonstrar as relações entre sinal e ruído nos radares durante a segunda guerra mundial (Margotto (2010)). Na literatura, esta técnica é usada constantemente para avaliar o desempenho dos classificadores e selecionar soluções que permitam um melhor aproveitamento da relação de acertos na classe positiva e alarmes falsos dos classificadores (Fawcett (2006)).

Ao se usar essas duas medidas como eixos de um gráfico, é possível variar um parâmetro e avaliar o mesmo com relação aos resultados de classificação, ou seja, é possível avaliar no gráfico qual valor do parâmetro será melhor para o dilema entre a taxa de acertos ou detecção e a taxa de alarmes falsos (Fawcett (2006); Metz (1978); Phung et al. (2009)).

Para realizar este teste, foi utilizada a base de imagens estáticas ImaSta (vide seção 3.1). Usando as 4 regras enunciadas anteriormente (F_1 à F_4) para classificar os pixels das imagens do conjunto descrito acima, a curva ROC é gerada levando em conta se houve ou não pixels classificados como fogo nas imagens, variando o valor de τ de 10 a 100 em intervalos de 10 unidades. Relembrando, um verdadeiro positivo ou acerto positivo é definido quando uma imagem possui fogo e possui pixels classificando-a como tal e um falso positivo ou alarme falso é definido quando uma imagem não possui fogo, mas contém pixels classificando-a como parte do grupo de imagens de fogo. A Curva ROC obtida para a base de imagens estáticas é mostrada na figura 3.11.

O valor da Curva ROC a ser selecionado deve garantir uma alta taxa de acertos positivos e uma taxa de alarmes falsos aceitável, sendo o valor 80 um bom candidato que garante uma taxa de acertos positivos de 94% e uma taxa de alarmes falsos de 35%, próximo ao obtido por Çelik e Demirel (2009) em seus testes com o valor 40 (>90% de acertos positivos e <40% de alarmes falsos). Essa diferença de valores do parâmetro τ se deve a base de dados usada (ImaSta) que possui imagens de controle bastante desafiadoras para o método. Os autores deixam claro que outros valores podem ser escolhidos de acordo com o necessário, opção que será explorada futuramente na seção 4.2 ao variar τ para capturar chamas amarelo claro.

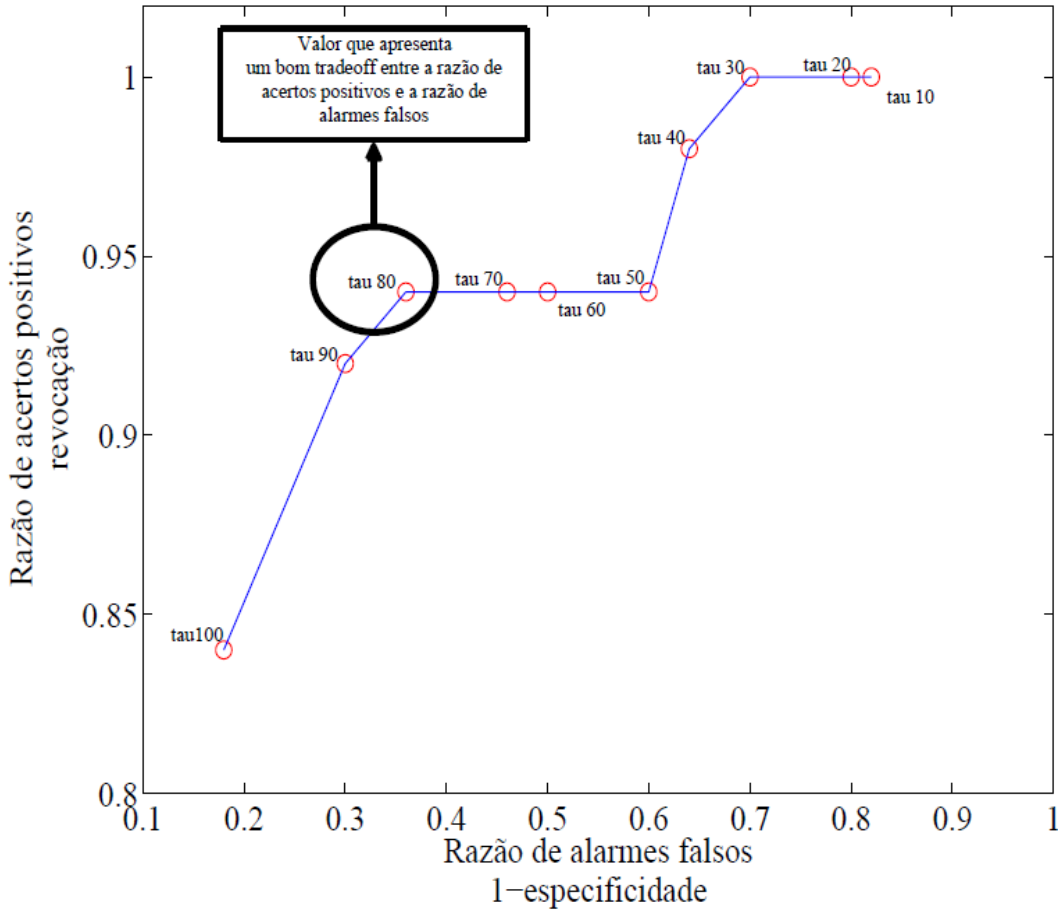


Figura 3.11: Curva ROC para a base de imagens estáticas

Em adição às regras propostas anteriormente, Celik e Demirel ainda fizeram uma análise estatística em cima da informação de crominância, em que imagens de fogo segmentadas manualmente são usadas para gerar um gráfico de distribuição ou histograma dos pixels no plano de crominância Cb-Cr, a partir dessa distribuição, polinômios foram gerados para delimitar a área ocupada por esses pixels de amostra, e um pixel é considerado fogo quando este se encontra dentro da área delimitada por esses polinômios. Infelizmente os polinômios informados no trabalho deles não geram resultados correspondentes à área de delimitação mostrada em seu trabalho, impedindo uma reprodução completa de seu experimento.

Para contornar o problema e manter-se o mais fiel possível ao modelo proposto por eles, foi utilizada uma estratégia bem similar ao usado por Toreyin et al. (2006) para a classificação de pixels por cores, onde ele usa um modelo de mistura de gaussianas para gerar esferas de contenção das cores do fogo no espaço RGB a partir de imagens de amostras. No caso, os dados obtidos do histograma são passados pelo algoritmo de agrupamento k-médias e a partir dos grupos formados, áreas circulares são geradas, delimitando a área do plano Cb-Cr que contém cores classificadas como cor de fogo.

O algoritmo de K-médias é amplamente utilizado na literatura por ser um dos algoritmos de aprendizado sem supervisão mais simples que resolve problemas de agru-

pamento, sendo usado para os mais diversos tipos de aplicações. O algoritmo calcula iterativamente um número de centróides (ponto médio do grupo) pré-estabelecido e verifica a distância euclidiana dos diversos pontos no espaço em relação aos centróides na formação dos grupos, procurando sempre minimizar a soma das distâncias euclidianas intra-grupo. A formulação é dada do seguinte modo: dado um conjunto de pixels ($\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_N$) onde cada pixel é um vetor com 2 dimensões que representam suas coordenadas no plano Cb-Cr, o algoritmo de k-médias objetiva particionar os N pixels em K conjuntos $\mathbf{S} : \mathbf{S}_1, \mathbf{S}_2, ..., \mathbf{S}_K$ (onde K≤ N) de forma a minimizar a soma da distância euclidiana segundo:

$$\arg\min \sum_{i=1}^{K} \sum_{\mathbf{p}_j \in \mathbf{S}_i} \|\mathbf{p}_j - \mathbf{c}_i\|^2 \tag{3.16}$$

onde $\mathbf{c}_i$ é um centróide (ponto médio do grupo) em $\mathbf{S}_i$ e $\|\mathbf{p}_j - \mathbf{c}_i\|^2$ a distância euclidiana entre o pixel e o centróide.

O histograma das amostras de pixels no plano Cb-Cr foi gerado a partir das 50 imagens de fogo da base ImaSta, nas quais as regiões de fogo destas imagens foram segmentadas manualmente, assim como feito por Çelik e Demirel (2009) (veja figura 3.12A). Os dados desse histograma foram então passados para o algoritmo de k-médias realizar o agrupamento.

Antes de qualquer coisa é importante salientar que a escolha do número de grupos pré-definido para o algoritmo de k-médias deve ser grande o suficiente para delimitar a área coberta pelas amostras de pixels da forma mais justa possível, evitando áreas não cobertas pelas amostras, mas deve ser pequeno o suficiente para não gerar um elevado custo computacional, sendo o valor de 100 grupos escolhido experimentalmente como um valor que cumpre essas restrições. Após o algoritmo de k-médias escolher os 100 grupos diferentes nesse plano, um raio baseado no valor do desvio padrão (2 vezes esse valor) de cada grupo foi calculado (Toreyin et al. (2006)), gerando 100 círculos a partir dos centróides que delimitam a área referente às cores do fogo, de maneira similar ao modelo de polinômios (veja figura 3.12B). Com esse modelo, o algoritmo de k-médias só é executado uma vez (à priori) com os dados do histograma de amostras de fogo no plano Cb-Cr, sendo obtidos como resultado uma matriz de centróides **C** e um vetor de raios **r**, que são declarados como constantes no algoritmo e usados para classificar os pixels, em que um pixel é considerado como fogo se estiver dentro da área de algum dos 100 círculos gerados pela técnica.

A figura 3.12 mostra a distribuição dos pixels de fogo obtidos das imagens da base ImaSta e como ficou a área delimitada pelos círculos, restringindo o espaço que contém os pixels de fogo.

Com esse modelo alternativo proposto, a regra F_5 de Çelik e Demirel (2009) passa a ser:

$$F_5(x, y) = \begin{cases} 1, & \text{se } \exists\ \mathbf{c}_i\text{: } \|\mathbf{p}\text{-}\mathbf{c}_i\|^2 < \mathbf{r}_i); \\ 0, & \text{caso contrário.} \end{cases} \tag{3.17}$$

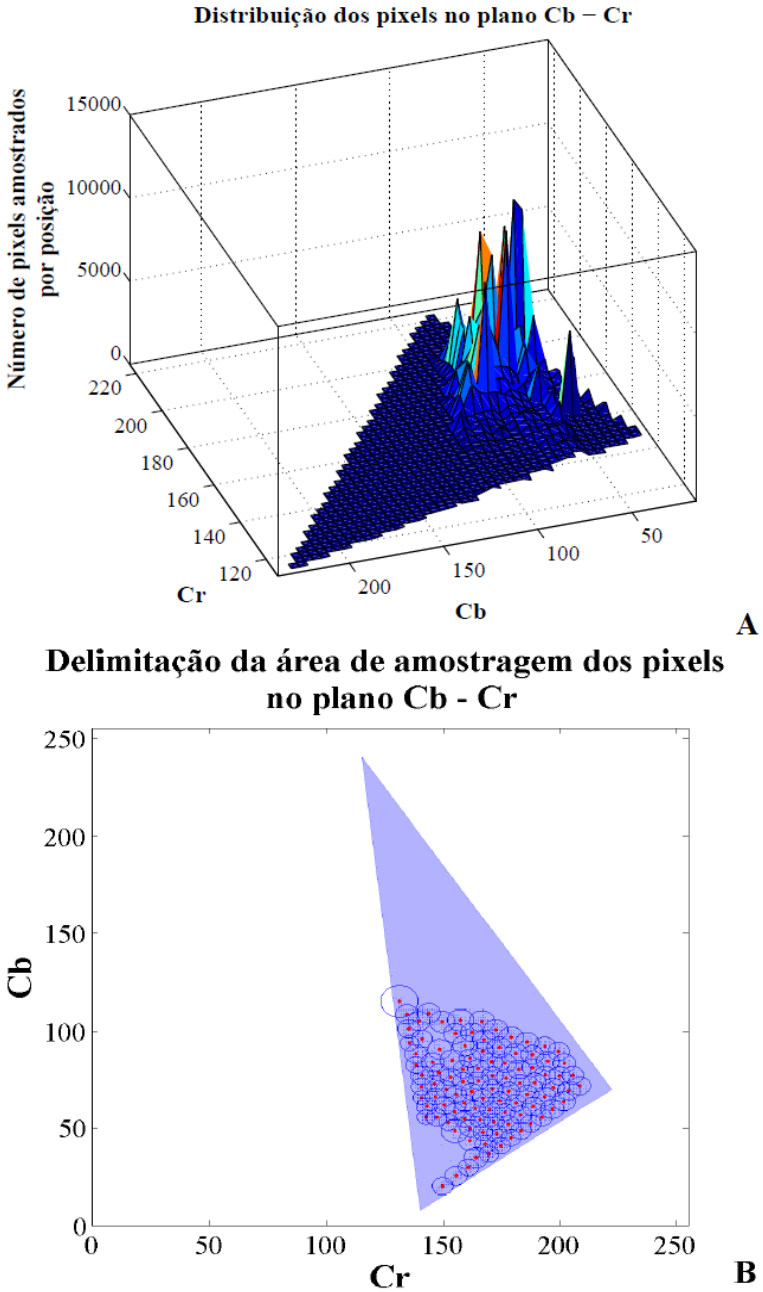


Figura 3.12: Delimitação do espaço de cor para classificação como fogo (A) Histograma no plano Cb-Cr; (B) Círculos delimitando a região de cor considerada como fogo.

onde **p** remete aos valores de intensidade no plano Cb-Cr de um pixel na posição x e y, $\mathbf{c}_i$ representa um dos centróides pertencentes a matriz **C**, $\mathbf{r}_i$ o raio referente a cada centróide e $\|\mathbf{p}\text{-}\mathbf{c}_i\|^2$ a distância euclidiana entre o pixel e o centróide no plano Cb-Cr.

A figura 3.13 mostra a segmentação obtida pelo modelo de polinômios proposta por Çelik e Demirel (2009) e pelo modelo de K-médias proposto utilizando a mesma imagem, evidenciando que o método de K-médias obtém um resultado superior na segmentação dos pixels pertencentes ao fogo.

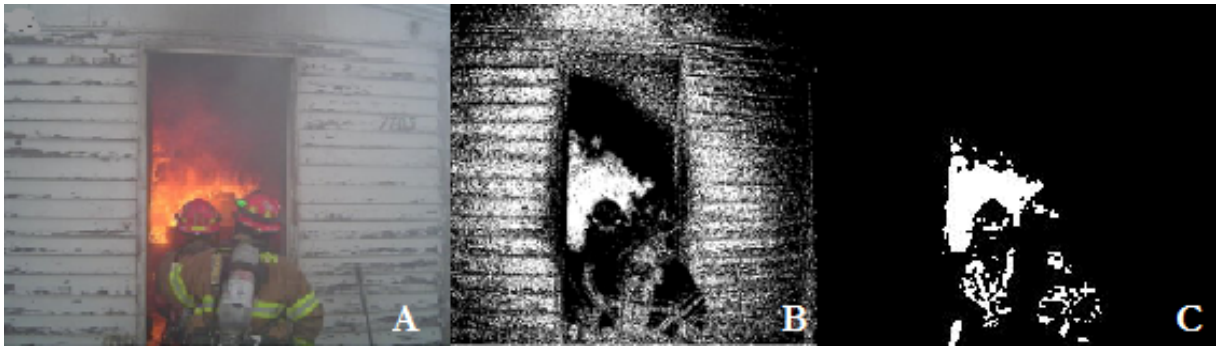


Figura 3.13: Segmentação com k-médias: (A) Imagens original ; (B) Imagem segmentada com polinômios, FONTE: Çelik e Demirel (2009); (C) Imagem segmentada com k-médias.

Finalmente, Çelik e Demirel (2009) completam seu classificador ao aplicar as 5 regras em cada pixel da imagem, classificando como fogo os pixels que satisfizerem todas as regras segundo:

$$F_{Celik}(x,y) = \begin{cases} 1, & \text{se } \sum_{i=1}^{5} F_i(x,y) = 5 \\ 0, & \text{caso contrário.} \end{cases} \tag{3.18}$$

O resultado obtido para cada regra e o resultado final da segmentação podem ser vistos na figura 3.14.

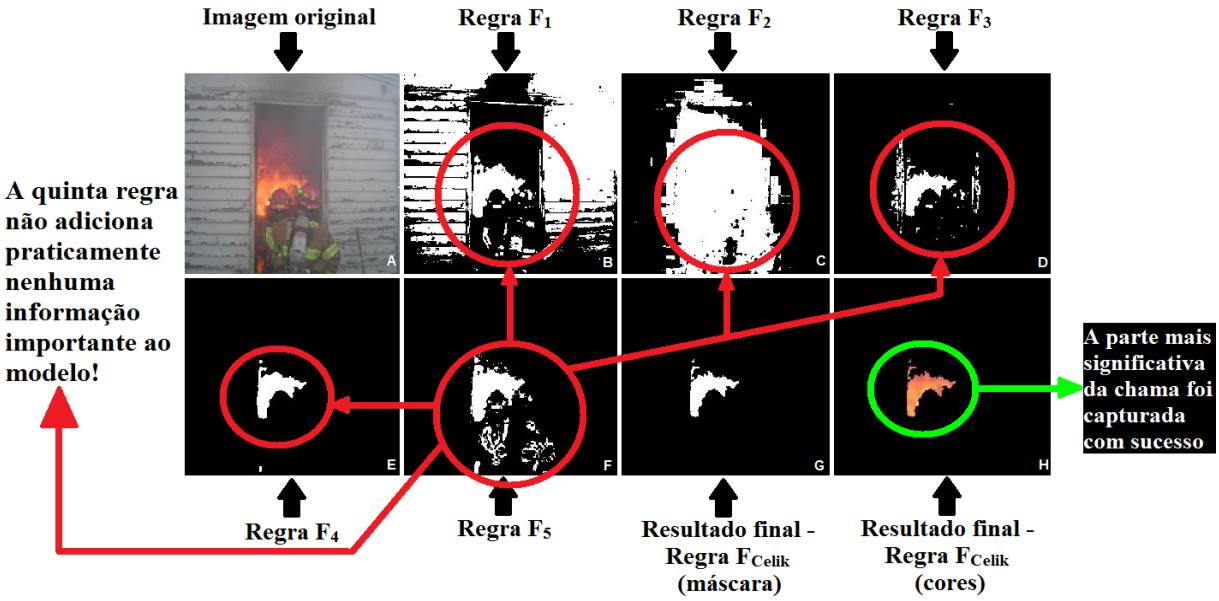


Figura 3.14: Segmentação usando Çelik e Demirel (2009): (A) Imagem original ; (B-F) Resultado da segmentação com cada uma das regras (F_1 a F_5); (G) Imagem segmentada com F_{Celik}; (H) Imagem colorida selecionada pela máscara de classificação F_{Celik}, mostrando que a parte mais significativa da chama na imagem foi capturada. Também é possível observar que a regra F_5 (F) possui informação redundante que está presente em todas as outras regras (B-E), podendo ser eliminada do modelo.

O método modificado a partir do método proposto por Çelik e Demirel (2009) será usado com as alterações descritas para regra F_5 posteriormente no trabalho (experimento da seção 4.1) para efeitos de comparação, em que se demonstram os ganhos obtidos pelas melhorias propostas neste trabalho em relação ao que o autor propõe em seu artigo.

Voltando a figura 3.13, foi mostrado que infelizmente a regra F_5 não adiciona praticamente nenhuma informação ao sistema de classificação. Não é preciso muito esforço para verificar isso, bastando comparar a figura 3.13F com as figuras 3.13B a 3.13E referentes as regras $F_1 - F_4$. A maior parte da informação colhida por F_5 já foi filtrada pela regra F_1 e as demais, sendo que a regra F_4 (correspondente ao parâmetro τ) mais restritiva terá a maior influência para o método. Este problema se repete para diferentes amostras de imagens.

Outro problema a ser levado em conta é o custo computacional exigido pelo modelo com k-médias, que embora não seja elevado, é significativo, uma vez que cada pixel será avaliado até 100 vezes para verificar se o mesmo encontra-se dentro da área de um dos círculos que delimitam a área de fogo no plano Cb-Cr, no qual esse tempo computacional poderia ser usado em outro processo mais eficiente.

Sob esse aspecto, foi decidido alterar esse método, propondo-se uma nova idéia baseada em diferentes trabalhos sobre classificadores mostrados em Guyon et al. (2006), nos quais técnicas de conjunto ou comitê (*ensemble*) de classificadores são usados de forma a melhorar a desempenho da tarefa de classificação ao custo de computações adicionais. Esta idéia de comitê pode ser separada em dois tipos distintos: homogêneo, quando múltiplas instâncias de um mesmo classificador são usadas em diferentes subconjuntos de dados, ou heterogênea quando classificadores diferentes são usados em um único processo, de forma a tirar vantagem de suas diferentes metodologias, modelo este que será utilizado neste trabalho. Este tipo de definição de comitê de classificadores tem muita semelhança com o esquema de regras adotado até o momento, em que a união de todas as regras (ou classificadores no caso do comitê) determina o resultado da classificação.

Baseado nessa metodologia, os seguintes critérios para gerar uma nova regra a partir de um classificador diferente foram propostos:

- C1 - O espaço de cor a ser utilizado deve ser diferente do usado por Çelik e Demirel (2009), sendo ainda capaz de separar em seus canais a luminância da crominância, assim como o YCbCr;
- C2 - Esse espaço de cor escolhido deve permitir uma melhor separação das chamas em seu canal de crominância que possui a informação de cor vermelha com relação ao canal "Cr" do espaço de cores YCbCr, como mostrado na figura 2.8D e enunciado por Çelik e Demirel (2009) em sua regra F_3, em que a chama geralmente apresenta uma intensidade de crominância vermelha mais alta que outros objetos na cena;
- C3 - E por último, a transformação do espaço de cor RGB para este espaço es-

colhido deve ter preferencialmente um custo computacional baixo, uma vez que este trabalho se propõe a usar resoluções mais altas que 320 x 240 pixels para a detecção de focos de incêndios em tempo real.

Dentre os espaços de cores disponíveis temos: o CIE Lab usado por Çelik (2010), que é uma transformação do espaço de cor CIE XYZ; o CIE Luv que também é uma versão transformada do CIE XYZ, sendo que não foi encontrado na literatura o uso deste espaço de cor para a detecção de focos de incêndio; e por último o YUV, usado por Marbach (2006). As transformações referentes a estes espaços de cores podem ser vistas no capítulo 2.

Todos os 3 espaços de cores acima satisfazem o critério C1. Para avaliar o critério C2, foi primeiro definido quais canais de cada espaço era responsável por mostrar as informações de crominância da cor vermelha. Para o CIE Lab o canal "a" é responsável por mostrar essa informação, o canal "u" para o CIE Luv e o canal "V" para o YUV.

Para definir um modelo de avaliação da informação contida em cada canal de crominância vermelha e compará-lo ao canal Cr do espaço de cor YCbCr, foram usadas as 50 imagens de fogo da base de dados ImaSta. Foi calculado para cada uma das imagens o valor de intensidade máxima obtida por cada canal (V_{max}) e a diferença entre este valor e a média ($\bar{V}$) obtida para todos os pixels da imagem no canal (V_c). Esta estratégia permite avaliar o quanto a cor vermelha se destaca do resto da imagem (contraste). Além disso, o tempo médio para transformação do espaço de cores RGB para cada um dos espaços citados acima também foi calculado, já que é necessário que o tempo de processamento seja o menor possível para o novo modelo criado. A média do canal que representa a crominância vermelha do espaço de cores ($\bar{V}$), o valor máximo (V_{max}) e o contraste (V_c) são obtidos respectivamente segundo:

$$\bar{V} = \frac{1}{img_w \cdot img_h} \sum_{x=1}^{img_w} \sum_{y=1}^{img_h} V(x,y); \tag{3.19}$$

$$V_{max} = \max V(x,y) \tag{3.20}$$

$$V_c = V_{max} - \bar{V} \tag{3.21}$$

onde $V(x,y)$ é o valor de intensidade do pixel (x,y) no canal de crominância vermelha, img_w a largura da imagem e img_h a altura da imagem.

A tabela 3.7 mostra a média dos resultados obtidos para o teste de quantificação do contraste da cor vermelha e tempo gasto usando as imagens de fogo da base de dados ImaSta na quarta e quinta coluna respectivamente usando as imagens com fogo da base ImaSta.

Pelos resultados obtidos é fácil ver que o canal "V" do YUV seguido pelo canal "u" do CIE Luv oferecem um maior contraste da cor vermelha em relação ao canal Cr do YCbCr. Esses dois canais passam no segundo critério C2 estabelecido anteriormente, no entanto, o tempo gasto para a transformação do RGB para CIE Luv é grande, fazendo com que ele venha a ser descartado no terceiro critério C3. O canal V do YUV

Médias de valores obtidos para as 50 imagens com fogo da base de dados ImaSta				
	V_{max}	$\overline{V}$	V_c	Tempo gasto para a transformação das imagens do RGB para o espaço de cores utilizado (ms)
Canal "V" YUV	227.66	155.06	72.59	2.26
Canal "u" CIE Luv	186.5	118.55	67.94	27.6
Canal "a" CIE LAB	184.02	141.31	42.70	4.7
Canal "Cr" YCbCr	209.28	149.52	59.75	2.34

Tabela 3.7: Resultados obtidos para os canais que representam a informação de crominância da cor vermelha para os diferentes espaços de cores nas 50 imagens com fogo da base de dados de imagens estáticas.

não só obteve os melhores resultados de contraste como também um resultado melhor no tempo de transformação, melhor até que o espaço de cor YCbCr, sendo este o canal escolhido para gerar a nova regra de classificação. A figura 3.15 mostra um exemplo visual da diferença de intensidade do fogo para cada canal testado.

As imagens da figura 3.15 ilustram os resultados obtidos na quarta coluna na tabela 3.7, no qual o canal V do YUV possui o maior contraste nas áreas da chama com cor vermelha, seguido pelo canal "u" do CIE Luv. O canal "a" do CIE Lab apresenta o pior resultado.

Usando os valores obtidos pelas equações 3.19 ($\bar{V}$) e 3.21 (V_c), uma nova equação é proposta aqui para calcular um limiar de segmentação T_{SV} baseado nas informações do canal V do espaço YUV. Esta equação tem um princípio bem similar ao que Çelik e Demirel (2009) e Chen et al. (2004) propõem, em que os pixels que compõem o fogo geralmente possuirão uma intensidade na crominância vermelha superior a média do resto da cena e provavelmente estarão mais próximos ao valor de saturação que o restante dos objetos. De forma a segmentar os pixels com cores mais próximas ao vermelho puro, o limiar T_{SV} consiste na média da cena capturada ($\bar{V}$) deslocada em função da diferença entre ela e o valor máximo de intensidade (V_c) da cena dividida por um valor de sensibilidade (β) que controla esse deslocamento, conforme pode ser visto na equação 3.22.

$$T_{SV} = |\bar{V} + \frac{V_c}{\beta}| \tag{3.22}$$

Para o método D383F, ao longo da realização dos experimentos neste trabalho, o valor do parâmetro β foi fixado experimentalmente em 1.8, valor este que alcançou uma ótima generalização para os diversos tipos de vídeos e imagens estáticas dentro do contexto de avaliar as bordas das chamas, aspecto que será explicado logo à frente.

A equação 3.22 permite que o limiar se altere dinamicamente de acordo com a cena capturada, gerando uma segmentação dos pixels com cores que geralmente variam de cores amarelo-alaranjado até o vermelho. A segmentação através deste limiar fornece, portanto, a informação de áreas com maior probabilidade de ser um foco de incêndio a partir das cores presentes na cena. A regra gerada usando a segmentação pelo canal



Figura 3.15: Exemplo da diferença de contraste entre os canais de diferentes espaços de cores. (A) Imagem original; (B) Imagem de intensidade do canal "V" do espaço de cores YUV; (C) Imagem de intensidade do canal "u" do espaço de cores CIE Luv; (D) Imagem de intensidade do canal "a" do espaço de cores CIE Lab; (E) Imagem de intensidade do canal "Cr" do espaço de cores YCbCr.

"V" do espaço de cor YUV é dada pela equação abaixo:

$$F_{5,D383}(x,y) = \begin{cases} 1, & \text{se } V(x,y) > T_{SV} \\ 0, & \text{caso contrário.} \end{cases} \tag{3.23}$$

Devido a esta característica importante de adaptação dinâmica, o uso deste modelo geralmente irá sempre manter as bordas das chamas como área segmentada, uma vez que as bordas das chamas são mais frias e variam de amarelo a vermelho puro (saturação no canal "V"), podendo ser utilizado para diferentes tipos de chama. Os resultados obtidos por este classificador/segmentador são usados em conjunto com as regras F_1 a F_4 do método proposto por Çelik e Demirel (2009), substituindo a quinta regra F_5 em que um modelo estatístico foi usado e funcionando como um comitê de dois classificadores heterogêneos.

A união de todas as regras e classificadores para o método D383F é análoga à equação 3.18,

no entanto, a informação de movimento obtida a partir da segmentação pela subtração de fundo deve ser levada em conta. Assim, a regra geral do método D383F para a classificação de um pixel é definida por:

$$F_{D383}(x,y) = \begin{cases} 1, & \text{se } \sum_{x=0}^{5} F_i(x,y) = 6 \\ 0, & \text{caso contrário.} \end{cases} \tag{3.24}$$

onde $F_{D383}(x,y)$ representa a classificação do pixel como sendo parte de uma chama ou não.

A figura 3.16 mostra em cores exemplos do funcionamento do classificador F_{D383} em diferentes vídeos da base de dados. Para mostrar o ganho obtido com o comitê de classificadores, são mostrados na terceira coluna em cor verde os pixels classificados como fogo ao se utilizar a segmentação por movimento (F_0) e as 4 regras de Çelik (F_1 a F_4), em que vários falsos positivos são gerados e em cor vermelha os pixels classificados como fogo ao adicionar a regra $F_{5,D383}$ às outras regras, ou seja, F_{D383}, em que apenas áreas das chamas são selecionadas, diminuindo a taxa de falsos positivos.

Somente a classificação por cores e movimento não é suficiente para um bom desempenho, já que é comum em ambientes externos objetos com cor de fogo se movimentarem com frequência devido à ação de agentes externos como galhos e folhas balançando ao vento. Para tratar este tipo de problema é necessária outra etapa que possa utilizar características singulares do fogo, sendo este o próximo passo, discutido na seção 3.4 a seguir.

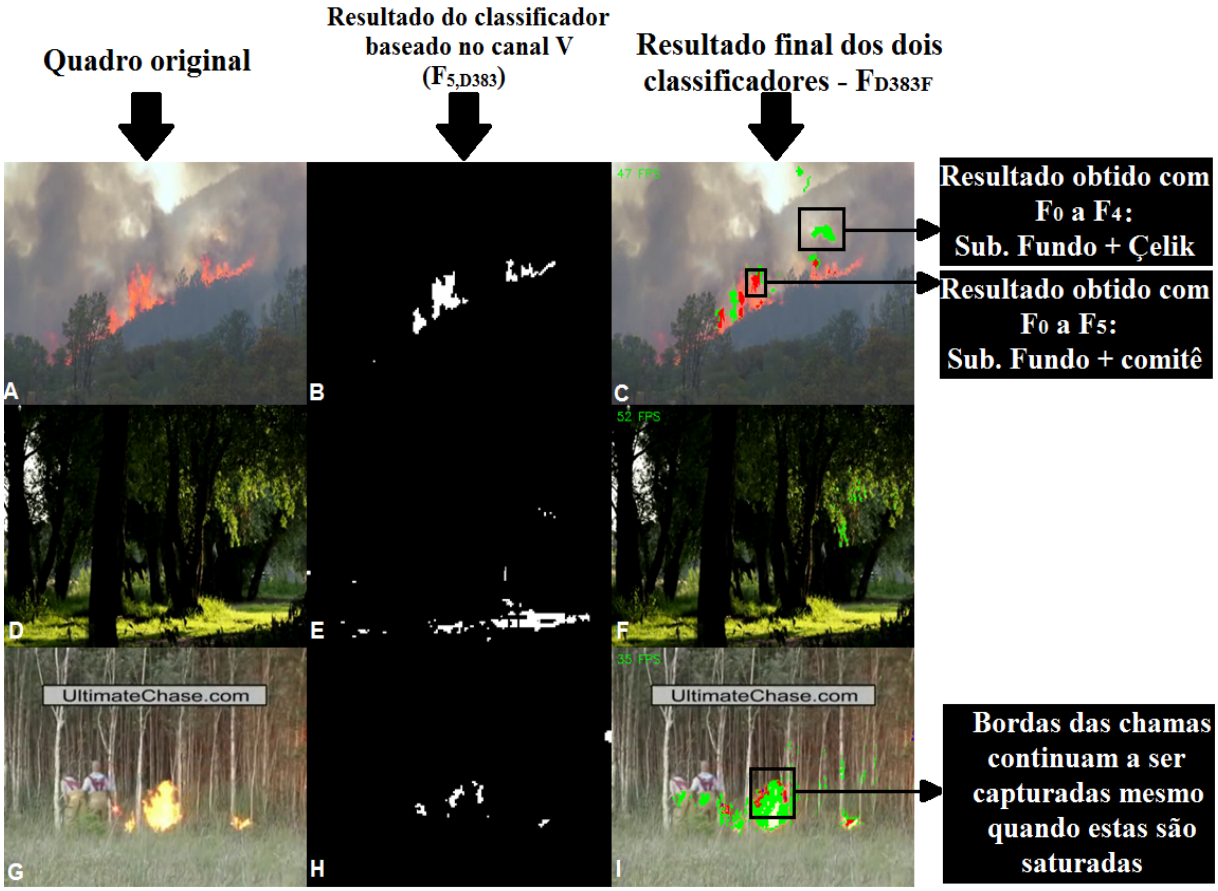


Figura 3.16: Exemplo de diferentes situações do funcionamento do classificador/segmentador F_{D383}. Cada linha possui respectivamente o quadro original, o quadro segmentado com as informações do canal "V" ($F_{5,D383}$) e o resultado final obtido. Na coluna de resultados, a cor verde representa os pixels classificados como fogo ao se utilizar a segmentação por movimento (F_0) e as 4 regras de Çelik (F_1 a F_4). A cor vermelha indica os pixels classificados como fogo utilizando a informação de movimento e as informações obtidas pelos dois classificadores (F_0 a $F_{5,D383}$), ou seja, F_{D383}. (A-C) No vídeo VD 1 é possível ver na segunda coluna praticamente toda a área da chama segmentada por $F_{5,D383}$ e na terceira coluna como a inserção da regra baseada no canal "V" ajuda a tratar falsos positivos capturados pelas regras propostas por Çelik; (D-F) No vídeo VDC 1 é possível ver como o ajuste dinâmico do limiar T_{SV} funciona na segunda coluna, mantendo segmentada parte da grama em amarelo como área de possível foco de incêndio, já que o amarelo é a cor mais próxima do vermelho saturado na cena; (G-I) No vídeo VD 4 é possível ver na terceira coluna o método capturando apenas as bordas das chamas que nesta cena estão com uma cor amarela clara, ignorando galhos e vestimentas dos bombeiros na cena que são classificadas como fogo pelas regras do método de Çelik.

3.4 Eliminação de falsos positivos através da persistência temporal e espacial dos objetos

Conforme Toreyin et al. (2007, 2006); Verstockt et al. (2012); Wang et al. (2010); Yuan (2010) citam, são necessários outros tipos de características para separar os objetos que possuem cor de fogo de uma chama verdadeira além dos vistos até aqui. Em todos esses trabalhos o movimento aleatório ao longo do tempo da chama é avaliado, nos quais alguns dos autores usam técnicas de processamento de sinais no domínio da frequência para classificar os pixels como sendo parte de uma chama ou não usando o efeito de cintilância (*flickering*). O método é bem robusto na maioria dos casos, falhando apenas em cenas específicas em que, por exemplo, o vento e as folhas em processo de abscisão foliar atuam em conjunto em certas distâncias, gerando padrões de pixels que são muito parecidos com os encontrados em uma chama. Schultze et al. (2006) falam sobre esse problema, onde os autores propõem o uso de outras técnicas em conjunto com técnicas de análise da frequência da cintilância quando uma cena apresenta diversos objetos que possuam essas características. Um exemplo do problema citado acima pode ser visto no vídeo VR 7 (figura 3.2).

Ao analisar vídeos com esse tipo de problema é possível verificar uma importante diferença quando comparado às chamas de um incêndio: geralmente estes tipos de objetos possuem um movimento semi-oscilatório e possuem na maioria das vezes um tempo de duração bem menor que as chamas de um incêndio. Outra característica importante nesta comparação é que o movimento de uma chama de incêndio é muito mais intenso em suas bordas, devido à turbulência gerada pela combustão, fazendo com que a chama altere seu volume e tamanho a cada instante e se mantenha em movimento continuamente.

Outro problema está no fato que avaliar os pixels no domínio da frequência exige um custo computacional alto devido à utilização de diversos quadros para que a informação de uma posição de pixel possa ser extraída. E quando se utilizam resoluções maiores, a área de interesse (*Region of interest* (ROI)) também cresce, tornando os métodos lentos para o uso em tempo real.

Baseado nestas características e no contexto proposto de atuar em tempo real usando diferentes resoluções, um novo modelo foi elaborado de forma a permitir o uso das características de movimento e existência temporal de uma chama para o método D383F.

Para que as características temporais e espaciais sejam usadas, é necessário primeiro obter algum tipo de informação que possa ser usado ao longo do tempo, sem que isso gere um custo computacional adicional de memória e processamento pelo uso de diversos quadros, sendo feito idealmente quadro a quadro. Dentro desse contexto um modelo de rastreamento "primitivo" dos objetos com cor de fogo foi elaborado, no qual os pixels classificados usando as etapas anteriores são primeiramente separados em grupos distintos e posteriormente extraídas as informações espaciais e temporais de cada grupo.

3.4.1 Agrupamento dos pixels com cor de fogo

Para reunir os pixels em conjuntos, novamente uma técnica de agrupamento foi utilizada, em que o contexto proposto deste trabalho exige que o algoritmo de agrupamento escolhido seja relativamente rápido e que não necessite de uma quantidade de grupos especificada a priori, já que não é possível saber quantos objetos com cor de fogo ocorrerão em uma cena. Isso descarta a técnica de k-médias e outros algoritmos semelhantes que necessitam desse parâmetro. A biblioteca computacional OpenCV oferece um algoritmo com essa característica em sua biblioteca, sendo um algoritmo simples de agrupamento de ordem quadrática O(N^2) que utiliza a idéia de estrutura de dados de conjuntos disjuntos (OpenCV (2013)). Um conjunto é dito disjunto quando ele não possui nenhum elemento pertencente a outro conjunto, ou seja, ele não possui interseção com outros grupos. No caso, desse trabalho, desejamos unir pixels e a decisão de dizer se um pixel pertence ou não a um conjunto será definida por uma operação de IGUALDADE baseada na distância euclidiana entre os pixels. Portanto, dois pixels são considerados como pertencentes ao mesmo conjunto quando estes estão próximos o suficiente um do outro na cena. Esta operação de IGUALDADE será mais bem detalhada adiante no trabalho.

Este tipo de algoritmo que trabalha com estas estruturas de dados é chamado de algoritmo de União-Procura (*Union-Find*) conforme Patwary et al. (2010) e Cormen et al. (2009) explicam, no qual a implementação usada pelo OpenCV utiliza árvores binárias para resolver o problema e cada árvore gerada pelo algoritmo representa um conjunto disjunto. Esse algoritmo de União-Procura com árvores binárias recebe o nome de florestas de conjuntos disjuntos e é composto pelas três operações primordiais a seguir:

- CRIAR-CONJUNTO (*MAKE-SET*) - Esta é a primeira operação realizada. Ela simplesmente cria um conjunto para cada elemento ou objeto (pixel), sendo que este objeto será o representante de seu grupo. Também é inicializado nessa etapa o *rank* ou posição, que se refere a uma das duas heurísticas utilizadas no algoritmo para melhorar seu desempenho. Esta heurística será explicada posteriormente em outra operação chamada UNIÃO. Abaixo o pseudocódigo referente a esta operação, onde o pai de um nó x é designado como $x.p$ (ponteiro) e o *rank* inicializado com valor 0:

  ```
  CRIAR-CONJUNTO(x)
    x.p = x
    x.rank = 0
  ```

- PROCURA-CONJUNTO (*FIND-SET*) - A segunda operação simplesmente procura e retorna o endereço de memória do representante do conjunto (raiz da árvore) sendo avaliado. Para melhorar o desempenho do algoritmo, essa operação executa a segunda heurística que realiza uma compressão do caminho, que consiste em fazer com que cada nó percorrido no caminho de procura aponte como pai a raiz da árvore. Veremos a seguir o pseudocódigo recursivo que realiza essa operação:

```
PROCURA-CONJUNTO(x)
  se x ≠ x.p então
    x.p = PROCURA-CONJUNTO(x.p)
  fim se
  retorne x.p
```

- UNIÃO (*UNION*) - A terceira operação efetua a união dos conjuntos reduzindo o número total de conjuntos disjuntos no espaço, esta operação é realizada quando um membro de um conjunto é considerado pertencente a outro segundo o critério de igualdade. Aqui, a heurística de *rank* é utilizada, no qual o conjunto de menor *rank* sempre é unido ao de maior *rank* e caso haja empate, um dos conjuntos é selecionado como pai, tem seu *rank* incrementado e o outro conjunto unido a ele. O pseudocódigo a seguir mostra implementação dessa operação, em que o procedimento CONEXÃO (LINK) é uma sub-rotina chamada pela operação UNIÃO que realiza a tarefa de atualização do nó pai e do *rank* quando necessário para as diferentes árvores:

```
UNIÃO(x,y)
  CONEXÃO(PROCURA-CONJUNTO(x),PROCURA-CONJUNTO(y))

CONEXÃO(x,y)
  se x.rank > y.rank  então
    y.p = x;
  senão
    x.p = y
    se x.rank == y.rank  então
      y.rank = y.rank + 1
    fim se
  fim se
```

Conforme dito anteriormente, é necessária uma operação de igualdade para definir se os pixels (ou nós) pertencem ao mesmo conjunto, assim dois pixels são considerados como pertencentes ao mesmo conjunto quando estes possuírem uma distância euclidiana quadrada menor do que o estipulado por um limiar chamado T_{dG}. No caso do método D383F, T_{dG} foi ajustado experimentalmente para conseguir um bom desempenho nos vários tipos de vídeos (chamas distantes e próximas) que compõem a base de dados, levando em conta a resolução e cena capturada do vídeo, em que a fórmula a seguir foi gerada através da resolução um sistema linear simples baseado nos melhores resultados obtidos para as resoluções máxima e mínima testadas:

$$T_{dG} = \left\lfloor \sqrt{img_w \times 0.227 + 65} \right\rfloor; \tag{3.25}$$

onde img_w se refere a largura da cena. O pseudocódigo da operação IGUALDADE usada pelo OpenCV pode ser visto abaixo:

IGUALDADE(T_{dg})
  para cada par de nós (x,y)
    se $x \neq y$ **e** d(x,y)< T_{dG} **então**
      UNIÃO(x,y)
    fim se
  fim para

Onde d(x,y) se refere a distância euclidiana quadrada entre os nós x e y. Como essa varredura de igualdade deve ser feita para todos os nós ou pixels, a complexidade do algoritmo se torna quadrática conforme os próprios desenvolvedores descrevem, o que pode fazer com que o método D383F tenha uma queda de desempenho quando houver muitos pixels provenientes das etapas de classificação anteriores, levando a implementação de outras técnicas para contornar este problema que serão detalhadas mais adiante no trabalho.

Para ilustrar o funcionamento do algoritmo de florestas de conjuntos disjuntos, tendo os conjuntos unitários disjuntos $\{a, b, c, d, e, f\}$ formados por um pixel cada e considerados parte de um mesmo agrupamento por satisfazerem o critério de igualdade, a figura 3.17 mostra sua união passo a passo, assumindo a seguinte ordem de abordagem dos agrupamentos seja: $\{a, b\}, \{d, e\}, \{a, c\}, \{b, e\}, \{e, f\}$. O passo 1 mostra os conjuntos disjuntos unitários em que cada um deles possui o *rank*=0. Os passos 2 e 3 executam a união de dois conjuntos unitários de *rank*=0 em que os *ranks* de "a" e "d" são incrementados durante o processo. O passo 4 executa a união de conjuntos com *ranks* diferentes, em que o conjunto de menor *rank* ("c") é anexado ao conjunto com maior *rank*. O passo 5 faz a união de conjuntos como mesmo *rank* através de nós filhos, em que os pais de cada nó filho são encontrados, levando ao acoplamento do conjunto {d,e} ao nó "a" e incremento do *rank* deste. O passo 6 executa a união de "f" por meio de um nó filho e mostra o funcionamento da heurística de compressão de caminho da operação PROCURA-CONJUNTO, em que o nó "e" passa a apontar para a raiz "a" e "f" com *rank* menor é anexado ao conjunto maior, gerando o conjunto completo.

A figura 3.18 mostra um exemplo do resultado obtido para o agrupamento de diferentes focos de incêndio no vídeo VR 5 ao se utilizar diferentes valores de T_{dG}, mostrando o quanto um grupo ou conjunto de pixels deve estar longe de outro para ser considerado um conjunto disjunto.

É conveniente neste ponto tratar alguns problemas que foram deixados de lado na parte da segmentação por movimento em função da almejada simplicidade e velocidade do algoritmo escolhido. Um exemplo são as alterações de iluminação repentinas que podem levar à captura de pixels que não estão em movimento até que a imagem de fundo se estabilize ou ainda podem ser frutos de ruídos gerados na captura da imagem pelo dispositivo. Esses pixels, se possuírem cor de fogo, chegarão a esta etapa do método, sendo necessário tratá-los devidamente para eliminar parte dos falsos positivos e garantir um bom desempenho do algoritmo.

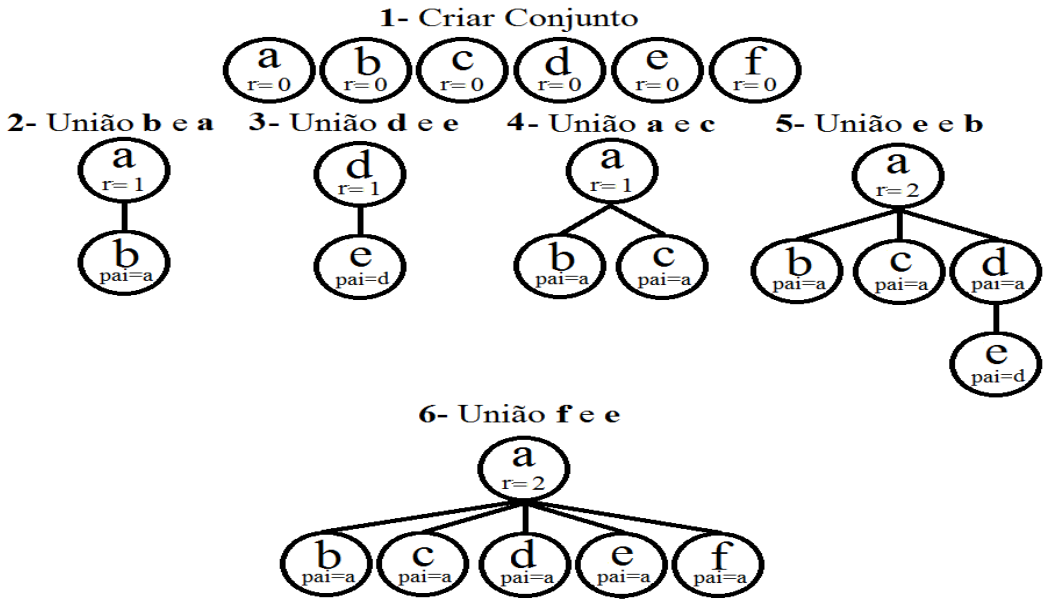


Figura 3.17: Geração de um conjunto pelo algoritmo de floresta de conjuntos disjuntos. (1) Operação Criar-Conjunto, no qual cada conjunto disjunto unitário é criado com *rank* = 0; (2 e 3) União de dois conjuntos unitários, o *rank* de "a" e "d" é incrementado; (4) União de conjuntos com *rank* diferentes, o menor é anexado ao maior; (5) União de conjuntos com mesmo *rank* através de nós filhos, o *rank* de "a" é incrementado e o outro conjunto é anexado ao mesmo; (6) União de "f" por meio de um nó filho, neste caso durante a operação PROCURA-CONJUNTO o nó "e" passa a apontar para a raiz "a" e "f" com *rank* menor é anexado ao conjunto maior, gerando o conjunto disjunto completo.

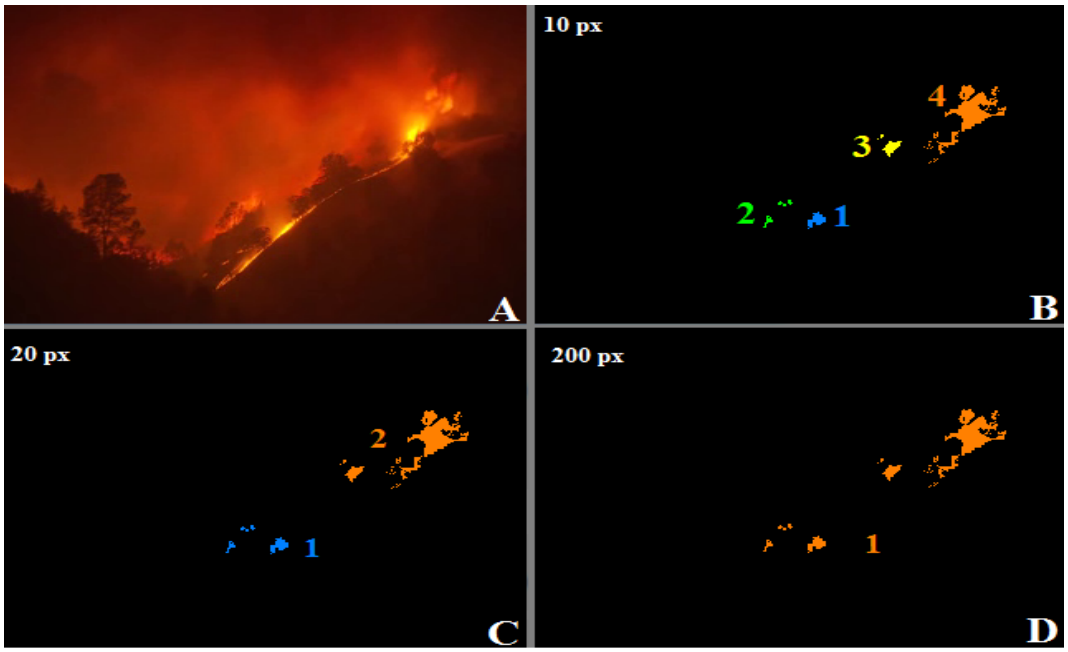


Figura 3.18: Agrupamento de diferentes focos de incêndio. (A) Quadro original; (B) Usando T_{dG}=10 pixels temos 4 focos distintos; (C) Usando T_{dG}=20 pixels temos 2 focos distintos; (D) Usando T_{dG}=200 pixels temos apena 1 foco de incêndio.

Como muitos destes pixels capturados em decorrência dos problemas citados acima possuem certa distância entre si e formam pequenos grupos com um número reduzido de pixels, é possível remover parte destes pixels facilmente nesta etapa de agrupamento através do uso de um parâmetro de limiar de tamanho mínimo do grupo T_{Gmin}. Para que um conjunto disjunto ou grupo G_i seja validado como parte do resultado do agrupamento **G** ($G_i \in \mathbf{G}$), ele deve possuir um certo número de elementos, problema este tratado por Çelik et al. (2007) em que eles declaram que o valor mínimo de um grupo deve ser de 5 pixels em vídeos de resolução igual a 176 x 144 pixels. No método D383F este mesmo valor será usado para vídeos em resoluções de 160 x 120 pixels e escalonado linearmente para as demais resoluções conforme a equação 3.26.

$$T_{Gmin} = \lfloor img_w \times 0.03125 \rfloor ; \tag{3.26}$$

onde img_w se refere a largura da cena.

Com esta técnica, parte dos falsos positivos já é tratado de acordo com a magnitude de seu tamanho, lembrando que isso terá pouco efeito sobre a detecção de incêndios florestais, visto que a chama de um incêndio sempre possuirá um grande volume de pixels relativos à sua área de abrangência. Assim, um conjunto disjunto só será parte do resultado de agrupamento segundo:

$$G_i \in \mathbf{G} \text{ se a quantidade de elementos de } G_i \geq T_{Gmin} \tag{3.27}$$

3.4.2 Características espaciais e temporais

Com os conjuntos ou grupos obtidos a partir dos pixels em movimento e com cor de fogo, outros tipos de falsos positivos podem ser tratados através das informações ou características espaciais e temporais de cada grupo, que serão detalhadas a seguir. Primeiro, o tratamento espacial dos grupos é feito, no qual cada grupo G_i tem seu ponto médio aritmético calculado, sendo este valor médio chamado de "centro de gravidade" do grupo ou Cg_i. Cabe frisar aqui que a letra "G" sempre que usada na discussão se refere a grupo e "g" a gravidade. A partir de Cg_i, é definido para cada G_i o quanto ele pode se mover, através de um parâmetro de limiar de movimento do grupo T_{mG}. Esse parâmetro é similar ao T_{dG}, ou seja, idealmente a resolução e o enquadramento da cena devem ser usados para determinar seu valor. Este valor em pixels é usado nas duas dimensões da cena (largura e altura), gerando uma caixa de contenção usada para observar por limiares o quanto um grupo se movimenta pela cena. A idéia do rastreamento proposto nesse trabalho se baseia na característica do fogo de se movimentar continuamente ao longo do tempo em torno do material combustível, ainda que seu tamanho e forma variem, seu centro de gravidade vai continuar dentro de uma área que se encontra "ancorada" à fonte de combustível. Em Çelik et al. (2007) os autores propõem uma idéia semelhante, em que eles geram uma área de observação retangular dos grupos com 2 vezes o valor da largura e altura do grupo para compensar o movimento do fogo e analisar a característica de contração e expansão do fogo ao longo do tempo. Em testes realizados nos vídeos das diferentes bases de dados (tabela 3.1), o valor de T_{mG} que obteve melhores resultados foi 2.5 vezes o valor de T_{Gmin}, sendo este um valor que conseguiu uma boa generalização para os

diferentes tipos de vídeos que fazem parte das bases de dados.

$$T_{mG} = \lfloor T_{Gmin} \times 2.5 \rfloor ; \tag{3.28}$$

Esta metodologia de acompanhamento do movimento dos grupos permite que um grupo se desloque espacialmente em uma área razoavelmente grande da cena ao longo do tempo, dando a liberdade necessária exigida pela característica de movimento aleatório de expansão e contração do fogo e que é usada por Çelik et al. (2007), Wen et al. (2010) e Zhou et al. (2010). Ela ainda permite excluir objetos com cor de fogo que se movam em uma trajetória linear e rapidamente pela cena, como um carro ou pessoa de passagem que podem facilmente apresentar a característica de expansão e contração na sua segmentação ao longo do tempo.

Uma vez descrito o tratamento espacial dos grupos, descreveremos agora seu tratamento ao longo do tempo. Para as informações temporais, é usada uma idéia baseada parcialmente na técnica de colônia de formigas, que adota uma abordagem probabilística para resolver problemas computacionais com o objetivo de achar caminhos ou soluções através de grafos. Em Dorigo et al. (1999), os autores ressaltam que idéias baseadas em insetos de comportamento social têm capturado a atenção de diversas áreas do conhecimento, devido ao nível de estruturação complexo de suas colônias quando comparado aos elementos individuais que formam as colônias, permitindo a criação de novas técnicas baseadas em idéias simples e eficientes para resolver diversos tipos de problemas.

A técnica de colônia de formigas foi proposta inicialmente por Dorigo (1992) e se baseia no comportamento das formigas em procurar um caminho de menor custo entre a colônia e a fonte de comida, usando um modelo bem simples no qual cada formiga, enquanto anda em um determinado caminho, deposita uma pequena quantidade de feromônio, formando uma trilha de feromônios para que as outras formigas possam usar.

Experimentos realizados por Goss et al. (1989) com uma raça de formigas argentinas em situações com caminhos longos e curtos entre a colônia e a fonte de alimento mostram o mecanismo de funcionamento dos feromônios, em que as formigas que escolhem o caminho mais curto conseguem fazer viagens mais rápidas e consequentemente colocar mais feromônios neste caminho, enquanto o caminho mais longo tem um número menor de viagens e em conjunto com a evaporação dos feromônios, se torna cada vez menos utilizado pelas formigas, sendo descartado posteriormente. Uma ilustração do experimento realizado por Goss et al. (1989) pode ser visto na figura 3.19.

Com base nestes trabalhos, uma nova metodologia foi proposta aqui, em que duas variáveis temporais baseadas na idéia de depósito e evaporação de feromônios foram adicionadas ao método D383F e são utilizadas em conjunto com as características espaciais de uma chama, na qual cada grupo é tratado analogamente a um caminho ou solução da técnica de colônia de formigas, e cada quadro possui formigas (uma por grupo) que visitam os grupos contidos na cena.

Como uma chama tem um tempo de duração extenso e movimento contínuo, existe

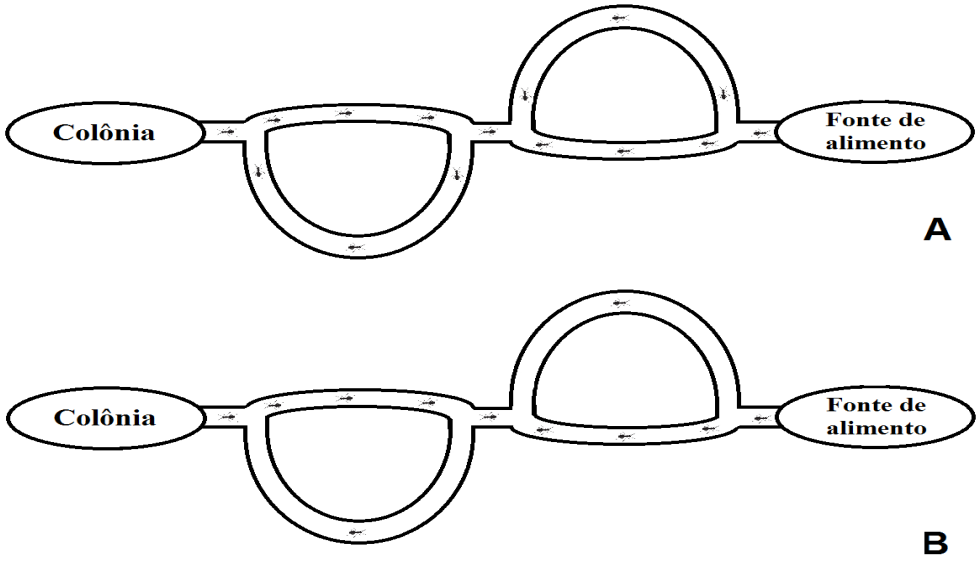


Figura 3.19: Experimento de Goss para colônias de formigas. (A) Caminhos de maior custo têm um espaçamento maior entre as formigas e viagens mais demoradas, tendo consequentemente uma quantidade menor de feromônios depositada; (B) Eventualmente a maior parte da formigas irá trafegar pelo caminho de menor custo.

certa similaridade com o experimento citado acima, em que a chama com seu movimento contínuo dentro de uma área em torno do combustível irá provocar um contínuo depósito de feromônios, que serão chamados a partir de agora no trabalho de existência ou *EX*. Com isso, a cada quadro, a *EX* do grupo que forma a chama é incrementada, assim como ocorre no caso do caminho curto das formigas. Para o caso de objetos com cor de fogo e de movimento semi-oscilatório como galhos e folhas, ou ainda, com movimento linear ao longo da cena, como pessoas ou carros, os grupos gerados por esses objetos terão um incremento menor de *EX* ao longo do tempo, não sendo capazes de superar a evaporação da *EX* e evitando possíveis falsos positivos. Estes tipos de objetos possuem certa similaridade ao caminho longo do experimento das formigas, onde há um depósito menor de feromônios. A evaporação de *EX* será controlada pela taxa de atualização ou *AT* dos grupos.

A partir das idéias propostas acima, uma estrutura de dados simples *MG* similar a uma tabela foi criada para guardar as informações relevantes de cada grupo detectado ao longo do tempo no vídeo. Cada elemento i desta estrutura corresponde a um grupo G_i de pixels detectados e agrupados anteriormente, contendo as seguintes informações: centro de gravidade do grupo (Cg_i), existência (EX_i) e atualização (AT_i). Uma linha da tabela MG pode ser representada da seguinte forma:

$$MG_i = [Cg_i, EX_i, AT_i] \text{ para i-ésimo grupo} \rightarrow \text{i de 1 até I} \tag{3.29}$$

EX_i contém a informação do tempo em quadros que MG_i está sendo capturado e persistido segundo suas características espaciais e temporais, enquanto AT_i informa há quantos quadros MG_i não recebe uma atualização.

Esta estrutura de dados permite um rastreamento primitivo dos objetos na cena, possibilitando que estes venham se unir e separar ao longo do tempo sem restrições, sem a necessidade de possuir rótulos ou identificação para o controle de rastreamento de objetos, etapa comum em sistemas de rastreamento de veículos ou pessoas. Em Bernardin e Stiefelhagen (2008) os autores introduzem métricas para avaliar o desempenho deste tipo de técnica e explicam os problemas geralmente encontrados quando os objetos da cena se sobrepõem, gerando falsos positivos que necessitam de várias metodologias para um rastreamento satisfatório dos métodos, demandando um custo computacional extra. No caso do método D383F, este tipo de rastreamento não se faz necessário, pois o interesse são as chamas, as quais tendem a se mover lentamente em relação à posição do combustível (<10 km/h) e diferentes focos tendem a se unir e separar constantemente quando muito próximos.

Para que EX_i seja incrementada, simulando o depósito de feromônios, é necessário que o grupo em movimento na cena tenha a existência atual constatada dentro da área determinada pelo parâmetro T_{mG} (característica espacial). Portanto, cada grupo G_a gerado pela técnica de agrupamento no quadro atual tem a distância de seu centro de gravidade Cg_a medida em relação aos demais I centros de gravidade anteriores Cg_i contidos em MG e se esta distância indicar que o grupo atual G_a está dentro da área de alcance de algum deles, ou seja, $|Cg_a - Cg_i| < T_{mG}$, Cg_i será atualizado para o novo valor contido em Cg_a (indicando que esse grupo é o mesmo grupo encontrado anteriormente e que seu centro de gravidade se alterou), sendo sua EX_i incrementada e sua AT_i zerada. Caso não exista nenhum grupo com Cg_i próximo, uma nova entrada MG_{I+1} é criada com o seu Cg_{I+1} recebendo o valor Cg_a, $EX_{I+1} = 1$ e $AT_{I+1} = 0$, onde I é número total de entradas de MG antes do acréscimo corrente.

Para formalizar o algoritmo, um pseudocódigo é mostrado seguir, usando como entrada o resultado do algoritmo de agrupamento:

```
ATUALIZA-MG-EXISTÊNCIA(G, MG, T_mG)
  n = 1
  A = comprimento(G)
  I = comprimento(MG)
  para a = 1 até A faça
    existe = falso
    Cg_a = calcula centro de gravidade (G_a)
    i = 1
    enquanto i ≤ I e existe == falso faça
      se |Cg_a − Cg_i| < T_mG então
        Cg_i = Cg_a
        EX_i = EX_i + 1
        AT_i = 0
        existe = verdadeiro
      fim se
      i = i + 1
    fim enquanto
    se existe == falso então
```

$Cg_{I+n} = Cg_a$
$EX_{I+n} = 1$
$AT_{I+n} = 0$
$n = n + 1$
fim se
fim para

Evidentemente, para que um grupo seja classificado como uma chama ele deve ter seu valor de existência comparado a um limiar de persistência espacial e temporal T_e. Este parâmetro idealmente deverá ser escolhido de acordo com a aplicação alvo de tal forma que galhos e folhas de árvores balançando ou pessoas passando pela cena sejam ignoradas pelo tempo de existência definidos pelo parâmetro T_e. Quando um grupo ultrapassa esse limiar, esse grupo é determinado como persistido espacialmente e temporalmente. Criamos então uma nova regra de classificação para o método D383F:

$$F_{Gi} = \begin{cases} 1, & \text{se } EX_i > T_e \\ 0, & \text{caso contrário.} \end{cases} \tag{3.30}$$

onde T_e nesse trabalho é usado no capítulo 4 com diferentes valores que permitem obter o melhor desempenho dentro do contexto proposto pelos experimentos. No caso de utilização real do algoritmo, espera-se que quanto maior o tempo de persistência menor será a incidência de alarmes falsos detectados pelo algoritmo, aumentando a confiabilidade da classificação/detecção.

Ao final do processamento de cada quadro, o parâmetro AT_i de cada MG_i recebe um incremento de valor 1, sendo depois calculada a evaporação parcial ou decremento de EX_i, que ocorre quando um grupo deixa de ser atualizado por algum tempo, no caso definido experimentalmente como 1/3 de T_e, obtendo um bom desempenho para as bases de vídeos utilizadas (tabela 3.1). Lembrando que a atualização de um grupo é feita em ATUALIZA-MG-EXISTÊNCIA quando a EX_i é incrementada e AT_i zerada. Por último, um grupo MG_i é apagado da tabela se o mesmo não for atualizado pelo tempo estipulado por T_e ($AT_i > T_e$, ou seja, evaporação completa), evitando que a tabela cresça e que grupos formados por objetos muito antigos façam parte do modelo, conforme mostrado abaixo em pseudocódigo executado a cada quadro de vídeo:

MG-EVAPORAÇÃO(MG, T_e)
$I = comprimento(MG)$
para $i = I$ **até** 1 **faça**
$AT_i = AT_i + 1$
se $AT_i > T_e/3$ **então**
$EX_i = EX_i - 1$
fim se
se $AT_i > T_e$ **então**
Apaga MG_i
fim se
fim para

É importante frisar que devido à metodologia de movimento dos grupos adotada, pode ocorrer uma situação em que o centro de gravidade de um grupo MG_i representando uma chama seja levado pelo movimento ascendente da mesma para longe da fonte de combustível e de forma abrupta retorne para perto dela. No próximo quadro, o Cg_a da chama estará fora de alcance da caixa de movimento delimitada por T_{mG}, não podendo atualizar Cg_i. Nesse caso a chama entrará na estrutura de dados como um novo grupo MG_{I+1} a ser persistido e duas situações podem ocorrer: o Cg_i poderá ser revisitado pela chama, em que o Cg_a da chama em algum momento entre no alcance da caixa de movimento de Cg_i, trazendo-o de volta para as proximidades da fonte combustível e continuando a persistir, onde a entrada MG_{I+1} é apagada quando esta superar T_e; Ou Cg_i poderá ser excluído se o Cg_a da chama não caminhar na direção de Cg_i dentro da janela de duração imposta pelo parâmetro T_e, sendo a entrada na estrutura de dados que representa a chama atualmente MG_{I+1} persistida quando sua existência alcançar novamente o valor T_e.

Finalmente, para melhor visualização dos resultados, os pixels que compõem os grupos detectados em cada quadro são usados para calcular uma caixa de contenção (coordenadas dos pixels mais externos de cada grupo), que é mostrada na cor amarela, indicando que o objeto que está sendo observado é um possível foco de incêndio. Neste caso a regra F_{Gi} ainda não foi satisfeita. Quando F_{Gi} é satisfeita, a caixa fica vermelha, indicando que este grupo de pixels é um foco de incêndio persistido segundo suas características espaciais e temporais na cena capturada. Portanto, a regra final para efetuar o alarme de incêndio é dada quando algum grupo obtiver a persistência temporal e espacial, indicando a existência foco de incêndio na cena.

$$ALARME = \begin{cases} 1, & \text{se } \exists\, F_{Gi} = 1; \\ 0, & \text{caso contrário.} \end{cases} \tag{3.31}$$

A figura 3.20 mostra o resultado obtido no vídeo VR 5, no qual podem ser vistos 3 focos de incêndios já persistidos e um deles em estado de observação.

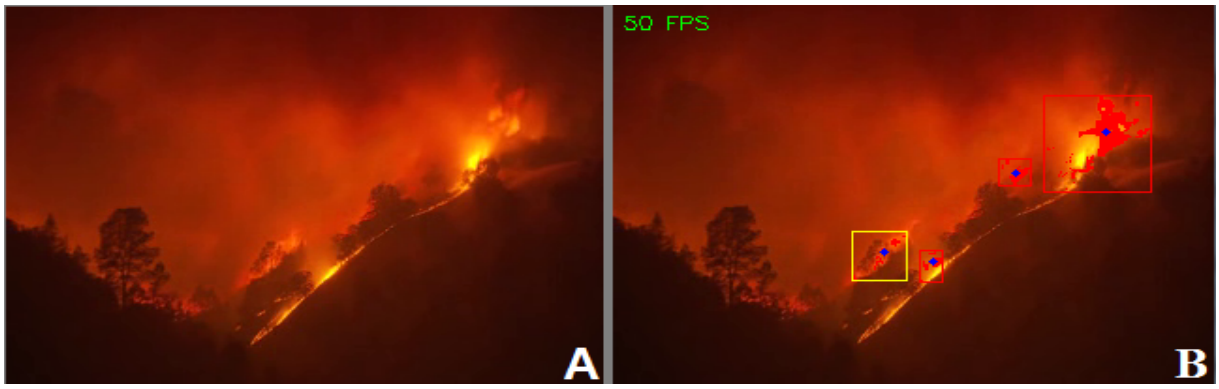


Figura 3.20: Resultado final obtido usando a persistência espacial e temporal. (A) Quadro original; (B) 4 focos de incêndio distintos. O círculo azul representa o centro de gravidade do grupo e os quadrados amarelo e vermelho indicam se o grupo foi ou não classificado como um foco de incêndio, respectivamente, segundo a regra F_{Gi}. Observa-se que um deles ainda não foi persistido temporalmente.

Deste modo, fica definido o processo de detecção dos focos de incêndio do algoritmo D383F. Metodologias suplementares para permitir o uso desta técnica em vídeos de resoluções mais altas são explicadas a seguir.

3.5 Técnicas para redução de custo computacional

Foram mencionadas anteriormente no trabalho as implicações do uso de imagens em alta resolução para a detecção de focos de incêndios, em que os seres humanos tiram vantagem de sua grande capacidade de processamento de informação visual e as máquinas têm um aumento no custo computacional para processar essas imagens, levando em conta que o ganho de desempenho de classificação pode não compensar devido às limitações de avaliação das mesmas.

Como a monitoração ambiental tem grande importância para o futuro, tendo em vista as previsões de consumo de recursos pela população humana nos próximos anos relatadas em WWF (2012), é interessante oferecer uma solução que permita a ambas as partes, humanos e máquinas, trabalhar da melhor maneira possível, sendo esta a motivação da utilização das metodologias mostradas a seguir.

3.5.1 Subamostragem de pixels

Uma das alterações propostas para reduzir o custo computacional do algoritmo é a subamostragem de pixels, em que neste caso específico procura resolver problemas relacionados principalmente à parte de classificação de pixels segundo sua cor em resoluções maiores que 352 x 288 pixels.

Um exemplo de problema que pode ocorrer facilmente em incêndios florestais é quando uma grande quantidade de fumaça gerada pelo fogo altera significativamente grande parte do movimento da cena, levando o algoritmo de subtração de fundo a segmentar praticamente toda a cena e passar toda essa informação para a classificação por cor.

De forma a resolver este problema, é realizada a subamostragem dos pixels em movimento da cena para resoluções maiores que 352 x 288 pixels. Para as resoluções até 1280 x 720 pixels, apenas os pixels ímpares na direção da largura da imagem são analisados, e na resolução de 1920 x 1080 pixels apenas os pixels nas posições ímpares em ambas as direções são processados. Isto implica em uma alteração da regra F_0 (3.11), que passa a possuir as seguintes condições:

$$F_0(x,y) = \begin{cases} 1, & \text{se } sub(x,y) > T_s \text{ e } sub(x,y) \in S_b \\ 0, & \text{caso contrário.} \end{cases} \tag{3.32}$$

onde S_b é a área definida como:

S_b	**Condição**
Toda imagem	$img_w \leq 352$
Colunas/ x ímpar	$352 < img_w < 1920$
Colunas/ x ímpar Linhas/ y ímpar	$img_w = 1920$

Tabela 3.8: Tabela de subamostragem de pixels.

Com esta técnica e a plataforma computacional relatada futuramente no capítulo 4 é possível manter o algoritmo funcionando em tempo real nas resoluções maiores que 352 x 288 pixels, com uma taxa de 10 quadros por segundo até 1280 x 720 pixels e na resolução de 1920 x 1080 pixels com uma taxa de 5 quadros por segundo.

No entanto, para executar essa subamostragem é necessário tomar cuidado com a perda de informação e alguns efeitos indesejados de *aliasing* espacial que podem ocorrer em função disso. Em palavras simples, o *aliasing* ocorre quando um sinal em alta frequência se comporta como um sinal em baixa frequência devido a uma amostragem pobre do fenômeno sendo observado. No caso de imagens, o efeito acontece quando a amostragem utilizada (número de pixels) não está sendo alta o suficiente para capturar a quantidade de detalhes dos objetos que formam a imagem, gerando efeitos como o padrão *moiré* (padrões indesejados em algumas texturas) ou *jaggies* (serrilhado em bordas) que podem ser vistos na figura 3.21.

Para evitar esse problema, segundo o teorema de Nyquist, a taxa de amostragem mínima a ser utilizada é definida como o dobro da frequência máxima do sinal analisado. A frequência máxima do sinal, conhecida como frequência de Nyquist (cuja abreviatura no trabalho será f_{Nyq}), é determinada com base em um limiar arbitrário de relevância geralmente tomado em função da intensidade ou da potência do sinal.

A figura 3.21A mostra exemplos do que acontece quando esse teorema não é respeitado, em que o número de pixels utilizados não é suficiente para representar a informação contida na parede de tijolos, sendo difícil após a subamostragem dizer se os objetos que formam aquela parede são realmente tijolos. Nesta imagem em especial, esta parede apresenta uma textura de alta frequência espacial (número grande de repetições ou ciclos) em que cada ciclo (tijolo) necessita de um número mínimo de pixels ou unidades de distância (u.d.) para que a informação possa ser representada.

Portanto, para a tarefa de avaliar se a subamostragem utilizada nesse trabalho está evitando a perda de informação e encontrar a frequência máxima do sinal das imagens (f_{Nyq}) e, consequentemente, a amostragem necessária para que a informação presente na imagem possa ser representada de maneira adequada, é necessário avaliar estas imagens no domínio da frequência, em que utilizamos uma ferramenta muito conhecida da área de processamento de imagens chamada transformada de Fourier. Essa técnica consiste na transformação da informação do domínio do espaço para o domínio da frequência espacial. Para um maior detalhamento da técnica e suas definições, indicamos a leitura de Gonzalez e Woods (2010). No nosso trabalho iremos focar somente na parte que diz respeito ao uso de imagens, onde se faz necessário usar a transformada

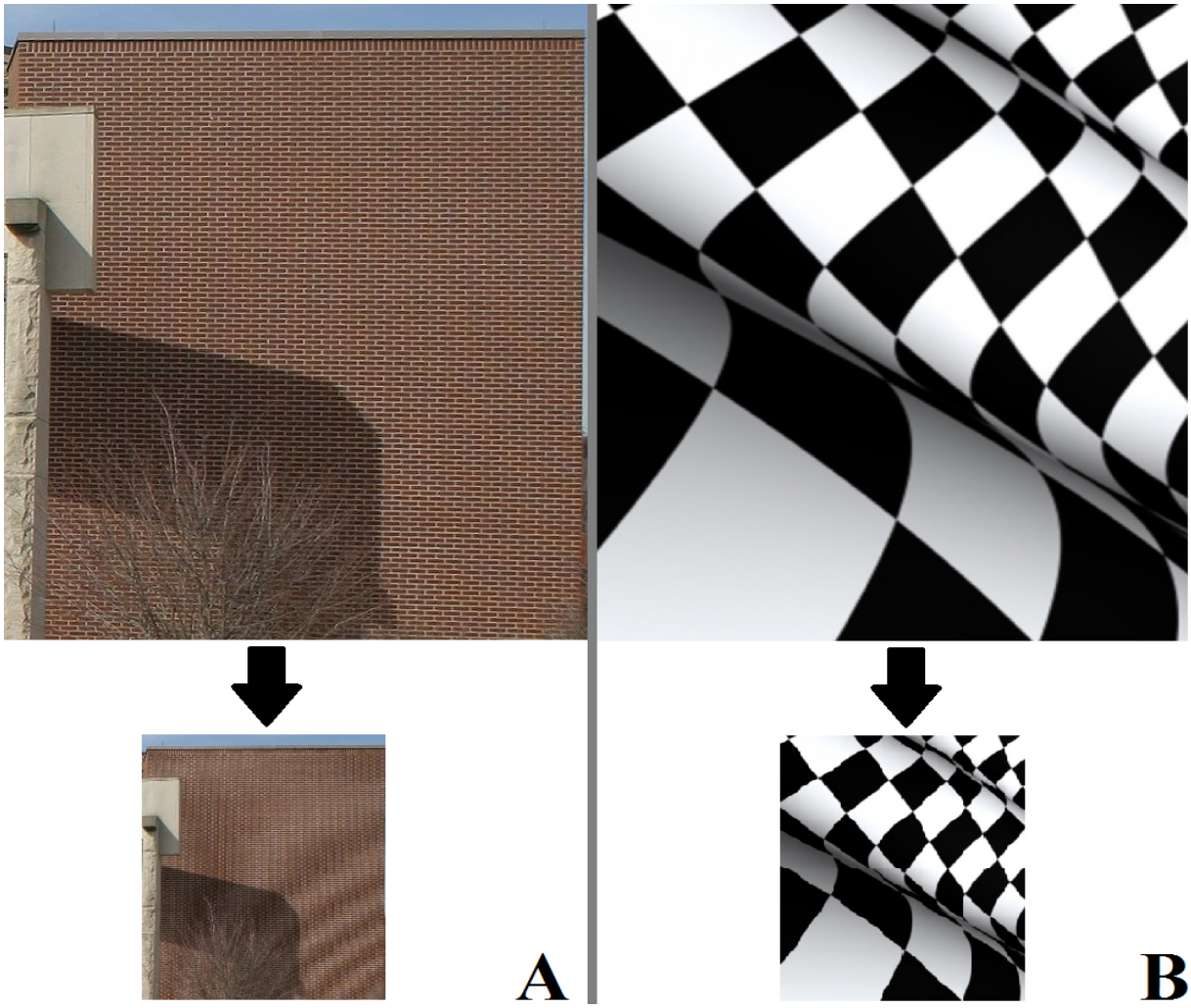


Figura 3.21: Efeito *moiré* e *jaggies* causados por subamostragem, (A) Efeito *moiré* causado pela subamostragem em uma parede de tijolos; (B) *Jaggies* em uma figura no estilo tabuleiro de xadrez causado pela subamostragem. Perceba que as bordas dos quadrados não são suaves e estão ligeiramente deformadas. FONTE: Wikipédia

discreta de Fourier (*discrete Fourier transform* - DFT), realizada computacionalmente através do algoritmo de transformada rápida de Fourier (*Fast Fourier Trasform - FFT*). A transformada discreta de Fourier $F(u,v)$ de uma imagem $f(x,y)$ em escala de cinza é definida como:

$$F(u,v) = \frac{1}{MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y) e^{[-j2\pi(\frac{ux}{M} + \frac{vy}{N})]} \tag{3.33}$$

para os índices u e v variando nas faixas $u = (0,1,2,...,M-1)$ e $v = (0,1,2,...,N-1)$, $j = \sqrt{-1}$, M e N sendo a altura e largura da imagem, respectivamente.

Um dos resultados obtidos dessa transformação é a imagem do espectro Fourier ou imagem de magnitude, que permite avaliar as características de frequência presentes na imagem original. O resultado da transformada de Fourier é geralmente em números

complexos e a imagem de magnitude é definida por Gonzalez e Woods (2010) como:

$$|F(u,v)| = [R^2(u,v) + I^2(u,v)]^{\frac{1}{2}} \tag{3.34}$$

onde R e I são respectivamente as partes real e imaginária do valor complexo.

A partir dessa imagem de magnitude podemos então avaliar quais as frequências têm maior relevância na imagem original e definir a partir disso a frequência máxima do sinal (f_{Nyq}) e a amostragem mínima desse sinal (taxa de amostragem de Nyquist) necessária para o processamento sem a perda de informação significante para a classificação, permitindo uma comparação desse valor com a subamostragem que adotamos.

Para realizar esse experimento, iremos utilizar quadros dos vídeos VR 1 e VR 8, ou seja, um vídeo com fogo e outro sem fogo da base VideoRes que representam os casos mais difíceis para a detecção pelo algoritmo. O vídeo VR 1 apresenta uma movimentação intensa de fumaça gerada pelo incêndio em grande parte da cena e o vídeo VR 8 é o vídeo de controle mais desafiador, apresentando movimento de folhas em torno do pôr-do-sol que se comportam como chamas, podendo gerar falsos positivos na detecção. Os quadros desses vídeos podem ser vistos na figura 3.22.



Figura 3.22: Quadros dos vídeos VR 1 e VR 8 a serem utilizados no experimento. (A) Quadro do vídeo VR1; (B) Quadro do vídeo VR 8.

Nas figuras 3.23A e B podemos ver como a frequência é representada na imagem de magnitude para os quadros dos vídeos VR 1 e VR 8 mostrados na figura anterior na resolução de 1920 x 1080 pixels, caso em que a subamostragem utilizada é a maior, sendo 50% (pixels de coordenadas ímpares nas direções horizontais e verticais). Nessas imagens de magnitude cada pixel representa uma frequência ou ciclos por u.d., onde a frequência do sinal aumenta à medida que se afasta da origem e quanto mais escuro, maior a amplitude do sinal.

No caso de imagens de ambientes externos, elas geralmente são formadas por uma quantidade considerável de objetos aleatórios observados a uma distância razoavelmente grande (árvores e arbustos de formatos variados, objetos de construção humana, etc.) em que predominantemente apresentam uma frequência espacial relativamente

baixa, ou seja, esses objetos ocupam porções significativas da imagem e não possuem grandes alterações de cores ou formas em sua extensão (lembre-se que geralmente as folhas de uma árvore observadas à longa distância formam um bloco verde maciço que define o objeto árvore na imagem). Com isso, as imagens de magnitude dessas imagens apresentam um formato característico de informação concentrada na origem, ou seja, o centro da imagem apresenta valores de amplitude altos que decrescem à medida que se afasta da origem. Isso deixa claro que as informações de alta frequência têm menos relevância nesse tipo de imagem, conforme pode ser facilmente visto nas figuras 3.23A e B.

Tendo isso em mente, a definição da frequência máxima relevante do sinal (f_{Nyq}) e a amostragem necessária para a representação da imagem sem a perda de informação foram feitas de maneira bastante simples, em que a imagem de magnitude foi normalizada e analisada em escala logarítmica (dB), conforme:

$$Atenuação_{dB} = 20 \times log_{10}(\frac{|F(u,v)|}{\max|F(u,v)|}) \tag{3.35}$$

Desta forma, definiu-se a frequência máxima do sinal com base em um limiar de relevância de 40 dB. Ou seja, frequências com magnitude apresentando atenuação superior a 40 dB (10000 vezes) são consideradas irrelevantes.

O resultado obtido pode ser visto nas figuras 3.23C e D, sendo mostrado em zoom um dos quadrantes nas figuras 3.23E e F respectivamente. Em cada uma destas figuras existe um plano de corte em -40 dB, abaixo do qual está grande parte do espectro. Consideramos todo o sinal abaixo desse valor como desprezível, o que deixa apenas uma pequena parte do espectro representando a informação relevante presente da imagem. A partir desse resultado podemos definir a frequência máxima do sinal (f_{Nyq}) correspondente ao limiar de -40 dB, selecionando o pixel mais afastado da origem em cada eixo da imagem de magnitude com valor maior que -40 dB como sendo a f_{Nyq} da imagem original e definindo a taxa de amostragem de Nyquist como o dobro da distância desses pixels em relação à origem. Ou seja, no caso da figura 3.23A, representando o resultado do quadro do vídeo VR 1, o valor da f_{Nyq} para o limiar de -40 dB no eixo horizontal (H) está a uma distância de 25 pixels da origem e no eixo vertical (V) está a 55 pixels (veja indicação na figura), logo a taxa de amostragem de Nyquist nesse caso seria o dobro, i.e. 50 ciclos/u.d. e 110 ciclos/u.d. respectivamente, formando a elipse interna da figura 3.23A. De maneira semelhante, o mesmo é feito para um quadro do vídeo VR 8, conforme se vê na figura 3.23B, com a taxa de amostragem de Nyquist de 140 ciclos/u.d e 110 ciclos/u.d. no eixo H e V respectivamente.

Outra informação importante que pode ser abstraída do espectro de Fourier é a potência de regiões da imagem, que pode ser feita facilmente conforme:

$$\text{Potência da região da imagem} = \frac{\sum_{u,v \in elipse}|F(u,v)|^2}{\sum_{u,v \in imagem}|F(u,v)|^2} \times 100 \tag{3.36}$$

Estes valores nos dão a idéia de quanta informação está sendo retida. Note que no caso

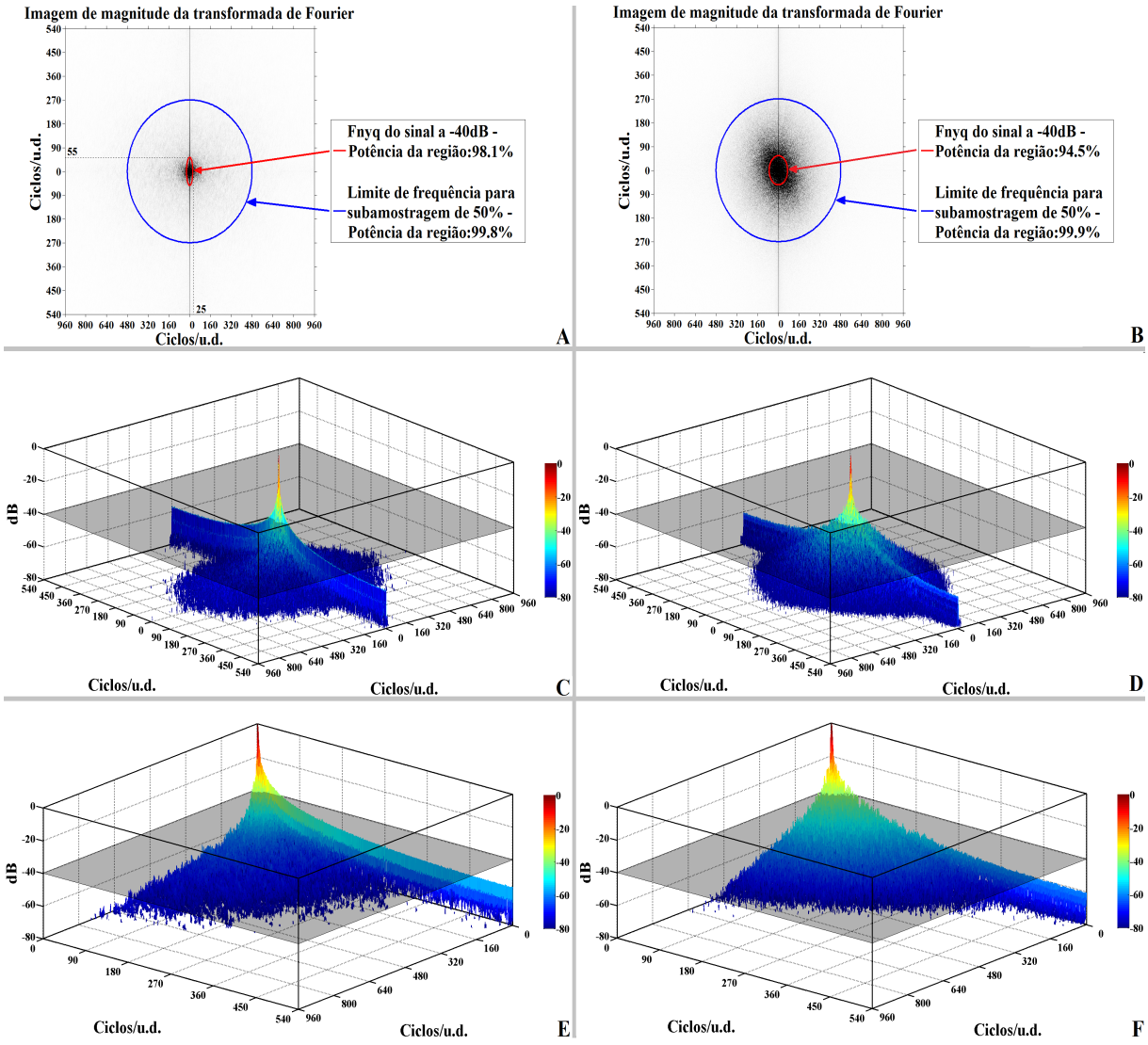


Figura 3.23: Imagens do espectro de Fourier e os gráficos de atenuação do sinal dos quadros dos vídeos VR 1 e VR 8 na resolução de 1920 x 1080 pixels. (A e B) Imagens de espectro ou magnitude da transformada de Fourier dos quadros. Elipses mostram na cor vermelha a frequência máxima (f_{Nyq}) do sinal considerado como relevante (atenuação <40 dB) e na cor azul a frequência máxima utilizada em nossa subamostragem de 50% da imagem, apenas com colunas e linha ímpares; (C e D) Gráficos de atenuação do sinal gerado a partir da imagem de magnitude. Um plano de corte mostra a parte do sinal mais relevante, permitindo avaliar qual a frequência máxima do sinal; (E e F) Zoom em um dos quadrantes dos gráficos de atenuação do sinal. Pode-se concluir que a subamostragem utilizada por nós ainda tem uma grande folga em relação à frequência máxima relevante do sinal.

do quadro do vídeo VR 1, mostrado na figura 3.23A, temos a área da elipse interna (f_{Nyq}) possuindo mais de 98% da potência da imagem original (perda de apenas 1,8%) e a área da elipse externa (referente a nossa subamostragem com colunas e linhas ímpares) possuindo 99.8% da potência (perda de apenas 0,2% da potência do sinal). No caso do quadro do vídeo VR 8, mostrado na figura 3.23B, temos 94,5% da potência da imagem original na elipse interna e 99.9% na subamostragem que utilizamos. Fica claro em

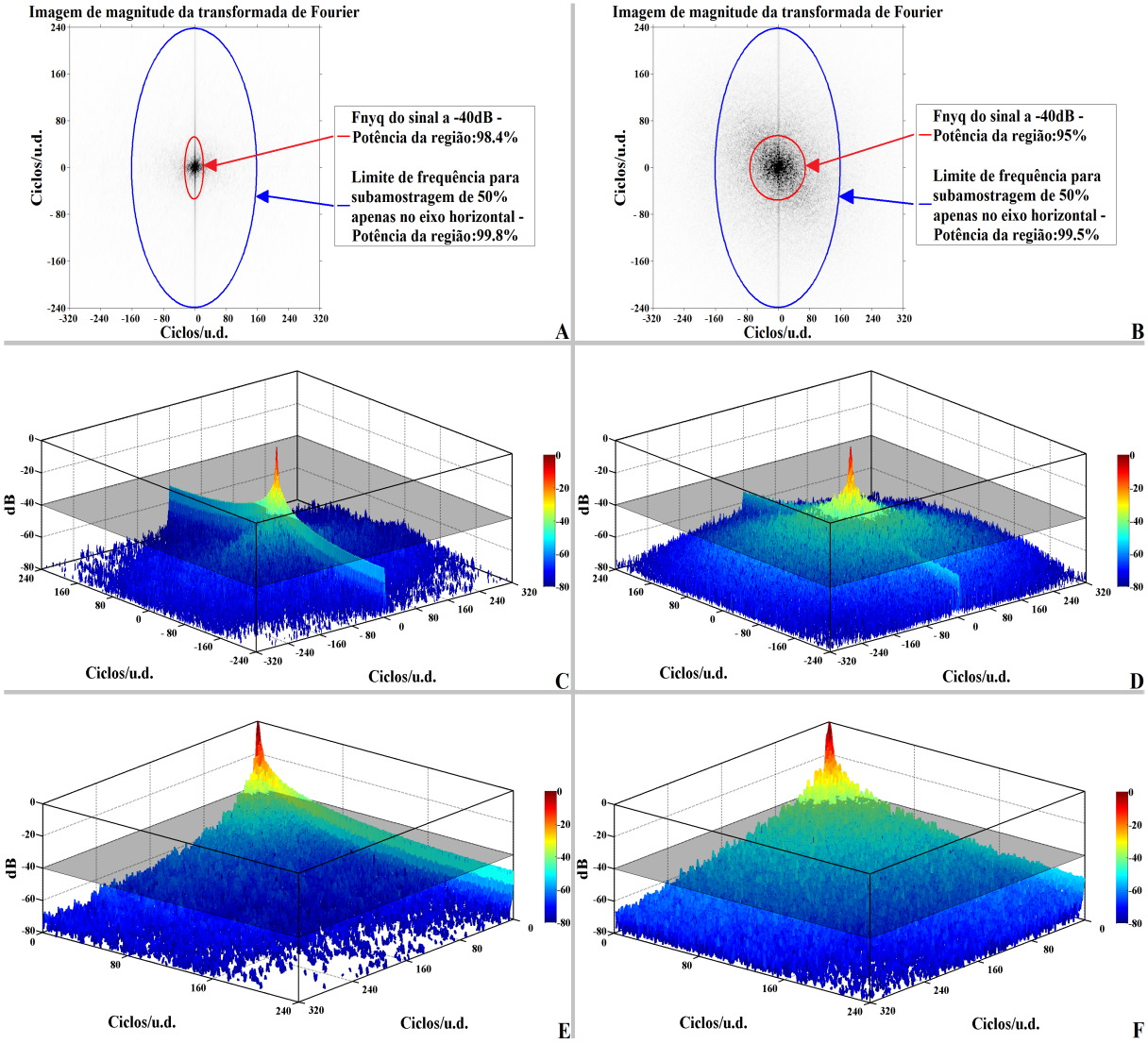


Figura 3.24: Imagens do espectro de Fourier e os gráficos de atenuação do sinal dos quadros dos vídeos VR 1 e VR 8 na resolução de 640 x 480 pixels. (A e B) Imagens de espectro ou magnitude da transformada de Fourier dos quadros; (C e D) Gráficos de atenuação do sinal; (E e F) Zoom em um dos quadrantes dos gráficos de atenuação do sinal; Vemos que a subamostragem utilizada apenas no eixo x ainda está respeitando a frequência máxima relevante do sinal.

ambas as imagens que o limite para a frequência máxima imposto pela subamostragem por linhas e colunas ímpares está bem distante da frequência máxima real do sinal e que utiliza quase a totalidade da informação contida na imagem. Em termos quantitativos em nosso exemplo, a amostragem de 960 x 540 pixels utilizada (elipse mais externa das figuras 3.23A e B), impõe um limite máximo de 480 ciclos/u.d e 270 ciclos/u.d. para os eixos H e V respectivamente, onde a taxa de amostragem de Nyquist para o sinal real seria de aproximadamente 50 x 110 pixels para o vídeo VR 1 e 140 x 110 pixels no VR 8. A figura 3.24 mostra as mesmas informações de imagem de magnitude e gráficos de atenuação do sinal referentes aos vídeos VR 1 e VR 8 na resolução de 640 x 480 pixels, caso em que apenas o eixo horizontal da imagem é subamostrado. Podemos ver que de maneira semelhante, a subamostragem utilizada ainda tem uma folga relativamente

alta em relação à amostragem mínima necessária definida pela freqüência máxima do sinal das imagens.

Portanto, podemos concluir pelos resultados mostrados que a subamostragem utilizada em nosso trabalho permite a classificação sem ter como consequência a perda de informação relevante nas imagens e os efeitos indesejados do *aliasing*. Posteriormente no trabalho, para mostrar que informação descartada não tem influência sobre o processo de classificação, um teste piloto na resolução de 1920 x 1080 pixels é mostrado, no qual todos os pixels da imagem são utilizados para gerar os resultados. Nesse teste, cada quádrupla de pixels (o pixel de coordenadas ímpar e seus vizinhos) tem sua média aritmética calculada e usada como se fosse apenas um pixel na etapa de classificação por cores, à custa de um pequeno incremento no custo computacional, mostrando que essa informação extra não adiciona nenhuma melhoria no desempenho de classificação.

Esta metodologia de subamostragem adotada ainda oferece outra vantagem de facilitar excluir objetos que não são fogo, uma vez que ao se aumentar a resolução não se aumenta apenas a informação da chama, mas também dos objetos que podem se comportar como tal, conforme dito anteriormente no trabalho. Como a chama de um incêndio geralmente vai possuir grandes dimensões e as características de interesse são mais significativas, é de se esperar que, mesmo com a subamostragem, um grande número de pixels referentes às chamas continue sendo capturado. Enquanto galhos e árvores balançando, que possuem uma magnitude menor, terão o número de pixels reduzido pela subamostragem, facilitando sua exclusão durante a validação dos grupos pelo número de elementos.

3.5.2 Ajuste automático do limiar de subtração de fundo

Como dito anteriormente na seção 3.4.1, a técnica de agrupamento possui uma complexidade quadrática, que pode fazer com que o algoritmo sofra uma queda de desempenho considerável no tempo de processamento se um objeto na cena tiver uma área grande de movimento e possuir cor de fogo. Um exemplo desse tipo de situação seria uma chama ou outro objeto qualquer com cor de fogo muito próximo da câmera. Tendo em vista este tipo de problema, foi desenvolvida uma metodologia para evitar que o algoritmo tenha um aumento significativo no tempo de processamento nessas condições.

A idéia em si é bem simples e usa a etapa de segmentação por movimento para resolver esse problema. Como a segmentação por movimento dita quantos pixels são avaliados posteriormente pelas etapas de classificação por cores e agrupamento na persistência temporal e espacial, o parâmetro T_s (3.11), que controla o quanto de movimento um pixel precisa ter para ser segmentado, tem seu valor alterado dinamicamente durante a execução do algoritmo em função do tempo de processamento gasto T_p, tempo esse medido ao final do processamento de cada quadro.

Para isso, uma variável de limiar tempo máximo T_{pmax} é usada para manter o algoritmo sempre rodando dentro de uma faixa de tempo limitada pela taxa de quadros por segundo do vídeo, por exemplo, um vídeo a 10 quadros por segundo permite que

cada quadro seja processado em no máximo 100 milissegundos. Toda vez que este tempo é extrapolado, implica que muitos pixels estão sendo capturados na etapa de segmentação e um ajuste se faz necessário. Este ajuste é feito por um parâmetro de passo chamado I_{Ts}, em que este irá incrementar ou decrementar o limiar T_s, com valores preferencialmente baixos entre 1 e 10.

Neste modelo, o valor de T_s atribuído pelo usuário será o valor mínimo possível para a segmentação, sendo armazenado no parâmetro T_{smin}, desta forma T_s irá variar dentro do intervalo $T_{smin} \leq T_s < 255$ automaticamente de acordo com tempo de processamento medido Tp.

O pseudocódigo abaixo demonstra o funcionamento do modelo proposto:

AJUSTE-T_s(T_p, T_{pmax}, T_{smin}, T_s)
se $T_p \leq T_{pmax}$ **então**
se $T_s > T_{smin}$ **então**
$T_s = T_s - I_{Ts}$
fim se
senão
se $T_s \leq 255$ **então**
$T_s = T_s + I_{Ts}$
fim se
fim se

Com esta implementação, mesmo em casos extremos, as bordas das chamas continuarão a ser detectadas normalmente, já que essas possuem uma maior magnitude de movimento, conforme dito anteriormente.

Esta metodologia, no entanto, não foi utilizada ao longo dos experimentos aqui realizados para não alterar os valores de T_s em tempo de execução e permitir que o tempo de processamento gasto em cada vídeo seja avaliado em sua forma plena, podendo esta metodologia ser usada no campo para evitar potenciais problemas de queda de desempenho, embora na maior parte dos vídeos da base de dados utilizada não tenha se mostrado necessário.

Síntese do capítulo

Foi mostrada ao longo desse capítulo toda a estrutura utilizada para gerar o algoritmo de detecção e alarme de incêndios D383F, podendo ser divido em 3 etapas principais: segmentação dos pixels por movimento, classificação dos pixels segundo sua cor e a persistência espacial e temporal dos objetos formados por estes pixels. A tabela 3.9 mostra resumidamente as regras utilizadas em cada etapa do algoritmo D383F.
Para a etapa segmentação por movimento foi usado um algoritmo básico de subtração de fundo estimado por médias adaptativas (*running average*) que se baseia em um filtro $ARMA_{(1,1)}$ para a atualização da imagem de fundo, sendo uma técnica simples e eficiente para o uso em ambientes externos com cenas relativamente estáticas. Foi mostrado

Etapa 1	$F_0(x,y) = \begin{cases} 1, & \text{se } \lvert img_{n+1}(x,y) - bg_n(x,y) \rvert > T_s; \\ 0, & \text{caso contrário} \end{cases}$	Segmentação dos pixels por movimento
Etapa 2	$F_1(x,y) = \begin{cases} 1, & \text{se } Y(x,y) > Cb(x,y); \\ 0, & \text{caso contrário} \end{cases}$	Regras baseadas no R>G>B
	$F_2(x,y) = \begin{cases} 1, & \text{se } Cr(x,y) > Cb(x,y); \\ 0, & \text{caso contrário} \end{cases}$	
	$F_3(x,y) = \begin{cases} 1, & \text{se } Y(x,y) > Y_{medio} \text{ e } Cb(x,y) < Cb_{medio} \text{ e } Cr(x,y) > Cr_{medio}; \\ 0, & \text{caso contrário} \end{cases}$	Maior brilho das chamas em relação a outros objetos
	$F_4(x,y) = \begin{cases} 1, & \text{se } \lvert Cb(x,y) - Cr(x,y) \rvert > \tau; \\ 0, & \text{caso contrário} \end{cases}$	Diferença de intensidade da chama nos canais de crominância Cb e Cr
	$F_{5,D383F}(x,y) = \begin{cases} 1, & \text{se } V(x,y) > T_{SV}; \\ 0, & \text{caso contrário} \end{cases}$	Classificador baseado na informação do canal "V" do espaço de cores YUV
	$F_{D383F}(x,y) = \begin{cases} 1, & \text{se } \sum_{i=0}^{5} F_i(x,y) = 6; \\ 0, & \text{caso contrário} \end{cases}$	Regra geral de classificação de pixels por movimento e cor
Etapa 3	$F_{Gi} = \begin{cases} 1, & \text{se } EX_i > T_e; \\ 0, & \text{caso contrário} \end{cases}$	Classificação de um grupo de pixels como foco de incêndio
	$ALARME = \begin{cases} 1, & \text{se } \exists F_{Gi} = 1; \\ 0, & \text{caso contrário} \end{cases}$	Alarme quando um grupo qualquer é classificado como foco de incêndio

Tabela 3.9: Tabela resumo das regras utilizadas em cada etapa do método D383F.

também que as cores são de grande importância na segmentação por movimento das chamas.

A parte de classificação dos pixels por cor foi feita utilizando o método proposto por Çelik e Demirel (2009), no qual uma alteração foi proposta para substituir uma regra que não adicionava informação relevante ao processo de classificação. A alteração consiste no uso de outro classificador baseado no espaço de cor (YUV), no qual a diferença de intensidade das chamas no canal crominância vermelha (canal V) é usado para aumentar a eficiência da classificação por cor.

A persistência espacial e temporal utiliza características bem singulares às chamas de um incêndio, em que o tempo de duração é extenso e sua movimentação em torno do combustível causada pela convecção dos gases é contínua. Como estas características são intrínsecas a um objeto representado por um conjunto de pixels na cena, é necessário realizar um agrupamento dos pixels utilizando um algoritmo de união de conjuntos disjuntos, sendo descartados nesta etapa grupos de tamanhos desprezíveis. A partir desses grupos ou objetos capturados, as informações temporais e espaciais são extraídas através da utilização de uma metodologia baseada em colônia de formigas, onde a idéia de depósito e evaporação de feromônios foi abstraída e utilizada, a fim de eliminar falsos positivos comuns em ambientes externos e melhorar o desempenho do

método D383F.

Finalmente, duas técnicas de redução do custo computacional foram propostas para permitir o uso do algoritmo em tempo real em resoluções maiores que 352 x 288 pixels. A subamostragem foi feita especificamente para resolver o problema de sobre carregamento em resoluções maiores que 352 x 288 pixels, garantindo que apenas a informação necessária seja usada para a detecção do incêndio e o algoritmo continue trabalhando em tempo real. O ajuste automático do limiar de subtração foi feito para evitar a perda de desempenho em casos específicos, nos quais praticamente toda a cena pode estar em movimento e possuir cor de fogo, evitando que o algoritmo fique sobrecarregado durante o agrupamento neste tipo de situação.

Com o método D383F desenvolvido, a próxima etapa consiste em realizar experimentos para avaliar o desempenho do mesmo, sendo este resultado comparado aos resultados obtidos pelos métodos de Çelik e Demirel (2009) e Toreyin et al. (2006).

Capítulo 4

Resultados

Esse capítulo é destinado a mensurar e avaliar os resultados obtidos pelo método D383F através de experimentos, utilizando para isso a base de dados de vídeos enunciada anteriormente na tabela 3.1. Estes experimentos também permitem a avaliação de algumas características funcionais dos métodos.

O método D383F foi desenvolvido em ambiente Windows usando a linguagem C++ em conjunto com biblioteca de visão computacional livre da Intel OpenCV (2013). Outros 3 métodos são utilizados para efeitos de comparação dos resultados obtidos, sendo estes métodos:

- Çelik Puro (CLKP) - Este método se refere à replicação do trabalho de Çelik e Demirel (2009), com as mudanças propostas na seção 3.3, em que a regra F_5 foi alterada para usar uma delimitação da área a partir dos resultados obtidos pelo algoritmo de k-médias, substituindo os polinômios usados pelos autores.
- Çelik com subtração de fundo (CLKS) - A técnica de subtração de fundo foi adicionada ao método CLKP para avaliar o ganho de desempenho relatado em Ko et al. (2010) e Wang et al. (2010). Esta alteração permite que o método de Çelik e Demirel (2009) se torne mais robusto, já que o CLKP, quando comparado a outros métodos na literatura, consiste de apenas uma etapa, que é a de classificação dos pixels por cor.
- Método de Toreyin et al. (2006) (TOR) - Método muito citado na literatura que consegue bons resultados sob diversas condições, sendo comercializado como parte do software VisiFire. Este método em especial permite uma comparação dos resultados obtidos com algo consolidado não só no meio acadêmico, mas também no mercado.

Para avaliar o desempenho e algumas das características do método D383F, foram realizados 3 experimentos distintos utilizando os diferentes conjuntos de vídeos que formam a base de dados da tabela 3.1:

- Experimento 1 - Neste experimento todos os métodos citados acima foram aplicados ao conjunto de vídeos VideoRes, que consiste primariamente de vídeos de áreas florestais. Este experimento visa avaliar os resultados de desempenho obtidos pelos métodos não só na tarefa de classificação, mas também com

relação à capacidade dos mesmos em realizar a tarefa em tempo real sob diversas resoluções, já que existe uma tendência observada de se utilizar câmeras de alta resolução para a monitoração ambiental, permitindo que as pessoas obtenham mais informações sobre o ambiente monitorado. Foram usadas como medidas de desempenho: a razão de verdadeiros positivos ou revocação (equação 3.3), a razão de falsos positivos ou alarmes falsos (equação 3.6), a acurácia (equação 3.2), a medida F (equação 3.4) e o tempo médio de processamento para cada vídeo.

- Experimento 2 - Neste experimento, os métodos D383F e TOR são analisados não só dentro do contexto de incêndios florestais, mas também em situações diversas como: chamas criadas pela ação do homem com diferentes materiais combustíveis, qualidade de imagem e ambientes urbanos (conjuntos de vídeos VideoDiV e VideoDivCtrl). Aqui, o uso da técnica de persistência elaborada para o método D383F foi feita de forma a controlar melhor a taxa de falsos positivos causados pela variação do parâmetro τ referente à regra F_4, em que um valor menor deve ser usado para que vídeos com chamas capturadas de cor amarelo-clara possam ser classificadas. Neste experimento temos a medição do desempenho por: a acurácia, a medida F, sendo mostrado também o número de verdadeiros positivos e alarmes falsos referentes aos quadros classificados com fogo para avaliar o custo em quadros na eliminação dos falsos positivos pela persistência.

- Experimento 3 - Este último experimento avalia a sensibilidade do método proposto para detectar chamas geradas a partir de combustíveis fósseis em diferentes distâncias (conjunto VideoDis). Estas chamas possuem maior energia e são capturadas pelas câmeras geralmente próximas do limite de saturação (cor branca), sendo possível mostrar como o comportamento do algoritmo pode ser ajustado para um contexto completamente diferente de monitoração. Neste experimento não há amostras da classe negativa nos vídeos utilizados, assim é usado como medida de desempenho a revocação (equação 3.3) ou razão de verdadeiros positivos.

Todos estes experimentos foram realizados utilizando o mesmo computador com um processador AMD $FX^{TM} - 8320$ de oito núcleos de 3.5 GHz com 8GB de memória RAM.

Com esse breve resumo de cada experimento, as próximas seções apresentarão os experimentos, sendo realizada a análise e comparação dos resultados de desempenho entre os diferentes métodos.

4.1 Resultados do experimento 1

Para o primeiro experimento, foi usado o conjunto de vídeos VideoRes. Este conjunto contém 10 vídeos de 10 segundos a 10 quadros por segundo em 8 resoluções diferentes, sendo 5 vídeos com fogo e 5 vídeos de controle sem fogo num total de 500 quadros com fogo e 500 sem fogo (vide seção 3.1). Os vídeos apresentam características diversas como: mudanças de iluminação, período do dia, chamas com pouca diferença de cor

em sua extensão e vídeos de controle que tentam enganar os métodos com diferentes situações em que objetos com cor de fogo simulam algumas características do fogo.

Este experimento busca avaliar o desempenho dos diferentes métodos dentro do contexto de incêndios florestais e também se os algoritmos conseguem executar sua tarefa em tempo real em diferentes resoluções.

Os parâmetros usados neste experimento para cada método são mostrados a seguir na tabela 4.1. A definição de tais parâmetros pode ser encontrada no capítulo 3.

Parâmetro	**Valor**	**Métodos aplicáveis**
τ Limiar da regra F_4	80 para todas as resoluções conforme descrito anteriormente na seção 3.3	CLKP, CLKS e D383F
T_s Limiar da subtração de fundo	[35, 35, 32, 31, 29, 27, 25 e 16] (valores por resolução da maior para menor respectivamente), definido experimentalmente.	CLKS e D383F
T_{Gmin} Limiar de tamanho mínimo do grupo	5 pixels, conforme Çelik et al. (2007) definem para uma resolução próxima de 160 x 120. Definido linearmente para as diferentes resoluções pela equação 3.26.	D383F
T_{dG} Limiar de distância mínima entre grupos	$100 \leq T_{dG} \leq 500$ (distância euclidiana quadrada da posição espacial entre pixels, equivalente de 10 a 22 pixels). Varia em função da resolução, definido experimentalmente, calculado linearmente para as resoluções pela equação 3.25	D383F
T_{mG} Limiar de movimento de um grupo	2,5 vezes o valor de T_{Gmin}. Suas dependências são análogas ao T_{dG}, sendo definido experimentalmente.	D383F
T_e Limiar de persistência temporal e espacial	10 quadros (1 segundo) para todos os testes. Definido experimentalmente.	D383F
S_T Limiar de sensibilidade para a detecção de fogo	Valor 3 para todos os experimentos a partir de testes realizados em vídeos e do resultado apresentado em Toreyin (2013), onde $S_T \in (1, 2, 3)$	TOR

Tabela 4.1: Tabela de valores para os parâmetros referentes a cada método.

As figuras 4.1, 4.2, 4.3, 4.4 e 4.5 mostram respectivamente os resultados da razão de verdadeiros positivos, razão de falsos positivos, acurácia, medida F e tempo médio gasto por vídeo obtidos no conjunto de vídeos VideoRes.

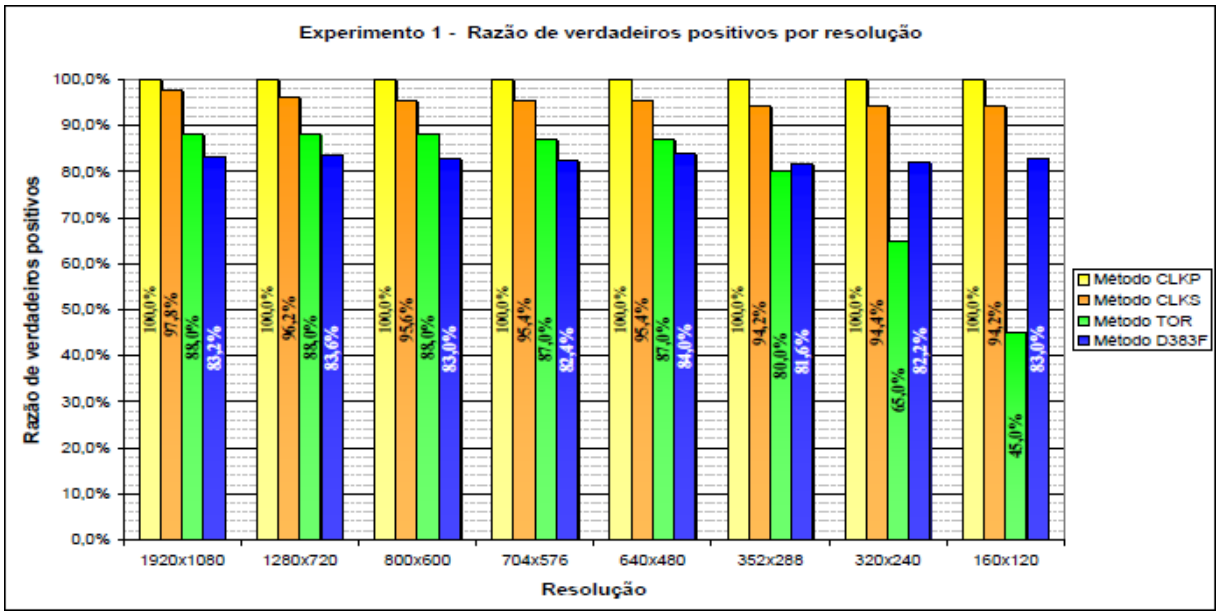


Figura 4.1: Razão de verdadeiros positivos no experimento 1 em 500 quadros de vídeo com fogo.

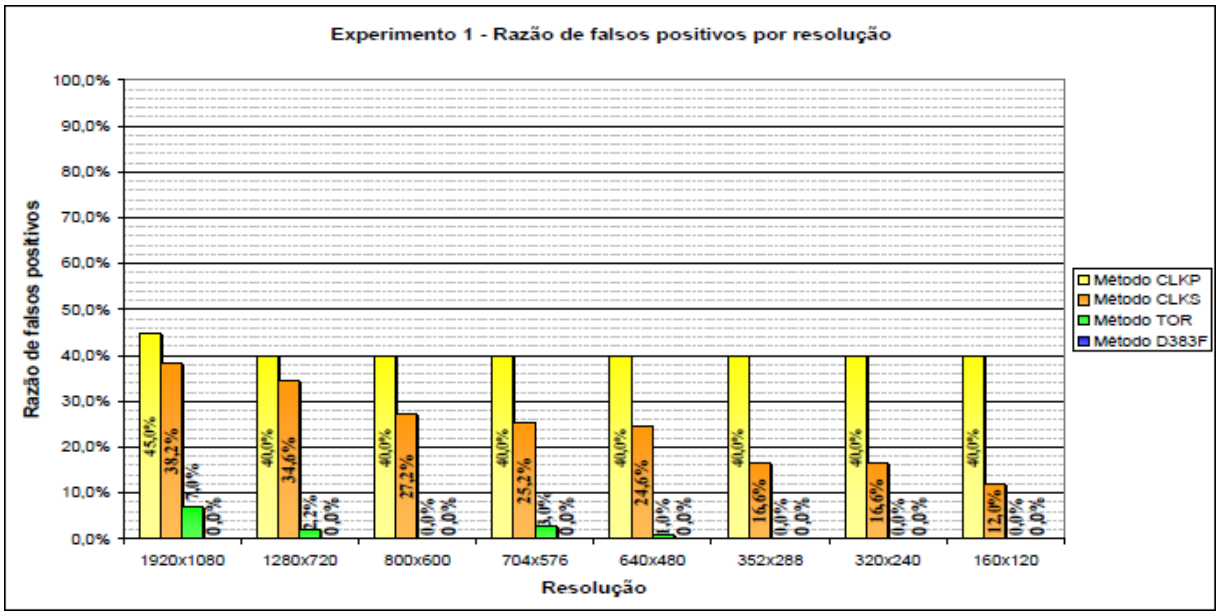


Figura 4.2: Razão de falsos positivos no experimento 1 em 500 quadros de vídeo sem fogo.

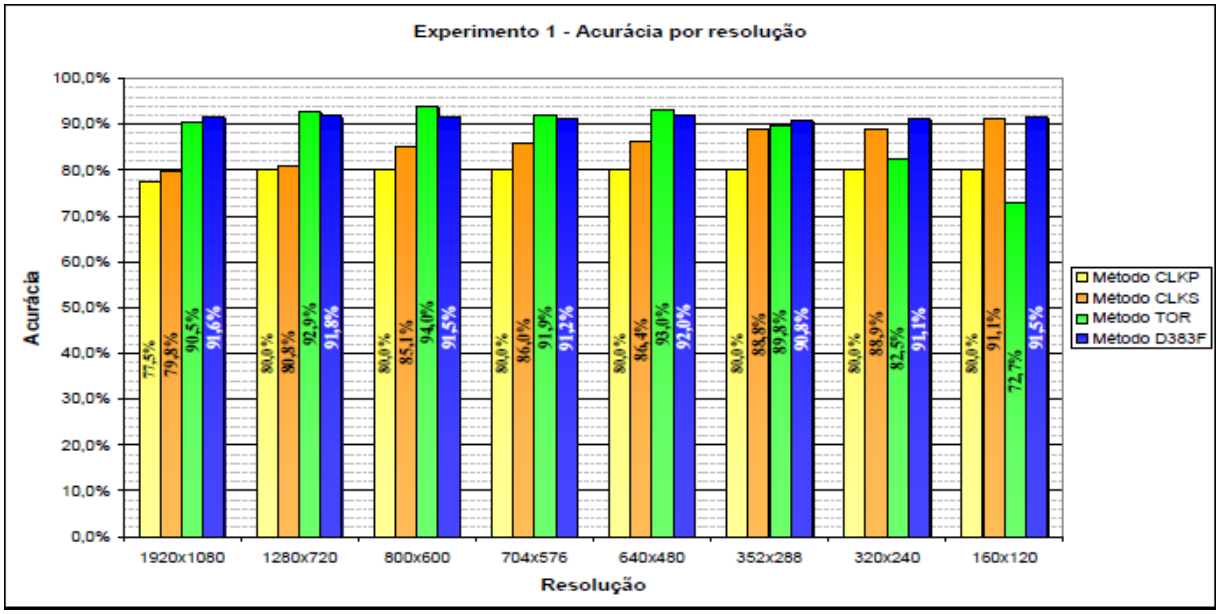


Figura 4.3: Acurácia dos métodos no experimento 1

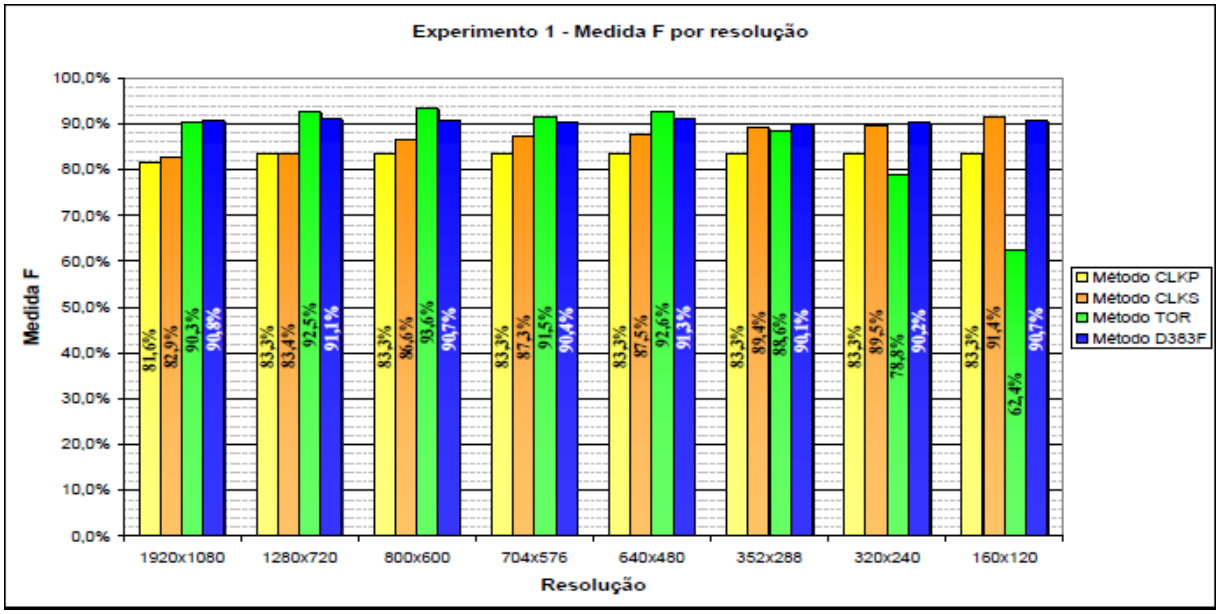


Figura 4.4: Medida F dos métodos no experimento 1

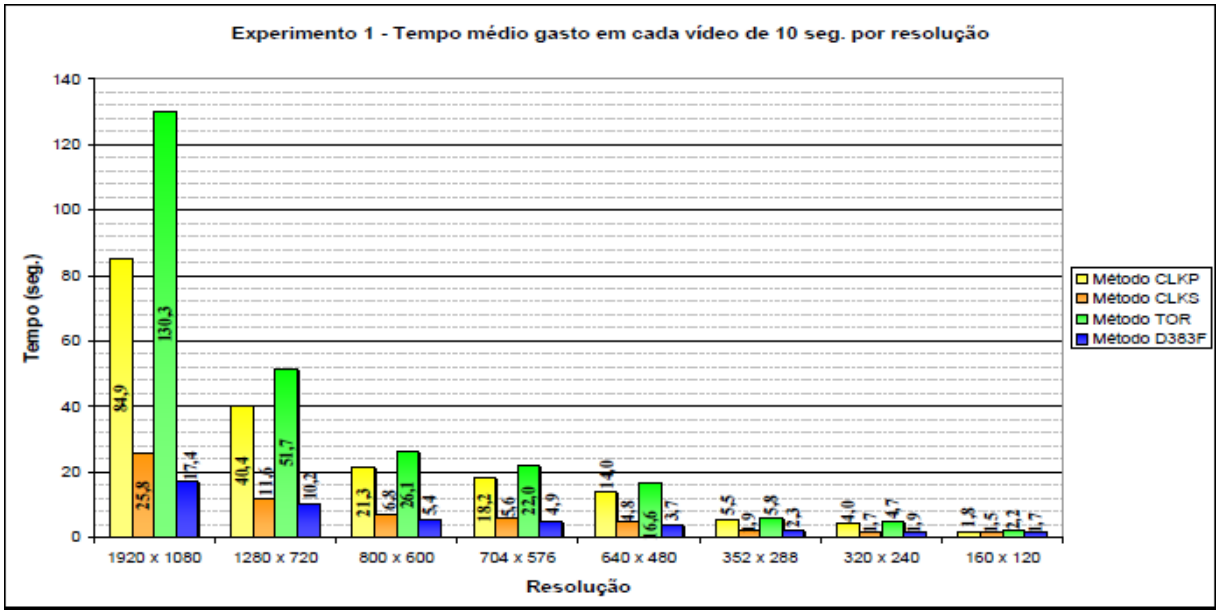


Figura 4.5: Tempo médio de processamento de vídeo dos métodos no experimento 1

Pelos resultados apresentados observa-se que os valores obtidos pelo método CLKP nas figuras 4.1 e 4.2, no qual nessa base relativamente difícil ele consegue 100% de razão de verdadeiros positivos e 40% de razão de falsos positivos respectivamente, estão próximos com o relatado em seu trabalho (99% de razão de verdadeiros positivos e 31.5% de falsos positivos), mostrando que a alteração proposta em sua quinta regra não afetou o desempenho de classificação do método de maneira significativa.

Nas figuras 4.2, 4.3 e 4.4, podemos ver que, com o uso da subtração de fundo no método CLKS foi possível reduzir consideravelmente a razão de falsos positivos detectados e conseqüentemente aumentar a acurácia e medida F do método, superando CLKP em todas as resoluções. É possível observar também que a diferença de desempenho de classificação entre CLKP e CLKS aumenta à medida que a resolução diminui. Isto acontece porque em resoluções mais baixas a quantidade de informação ou número de pixels para detectar o movimento dos objetos é diminuída e isso acaba por ajudar a eliminar os falsos positivos na fase de segmentação de movimento, uma vez que os objetos que geram os falsos positivos geralmente possuem uma magnitude de tamanho e movimento menor que a do fogo. Outro fator importante que pode ser observado na figura 4.5, é o impacto da subtração de fundo no tempo de processamento médio de cada vídeo para CLKS, que reduziu drasticamente em relação a CLKP, mostrando a força da técnica para problema em questão conforme Ko et al. (2010) e Wang et al. (2010) afirmam em seus trabalhos.

O método TOR obteve um desempenho muito bom em grande parte das resoluções, como pode ser visto nas figuras 4.3 e 4.4, ficando na maioria das vezes com as melhores acurácia e medida F. Esse resultado era esperado com base nos relatos na literatura em geral e corrobora por sua adoção como sistema comercial. No entanto, em resoluções menores que 352 x 288 pixels os seus resultados sofreram um queda significativa. Isso se deve em parte à diminuição da quantidade de informação disponível para a

classificação, principalmente no caso dos vídeos VR 1 e VR 4. Em VR 1 a chama apresenta uma cor muito uniforme, similar a um objeto de cor sólida. Como o algoritmo de Toreyin usa a transformada de wavelet para avaliar a variação espacial das cores da chama e discernir a mesma de um objeto sólido, esse tipo de chama apresentado em VR 1 acaba dificultando a detecção pelo algoritmo. Em VR 4, o incêndio se encontra muito distante e à noite. Novamente não há informações suficientes para que a transformada de wavelet possa detectar com precisão o foco de incêndio. Nesse vídeo, os focos de incêndio ficam apenas em situação de observação, não sendo classificados como um incêndio pelo algoritmo.

Esse tipo de comportamento indica que a sensibilidade do método TOR está baixa para esses tipos de situações, que podem facilmente ocorrer dentro do contexto proposto de monitoração ambiental.

Fica evidente que os problemas de desempenho de classificação acima são facilmente resolvidos com o aumento da resolução conforme os próprios resultados mostram. Entretanto, os desenvolvedores não indicam a utilização de resoluções superiores a 320 x 240 pixels. Com base nos resultados obtidos na figura 4.2, é possível observar que, embora o algoritmo se saia muito bem em resoluções superiores à recomendada, o mesmo começa a detectar alguns falsos positivos nos vídeos de controle, e principalmente, ao observarmos a figura 4.5, vemos que TOR começa a exigir um tempo de processamento maior que o disponível para a execução em tempo real da detecção e classificação, devido ao aumento significativo na área que corresponde ao corpo da chama. Fica subentendido a partir desses resultados que os desenvolvedores preferem manter seu algoritmo usando uma resolução mais baixa com uma sensibilidade baixa, evitando grande parte dos falsos positivos em diversas situações. Ainda que isso resulte em uma restrição na detecção de incêndios em longas distâncias na resolução de 320 x 240 pixels, isso o mantém funcionando em tempo real com uma taxa de amostragem alta o suficiente para que a cintilância do fogo possa ser captada eficientemente, segundo eles mesmos afirmam em Toreyin et al. (2006).

O parâmetro S_T (veja tabela 4.1) poderia ser mudado para um valor maior de sensibilidade (S_T=2) de forma a resolver o problema de detecção para chamas em longas distâncias citado acima, mas ao realizar essa ação para este experimento, isso acabou acarretando falsos positivos constantes nos vídeos de controle, não havendo ganho algum para o algoritmo TOR no desempenho de classificação.

Continuando a análise dos resultados, o método D383F se mostra superior a todos os métodos em acurácia e medida F em todas as resoluções menores que 640 x 480 pixels (salvo uma exceção na medida F na resolução de 160 x 120 pixels em que CLKS apresenta uma leve vantagem), dista menos de 3% do melhor método (TOR) nas resoluções intermediárias e é superior também na resolução de 1920 x 1080 pixels. É necessário mencionar o fato de todos os falsos positivos serem eliminados, conforme se vê na figura 4.2, no qual o algoritmo cumpre umas das metas para o qual foi projetado, que é diminuir significativamente o número de falsos positivos. Em contrapartida, pode-se observar na figura 4.3 que a razão de verdadeiros positivos detectados pelo método

D383F é sempre menor que todos os outros na maior parte dos casos. Isto ocorre devido ao custo da persistência (T_p = 10) em que pelo menos 50 quadros (10 por vídeo) são gastos para executar a classificação de um foco de incêndio.

Embora o método D383F perca por pouco (<3%) em desempenho de classificação na acurácia e medida F para TOR em resoluções maiores ou iguais a 640 x 480 até 1280 x 720 pixels, ele se mostra superior quando observado do ponto de vista de tempo de processamento médio de cada vídeo (figura 4.5), conseguindo nestas resoluções o melhor tempo e superando o método TOR brutalmente neste quesito. O algoritmo D383F é capaz de processar imagens até 1280 x 720 pixels em tempo real, bem como ser usado com a resolução de 1920 x 1080 pixels a uma taxa de 5 quadros por segundo.

Outra observação importante a ser feita é o fato que esse conjunto de vídeos coloca o método D383F no limite de seu funcionamento, uma vez que o tempo de persistência (T_e = 10 quadros ou 1 segundo) para essa base foi demasiadamente pequeno, ficando no limite de detectar o movimento de folhas e galhos balançando em movimento oscilatório ao vento. Em condições reais de uso, esse tempo poderia se estender a minutos, eliminando praticamente todos os falsos positivos causados por esses objetos na cena e garantindo que um alarme de incêndio seja relatado de forma mais precisa ao custo de algumas dezenas ou centenas de quadros para a persistência temporal. Ainda ao observar as figuras 4.3 e 4.4 com atenção é possível ver que a acurácia e a medida F do método D383F se mantêm constantes e acima de um patamar alto (> 90%) para todas as resoluções testadas, mesmo sendo feita a sub amostragem de pixels nas resoluções maiores que 352 x 288 pixels.

Nessa técnica de sub amostragem, para as resoluções de 640 x 480 a 1280 x 720 pixels, apenas os pixels das colunas ímpares (referente ao eixo de largura da imagem) são avaliados e na resolução de 1920 x 1080, apenas os pixels das colunas e linhas ímpares são avaliados (equação 3.32). Este tipo de abordagem objetiva diminuir o tempo computacional gasto ao máximo utilizando o mínimo de informação possível. Fica claro nesse caso que muitos pixels estão sendo descartados nesta técnica, e isto poderia estar acarretando em um desempenho de classificação mais pobre do método D383F. Assim, um teste piloto foi realizado em todos os vídeos na resolução de 1920 x 1080 pixels (nesta resolução 3/4 dos pixels são descartados), sendo tal procedimento comparado com uma técnica auxiliar de sub amostragem em que a média dos 4 pixels vizinhos (coordenada ímpar e os 3 pixels vizinhos que não são utilizados pela sub-amostragem original) é calculada e esse valor médio é então usado para a classificação.

A representação das duas técnicas pode ser vista na figura 4.6, em que uma imagem hipotética com 64 pixels de tamanho é usada para mostrar a informação utilizada da imagem. No caso, é possível ver na figura 4.6A, os pixels de coordenadas ímpares utilizados pelo algoritmo D383F, que correspondem a 1/4 da imagem total, já na figura 4.6B, um pixel a ser utilizado corresponde a média aritimética da informação contida nos 4 pixels vizinhos, utilizando toda a informação presente na imagem.

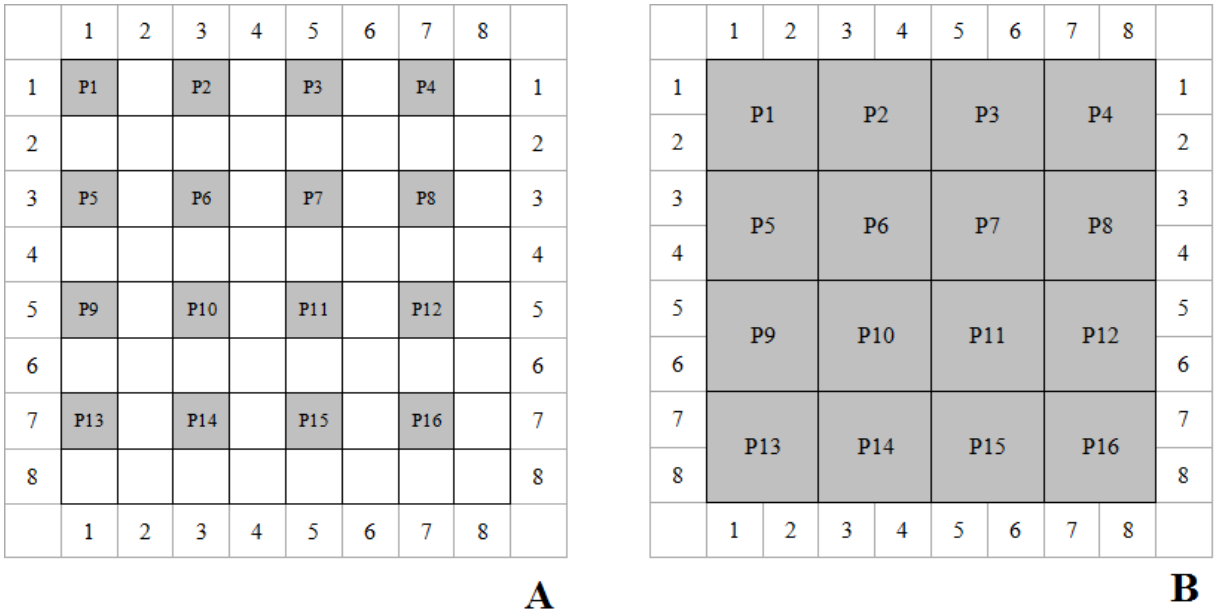


Figura 4.6: Pixels utilizados nas técnicas de sub-amostragem. (A) Técnica de coordenadas ímpares: Pixels utilizados na classificação correspondem a apenas 1/4 da informação; (B) Técnica de média dos 4 vizinhos: Um pixel a ser utilizado corresponde à média da informação contida nos 4 pixels vizinhos, utilizando toda a informação presente na imagem.

O teste piloto usando estas duas técnicas de sub-amostragem permite avaliar se a informação dos pixels descartados tem influência significativa sobre o desempenho de classificação obtido pelo método D383F, sendo o resultado deste teste mostrado na tabela 4.2.

Sub-amostragem	**Razão de verdadeiros positivos**	**Razão de falsos positivos**	**Acurácia**	**Medida F**	**Tempo médio (s)**
Coordenadas ímpares	83,2%	0	91.6%	90.8%	17.3
Média dos 4 pixels vizinhos	83,2%	0	91.6%	90.8%	18

Tabela 4.2: Tabela de resultados do teste piloto para a técnica de sub amostragem.

Conforme se vê na tabela 4.2, os resultados não se alteraram, e como o esperado, o tempo médio de processamento foi um pouco maior quando o cálculo de médias foi utilizado, em especial nos vídeos VR 1 e VR 2, nos quais há uma intensa movimentação na cena ocasionada pela fumaça. O tempo gasto nesses vídeos usando a média dos 4 pixels foi de 1.6 e 2 segundos respectivamente a mais que quando se usam apenas as coordenadas ímpares. Com base nesse resultado, pode-se dizer que a perda de informação relacionada à classificação alvo é praticamente irrelevante ao se utilizar a técnica de sub-amostragem por coordenadas ímpares, pois mesmo com o descarte de grande parte dos pixels, as chamas ainda possuem magnitude suficiente para continuarem sendo detectadas pelo método sem nenhum problema, sendo esta técnica mais barata e simples computacionalmente.

Outro fator que deve ser observado e que foi mencionado anteriormente no trabalho, é que computacionalmente, uma resolução maior não afeta os algoritmos da forma esperada conforme se vê nas figuras 4.3 e 4.4, ou seja, melhorando seu desempenho de classificação. Exceto para o caso de Toreyin, que em resoluções maiores de 360 x 240 pixels, seu desempenho de detecção aumenta devido à quantidade maior de informação para a detecção da cintilância da chama, entretanto isso leva o algoritmo a um tempo de processamento maior que o disponível para o processamento em tempo real. Como os algoritmos trabalham apenas com características "primitivas", as quais são muito significativas no objeto de interesse (fogo), o aumento de resolução não leva a um aumento linear da significância destas características, mas leva a um aumento destas características em outros objetos que podem ser confundidos com fogo, aumentando o número de falsos positivos e afetando o desempenho do algoritmo, conforme pode ser visto no método CLKS e em menor relevância no TOR.

Em suma, pode-se dizer pelos resultados obtidos para este experimento que:

- A subtração de fundo realmente ajuda na classificação e detecção de eventos em vídeo reduzindo significativamente o custo computacional;
- O método TOR é robusto como esperado, conseguindo melhor desempenho em grande parte dos testes, salvo o tempo computacional;
- O método D383F supera todos os outros métodos em pelo menos uma medida, perdendo por pouca diferença nas medidas não superadas, mantendo um desempenho estável acima de 90% em acurácia e medida F e se mostrando tão robusto quanto o método TOR, que é usado em uma aplicação comercial e muito citado na literatura. Além disso, conseguiu tratar de forma efetiva os vídeos de controle, evitando a captura de falsos positivos. Outro fato importante a ser levado em conta é sua capacidade de trabalhar com resoluções superiores a 320 x 240 pixels em tempo real sem sacrificar significativamente desempenho na detecção e classificação de focos de incêndios.

4.2 Resultados do experimento 2

Para o segundo experimento, os conjuntos de vídeos VideoDiv e VideoDivCtrl foram utilizados, somando um total de 30 vídeos que representam diversas situações dentro e fora do contexto florestal para avaliar o comportamento e robustez dos métodos nestas situações. Esses vídeos estão todos na resolução de 320 x 240 pixels a 10 quadros por segundo, sendo 10 deles obtidos da base de dados disponibilizada por Toreyin (2013) e outros 4 pelo grupo CAVIAR. O conjunto VideoDiv possui 4330 quadros de vídeo com fogo e o VideoDivCtrl possui 4550 quadros de vídeo sem fogo. Os vídeos apresentam, como no primeiro experimento, diversas características como: mudanças de iluminação, período do dia, chamas produzidas por outros materiais combustíveis com intervenção humana, locais com constante circulação de pessoas, etc. Detalhes dos vídeos podem ser vistos na seção 3.1.

Com base no que foi analisado no experimento anterior, neste experimento apenas

o método D383F e o método TOR são utilizados, isto devido a seus melhores resultados, e também porque o valor de τ foi reduzido de forma a capturar chamas que estão mais próximas do limite de saturação da câmera, como as chamas dos vídeos VD 1, 2, 3 e 4. Caso os métodos CLKP e CLKS fossem utilizados, conforme explicado e exemplificado na seção 3.3 com a figura 3.11 e a figura 3.16, um grande número de falsos positivos seria contabilizado para estes métodos devido a esse valor mais baixo de τ.

Usando esta característica gerada pelo parâmetro τ, o funcionamento do parâmetro T_e referente a persistência temporal e espacial no método D383F é mostrado neste experimento, no qual τ foi usado com valor igual a 20 e T_e com os valores de 10 e 30 quadros. Os parâmetros restantes tiveram seus valores mantidos, conforme mostrado na tabela 4.1 do experimento 1.

Não foi considerada a medida de tempo médio de processamento dos vídeos, uma vez que no experimento anterior já ficou evidente que o método TOR é o mais caro computacionalmente (vide figura 4.5) e que ele é perfeitamente capaz de processar vídeos na resolução de 320 x 240 pixels em tempo real.

Os resultados obtidos para este experimento podem ser vistos nas figuras 4.7, 4.8, 4.9 e 4.10.

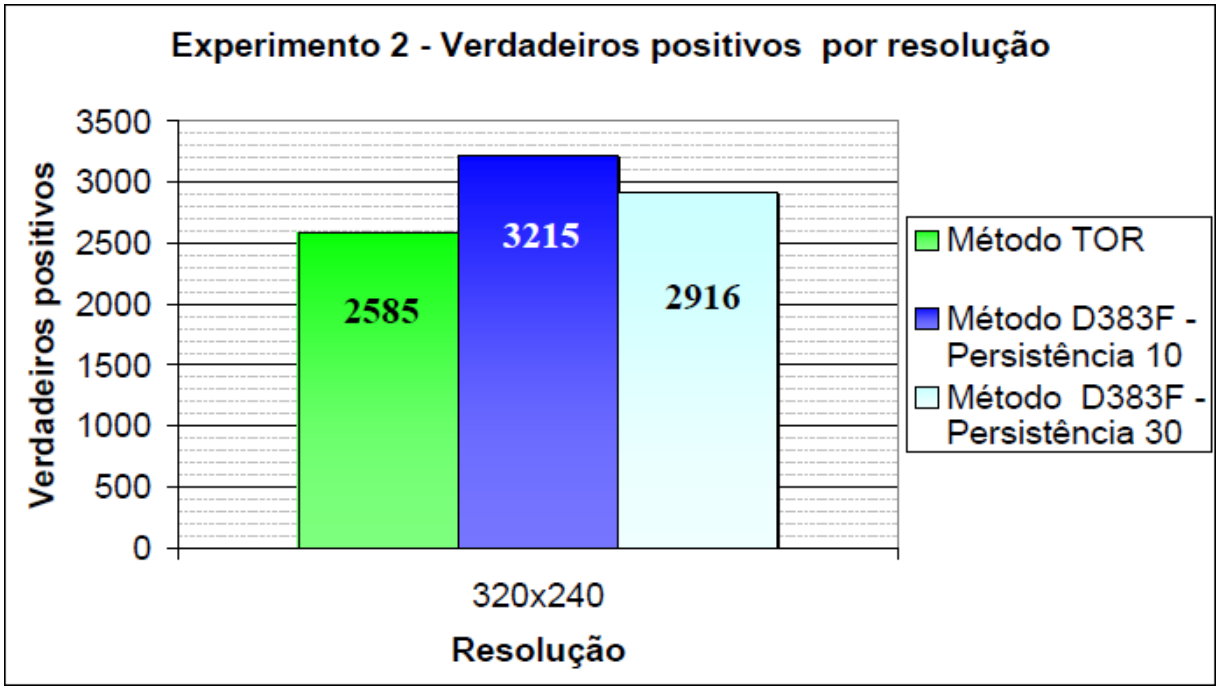


Figura 4.7: Número de verdadeiros positivos no experimento 2 para um total de 4330 quadros de vídeo com fogo.

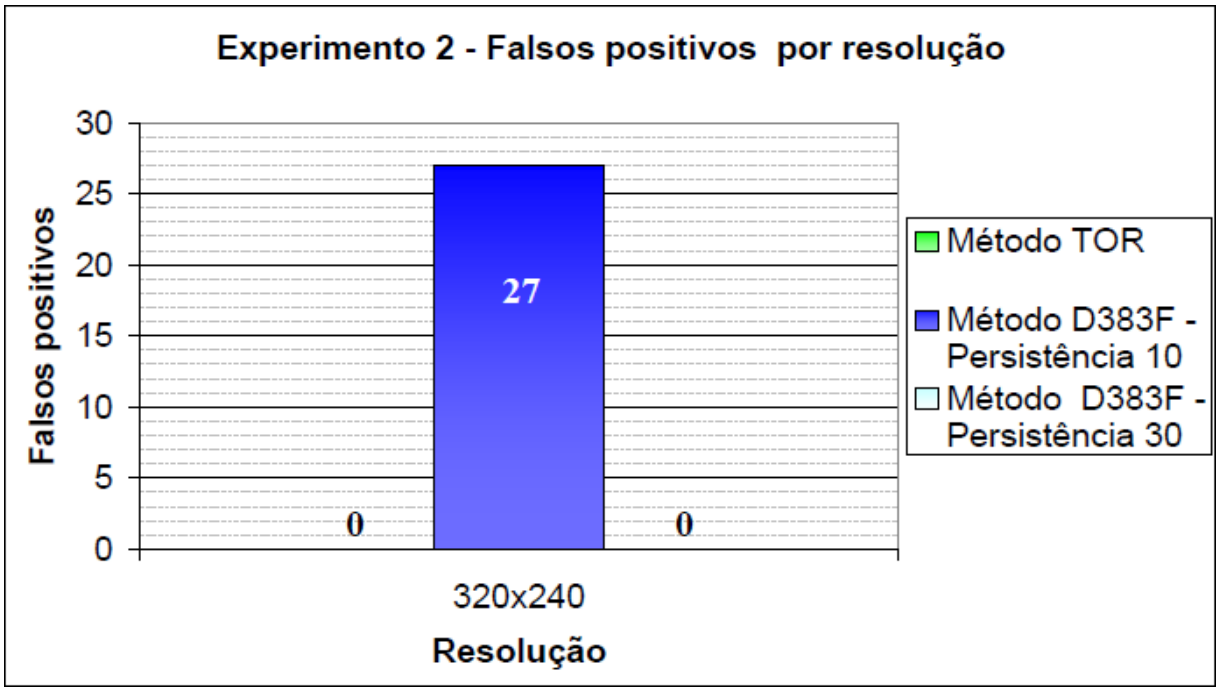


Figura 4.8: Número de falsos positivos no experimento 2 para um total de 4550 quadros de vídeo sem fogo.

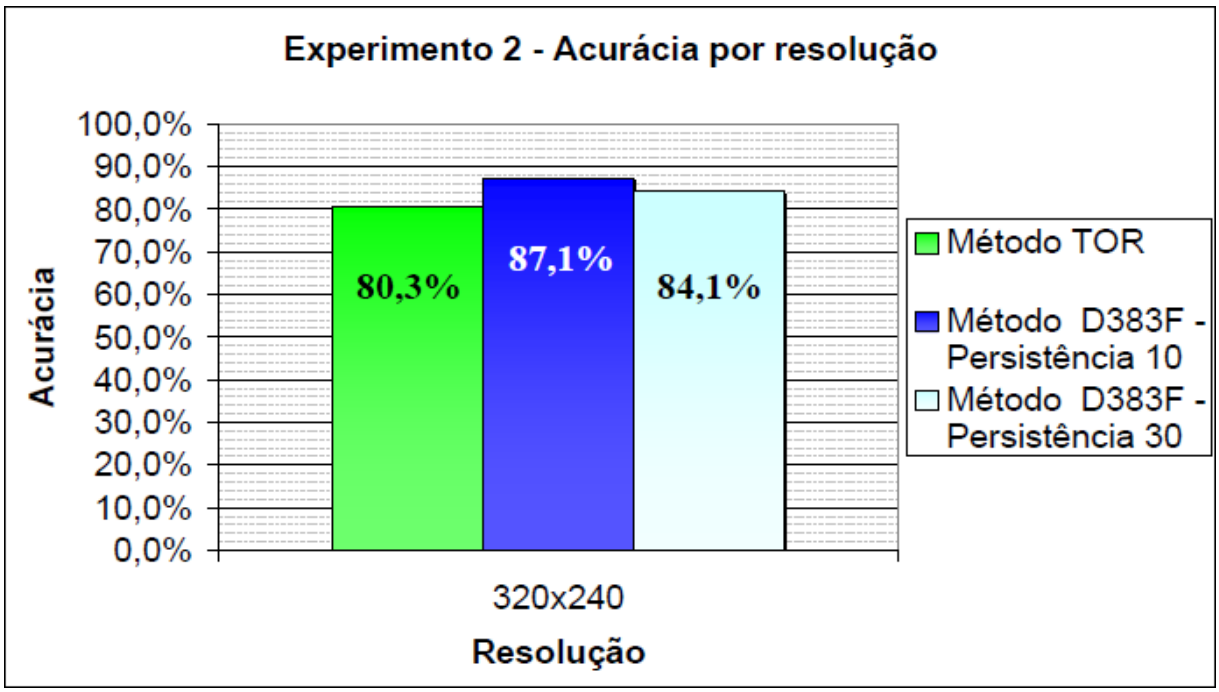


Figura 4.9: Acurácia dos métodos no experimento 2

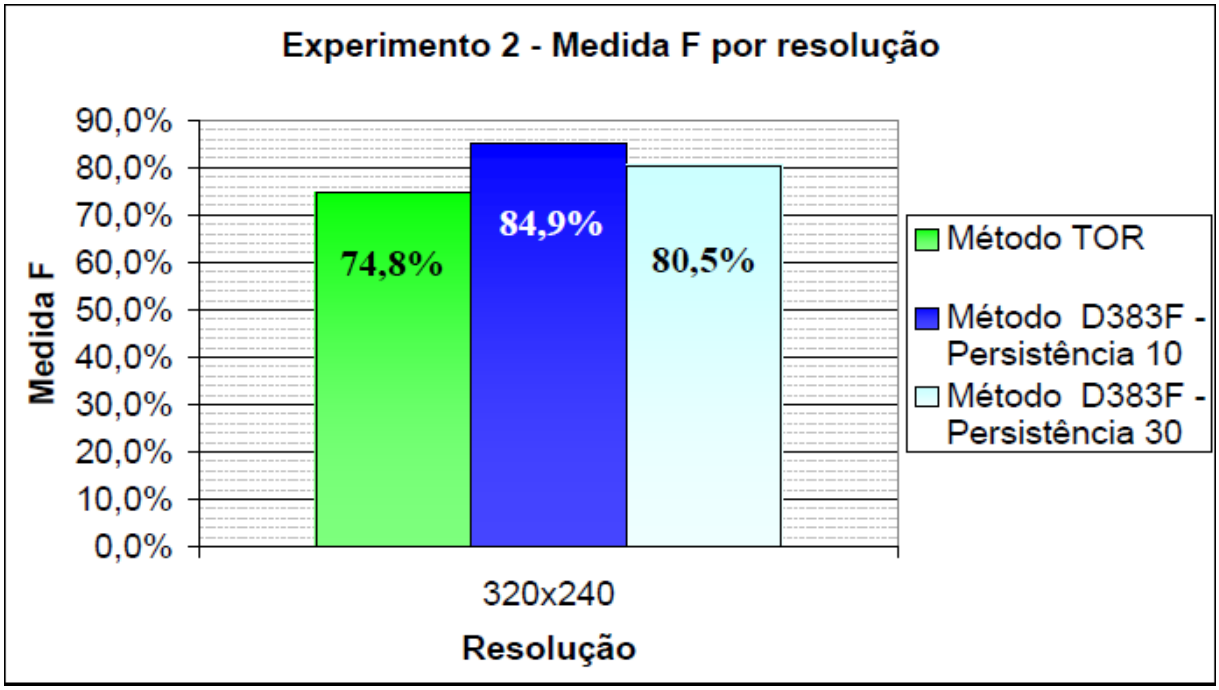


Figura 4.10: Medida F dos métodos no experimento 2

Pelos resultados mostrados podemos ver que o método D383F se sai melhor no geral que método TOR tanto com T_e com o valor 10 ou 30. No entanto, como esperado para o valor de $\tau = 20$, alguns falsos positivos foram obtidos quando $T_e = 10$ em alguns dos vídeos de controle, sendo um caso especial o vídeo VDC 2, em que uma pessoa com um decalque laranja na camisa caminhando normalmente pelo saguão passa na área com incidência de luz solar. Aqui poderia ser indagado porque o parâmetro T_{mG} (persistência espacial) não foi capaz de filtrar esse falso positivo, e a resposta reside no fato que o vídeo tem uma câmera em uma posição bem alta, e portanto, distante do alvo, e que o parâmetro T_{mG} está definido de forma a tratar o caso mais geral. Um valor de T_{mG} para tratar este caso específico seria menor que 5 pixels de distância. Esse valor baixo prejudicaria outros vídeos como o VD 2 que tem a chama bem próxima da câmera e se movimenta por quase todo o quadro. Os outros falsos positivos foram obtidos nos vídeos VDC 12 e 14, com um carrinho de compras e um carro respectivamente.

Todos esses eventos têm algo em comum, são objetos com cor de fogo se movimentando relativamente devagar quando avaliados em relação ao restante dos vídeos. Esse tipo falso positivo é comum até em detectores de movimento, nos quais uma pessoa se movendo suficientemente devagar consegue enganar o sistema. A característica mais importante com relação a esses eventos está no fato que são eventos de curta duração, ou seja, não duram mais que dois segundos, sendo perfeitamente possível eliminar a maioria desses falsos positivos ao levar em conta que uma chama tem uma duração significativamente maior que esses eventos. Assim, trocando o valor de T_e para 30, o que equivale a um tempo de 3 segundos, todos esses falsos positivos são eliminados à custa de um atraso maior no número de quadros de existência (20 quadros ou 2 segundos em relação a $T_e = 10$) para que o alarme de detecção de chamas seja emitido.

Em casos de monitoração real, esse valor pode ser estendido a minutos, eliminando grande parte dos falsos positivos conforme mencionado no experimento anterior.

O cálculo do número aproximado de quadros gastos a mais na detecção de chamas (verdadeiros positivos não considerados nos vídeos com fogo - figura 4.7) pode ser facilmente realizado multiplicando a diferença de T_e nos dois testes ($30 - 10 = 20$) pelo número de vídeos com fogo ($20 \times 15 = 300$). Utilizando os valores da figura 4.7 referentes aos testes com os diferentes valores de T_e para o método D383F, a diferença é $3215 - 2916 = 299$, muito próximo ao esperado. E mesmo com esse valor expressivo, o método D383F ainda se mantém com um desempenho de classificação melhor que TOR, embora ele não se saia melhor em todos os vídeos do conjunto. Em vídeos de média e curta duração (vídeos com menos de 300 quadros), como VD 4, 5 e 6, o método D383F leva uma desvantagem considerável no número de verdadeiros positivos detectados para $T_e = 30$, perdendo por uma média nestes vídeos específicos de 23 quadros aproximadamente. Mas leva uma vantagem considerável quando os vídeos tem uma duração maior, apresentam chamas saturadas como VD 2, 3 e 15 e chamas que ficam bem pequenas como é o caso de VD 1, 2 e 3, evidenciando o que foi dito no experimento anterior sobre a sensibilidade do método TOR. Fica evidente que em casos de funcionamento contínuo do algoritmo D383F, a desvantagem causada no número de verdadeiros positivos pela persistência espacial e temporal tende a desaparecer.

Ainda nesse experimento, outro fato importante que deve ser avaliado é a diferença significativa existente entre os resultados obtidos para a acurácia e medida F, conforme pode ser vista nas figuras 4.9 e 4.10. A diferença de desempenho de classificação na medida F entre TOR e o método D383F com $T_e = 10$ é mais de 10 pontos percentuais (74,8% para TOR e 84,9% para o D383F) contra menos de 7 pontos percentuais na acurácia (80,3% para TOR e 87,1% para D383F. Esta diferença de valores de desempenho entre a acurácia e a medida F se deve ao fato da classe fogo ser ligeiramente desbalanceada neste experimento, e a medida F mostra que o método D383F foi bem mais eficiente na avaliação da classe minoritária fogo do que TOR.

Mesmo nesta base de dados em que vídeos fora do contexto florestal são usados, os resultados obtidos pelos métodos estão próximos do que foi obtido no experimento anterior, mostrando que os mesmos são robustos para diferentes situações.

4.3 Resultados do experimento 3

O último experimento utiliza o conjunto de vídeos VideoDis que possui 3 vídeos da base de dados disponibilizada pelo grupo CVPR Lab. Os vídeos são parte de um experimento controlado em que uma chama produzida pela queima de querosene é colocada em diferentes distâncias da câmera. Estes vídeos estão na resolução de 320 x 240 pixels e têm uma taxa de 24 quadros por segundo, em um total de 15263 quadros que possuem chamas.

Nesse conjunto de vídeos não há amostras da classe negativa, sendo utilizado como

medida de desempenho apenas a revocação (equação 3.3), com cada vídeo avaliado separadamente para avaliar o quanto a distância influencia nos resultados.

Os parâmetros são os mesmos usados no primeiro experimento, em conjunto com as alterações feitas no segundo experimento com $\tau = 20$ e $T_e = 30$ para o método D383F.

Este experimento busca avaliar o uso dos métodos para um tipo de chama diferente, colocando sua capacidade de detecção à prova tanto no quesito de tipos de chama como distância máxima para a detecção. Os conceitos relatados no capítulo anterior que possibilitam tratar as bordas das chamas como uma região de importância são primordiais aqui, já que a chama capturada pela câmera nestes experimentos está saturada (branca) e não possui cor de fogo, mas sua borda que é mais fria consegue gerar cores amarelo-alaranjadas e, por causa desse efeito, podem ser detectadas pelo método D383F. A figura 4.11 mostra marcações retangulares nas imagens das bordas da chama capturadas em cada um dos testes de distância para o método D383F.



Figura 4.11: Bordas capturadas de uma chama saturada em diferentes distâncias. (A) VDT 1 - Chama a 10 metros; (B) VDT 2 - Chama a 20 metros; (C) VDT 3 - Chama a 30 metros;

A tabela 4.3 mostra os resultados da revocação ou razão de verdadeiros positivos obtidos por vídeo pelos métodos D383F e TOR.

		Vídeos (N° total de quadros do vídeo)		
		VDT 1 (4478)	**VDT 2 (3879)**	**VDT 3 (6906)**
Método	**D383F**	31,9%	42%	48,2%
	TOR	1,1%	0%	0%

Tabela 4.3: Tabela de resultados com a razão de verdadeiros positivos obtida para cada vídeo do experimento de distância.

No caso, vê-se nos resultados que à medida que a chama se afasta, melhor é o desempenho do método D383F. Esse resultado é causado pela menor densidade de fluxo luminoso que atinge a câmera em função da distância da chama, ou seja, quanto mais longe a fonte de luz, menos raios de luz chegam aos receptores da câmera, diminuindo a saturação destes na área da chama. Isso permite à câmera capturar as cores de menor intensidade de brilho nas bordas das chamas que possuem a cor amarelo-alaranjada, melhorando o desempenho do algoritmo nesse caso específico.

O resultado obtido para o método TOR no vídeo VDT 1 não indica uma detecção real na chama em si, a qual ele se mostrou incapaz de classificar. Isto se deve em grande parte ao fato que a chama está tão próxima da saturação nestes vídeos que a análise por transformada de wavelet não consegue verificar a variação de cores entre os pixels que formam a chama. No entanto, ele conseguiu captar o reflexo das chamas no concreto e classificar 50 quadros como possuindo fogo. É normal pensar que este tipo de comportamento indique um falso positivo, mas não é bem assim, Liu et al. (2012) citam em seu trabalho a importância de um sistema de detecção de fogo ser capaz de detectar chamas quando estas estão obstruídas por algum tipo de objeto, principalmente quando este sistema deve ser usado dentro de um ambiente urbano. No caso deste experimento em especial o método TOR conseguiu captar essa reflexão e poderia facilmente ser usado em um ambiente urbano onde haja objetos na cena que obstruam possíveis locais com probabilidade de incêndio.

O método D383F com os parâmetros usados também foi capaz de capturar essas reflexões no vídeo VDT 1, tendo a mesma capacidade de ser usado em um ambiente urbano com obstrução do fogo por parte dos objetos. A figura 4.12 mostra as capturas de ecrã dos métodos identificando com marcações retangulares à área da reflexão da chama no concreto classificado como foco de incêndio.



Figura 4.12: Métodos capturando a reflexão da chama no concreto no vídeo VDT 1. (A) método D383F; (B) Método TOR

É necessário frisar que isso só é possível quando a superfície possui uma certa refletância para que as chamas possam ser captadas, e principalmente, a área correspondente à cena observada deve ser bem próxima, já que distâncias maiores como no caso do vídeo VDT 2 com 20 metros do alvo, a menor densidade de fluxo luminoso sobre a câmera é suficiente para atenuar a recepção da luz refletida no concreto e ambos os métodos não conseguirem mais identificar essa reflexão como um foco de incêndio.

Esse experimento mostra que o método D383F pode ser facilmente ajustado para detectar focos de incêndio fora do contexto proposto de áreas florestais, sendo capaz de detectar chamas de aproximadamente 1.5 metros de altura a 30 metros de distância com relativa facilidade (VDT 3), bem como ser usado em ambientes fechados nos quais diversos objetos podem obstruir a chama e ser capaz de detectar a chama pela reflexão da mesma em outros objetos da cena, observadas certas restrições.

Síntese do capítulo

Foi possível observar pelo experimento 1 que as mudanças propostas e que formam as etapas do método D383F fazem dele um método robusto para a tarefa de classificação da existência ou não de focos de incêndio, se equiparando a métodos sedimentados na literatura e no mercado como é o caso do método TOR. O método D383F não só foi capaz de manter um bom desempenho de classificação na casa dos 90% nas medidas de desempenho usadas e eliminar todos os falsos positivos, como também foi capaz de realizar a classificação em tempo real usando resoluções mais altas referentes à tecnologia disponível nas câmeras atualmente.

No experimento 2 é possível ver como ajustes finos na parametrização da persistência temporal e espacial podem levar a resultados mais consistentes, evitando falsos positivos pela inclusão de uma inércia temporal proposital de alguns segundos que incorre em um custo de alguns quadros que podem eventualmente conter verdadeiros positivos. Mesmo nesse caso em que vídeos fora do contexto florestal são usados, o desempenho de classificação alcançado se manteve próximo ao obtido no experimento 1 e superior ao obtido por TOR.

O último experimento levanta a questão do uso do método D383F no contexto urbano e como as metodologias propostas para o mesmo permitem a ele ser capaz de exercer sua função mesmo em casos em que o método TOR falha, com chamas produzidas pela combustão de combustíveis fósseis e que foram capturadas pela câmera perto do limite de saturação. Esta situação mostra que o método tem um dinamismo capaz de suprimir com relativa facilidade deficiências na captura causadas pelo grande número de variáveis que fazem parte de uma cena em ambiente externo.

Com esses experimentos foi exposta a robustez e potencial do método D383F gerado neste trabalho, ficando a par em desempenho de classificação com métodos citados na literatura e dinâmico o suficiente para utilização em diversos contextos, seja ambiental ou urbano.

Capítulo 5

Conclusão

Foi apresentado neste trabalho o método D383F de detecção de focos de incêndios usando visão computacional dentro do contexto de ser capaz de lidar com as constantes alterações de iluminação e de cena impostas no ambiente externo monitorado por uma câmera instalada em uma torre da linha de transmissão de energia elétrica, bem como ser usado em diferentes resoluções de qualidade maior que as relatadas na literatura.

Com base na extração de características relevantes dos vídeos de chamas presentes em incêndios florestais e nas limitações de análise e processamento destes vídeos que os computadores estão sujeitos, o algoritmo possui 3 etapas principais:

- Na primeira etapa, o movimento na cena é avaliado com uso da técnica de subtração básica de fundo estimada por médias adaptativas (*running average*) que se baseia em um filtro $\text{ARMA}_{(1,1)}$ para a atualização da imagem de fundo. Esta técnica é invariante a mudanças de iluminação e tem custo computacional baixo, sendo utilizado o esquema de cores RGB para maximizar o desempenho da segmentação do movimento do fogo sob condições em que a escala de cinza falharia. Nessa etapa também foram incluídas técnicas para redução de custo computacional. Nos experimentos realizados nesse trabalho, foi utilizada a técnica de sub-amostragem de pixels, permitindo ao algoritmo D383F ser capaz de manter a detecção em tempo real em resoluções superiores a 352 x 288 pixels;

- A segunda etapa consiste na classificação dos pixels segundo sua cor usando o método proposto por Çelik e Demirel (2009) que usa o espaço de cores YCbCr para separar a informação de crominância da luminância. Uma alteração neste método foi feita, na qual um segundo classificador baseado nas informações de cromaticidade obtidas do canal "V" do espaço de cores YUV é usado, realizando uma segmentação de pixels dinamicamente e mantendo áreas de possíveis focos de incêndios em constante observação. Esses dois classificadores compõem um comitê de classificadores heterogêneos que trabalham em conjunto, melhorando o desempenho nessa etapa;

- A terceira etapa se baseia em extrair as características singulares do fogo de movimento contínuo em torno do material combustível e tempo de duração da combustão e usá-las para a persistência espacial e temporal dos objetos. Isto é feito

utilizando os grupos gerados pelo algoritmo de florestas de conjuntos disjuntos a partir dos pixels classificados nas etapas anteriores. Aqui, foi proposta uma técnica baseada parcialmente no comportamento de colônias de formigas que permite o uso das informações espaciais e temporais do fogo para a eliminação de falsos positivos, aumentando a confiabilidade do método D383F.

Nos experimentos realizados, observou-se que o método D383F obtém bons resultados, superando cada método em pelo menos um critério avaliativo e nos critérios não superados, o método D383F se mantém muito próximo ao melhor desempenho obtido.

No primeiro experimento com vídeos em âmbito florestal (VideoRes), sua acurácia e medida F superam os 90% em todas as resoluções testadas, possuindo um desempenho próximo ou superior ao método (TOR) que representa até o momento o estado da arte nessa área. Ainda o método D383F conseguiu eliminar todos os falsos positivos da base de vídeos VideoRes e pelos resultados obtidos no tempo de processamento de cada vídeo, fica evidente que o método é capaz de executar a tarefa de detecção em tempo real em resoluções HD de forma simples e eficiente.

Usando as bases de dados de vídeos mistos em ambientes florestais e urbanos, VideoDiv e VideoDivCtrl, o segundo experimento mostrou como o uso da persistência temporal e espacial pode influir nos resultados obtidos para eliminar falsos positivos, superando os resultados obtidos pelo método TOR e alcançando um desempenho de classificação geral próximo ao obtido no primeiro experimento.

O terceiro experimento mostrou a sensibilidade do método D383F para detectar chamas fora do escopo de incêndios florestais proposto. Em distâncias variadas, o método mostra um desempenho aceitável para monitoração em ambientes urbanos que possuem materiais combustíveis diferentes dos encontrados em um incêndio florestal, superando novamente o método TOR. Ainda, a habilidade de ser usado para detecção de focos de incêndio obstruídos por outros objetos foi demonstrada, indicando outra característica de robustez do método gerado.

Foi demonstrado, portanto, que através de regras simples e características visuais "primitivas" é possível gerar um método robusto e dinâmico para executar a tarefa de detecção de focos de incêndios em tempo real não só dentro do escopo florestal proposto, mas de maneira geral em qualquer ambiente que necessite de monitoração sem a presença local de seres humanos.

Dentro do que foi proposto, o uso de imagens para a monitoração ambiental ainda oferece uma segunda vantagem, permitindo o uso destas informações digitais de detecção classificação e alarme para estudos posteriores. Um exemplo seria o desenvolvimento de uma nova técnica de detecção automática, criando um novo espaço de cor que permita a extração de características específicas de cor do fogo, obtendo um melhor desempenho do algoritmo na etapa de classificação por cor. Ainda, estudos do comportamento do fogo em incêndios florestais podem ser levantados com estas imagens, sendo possível estimar e inferir as condições e características do ambiente que levaram à aquele evento, diminuindo a necessidade de um estudo em campo em regiões de

difícil acesso, economizando recursos e propiciando o estudo e desenvolvimento de pesquisas nesta área.

Referências Bibliográficas

Angayarkkani, K. e Radhakrishnam, N. (2010). An intelligent system for effective forest fire detection using spatial data. *International Journal of Computer Science and Information Security*, 7:203–208.

Batista, A. C. (1995). *Avaliação da queima controlada em povoamentos de Pinus taeda L. no Norte do Paraná*. PhD thesis, Universidade Federal do Paraná.

Bernardin, K. e Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The clear mot metrics. *EURASIP Journal on Image and Video Processing*, 2008:10.

Beutling, A. (2009). *MODELAGEM DO COMPORTAMENTO DO FOGO COM BASE EM EXPERIMENTOS LABORATORIAIS E DE CAMPO*. PhD thesis, Universidade Federal do Paraná.

Boult, T., Micheals, R., Gao, X., e Eckmann, M. (2001). Into the woods: visual surveillance of noncooperative and camouflaged targets in complex outdoor settings. *Proceedings of the IEEE*, 89(10):1382–1402.

Bradski, G. R. (2008). *Learning OpenCV*. O'Reilly Media Inc., 1st edition edition.

Braga, A. P. (2007). *Redes neurais artificiais: teoria e aplicações*. LTC Editora.

BrasilEscola (2013). Hormônios vegetais. http://www.brasilescola.com/biologia/hormonios-vegetais.htm.

Chamberlin, G. e Chamberlin, D. (1980). *Colour: Its measurement, computation and application*. Heyden.

Chen, T.-H., Wu, P.-H., e Chiou, Y.-C. (2004). An early fire-detection method based on image processing. Em *Image Processing, 2004. ICIP '04. 2004 International Conference on*, volume 3, pgs. 1707–1710 Vol. 3.

Cheng, X., Wu, J., Yuan, X., e Zhou, H. (1999). Principles for a video fire detection system. *Fire Safety Journal*, 33(1):57 – 69.

Collins, R. T., Lipton, A. J., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., Hasegawa, O., Burt, P., e Wixson, L. (2000). A system for video surveillance and monitoring. Technical report, Carnegie Mellon University.

Cormen, T. H., Leiserson, C. E., Rivest, R. L., e Stein, C. (2009). *Introduction to Algorithms, Third Edition*. The MIT Press, 3rd edition.

Crutzen, P. e Goldammer, J. (1993). *Fire in the environment: the ecological, atmospheric, and climatic importance of vegetation fires*. Dahlem Workshop Reports – Environmental Sciences. Wiley.

Donohoe, G., Hush, D. R., e Ahmed, N. (1988). Change detection for target detection and classification in video sequences. Em *Acoustics, Speech, and Signal Processing, 1988. ICASSP-88., 1988 International Conference on*, pgs. 1084–1087 vol.2.

Dorigo, M. (1992). *Optimization, Learning and Natural Algorithms*. PhD thesis, Politecnico di Milano, Italy.

Dorigo, M., Caro, G. D., e Gambardella, L. M. (1999). Ant algorithms for discrete optimization. *Artificial Life*, 5:137–172.

Eisberg, R. M. e Resnick, R. (1999). *Física quântica: átomos, moléculas, sólidos, núcleos e partículas.* Campus.

Çelik, T. (2010). Fast and efficient method for fire detection using image processing. *ETRI Journal*, 32:881–890.

Çelik, T. e Demirel, H. (2009). Fire detection in video sequences using a generic color model. *Fire Safety Journal*, 44:147–158.

Çelik, T., Demirel, H., Ozkaramanli, H., e Uyguroglu, M. (2007). Fire detection using statistical color model in video sequences. *J. Vis. Comun. Image Represent.*, 18(2):176–185.

Equasys (2013). Color formats. http://www.equasys.de/colorformat.html.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recogn. Lett.*, 27(8):861–874.

Ferreira, M. A. L. (2007). *Detecção de movimento através de subtração de fundo para vigilância eletrônica automática*. PhD thesis, PUC-Rio.

Gonzalez, R. e Woods, R. E. (2010). *Processamento Digital de Imagens*. 3º edição edition.

Goss, S., Aron, S., Deneubourg, J., e Pasteels, J. (1989). Self-organized shortcuts in the argentine ant. *Naturwissenschaften*, 76(12):579–581.

Guyon, I., Gunn, S., Nikravesh, M., e Zadeh, L. A. (2006). *Feature Extractions Foundations and Applications*. Springer.

Harleam, M. (2012). Precision and recall ratio. http://abdulmaleaq.blogspot.com.br/2012/12/precision-and-recall-ratio.html.

Hou, J., Qian, J., Zhang, W., Zhao, Z., e Pan, P. (2011). Fire detection algorithms for video images of large space structures. *Multimedia Tools and Applications*, 52:45–63. 10.1007/s11042-009-0451-0.

Institute for Statistics Education (2013). Autoregression and moving average (arma) models. http://www.statistics.com/index.php?page=glossary&term$_id$ = 710.

Intel (2013). Color models. http://software.intel.com/sites/products/documentation/hpc.

Ito, K., Ihara, H., Tatsuta, S., e Fujita, O. (1992). Quantitative characterization of flame color and its application : Introduction of cie 1931 standard colorimetric system. *JSME international journal. Ser. 2, Fluids engineering, heat transfer, power, combustion, thermophysical properties*, 35(2):287–292.

Jahne, B. e Haussecker, H. (2000). *Computer vision and applications: a guide for students and practitioners*. Academic Press, Inc., Orlando, FL, USA.

Joglar, F., Mowrer, F., e Modarres, M. (2005). A probabilistic model for fire detection with applications. *Fire Technology*, 41(3):151–172.

Ko, B., Cheong, K.-H., e Nam, J.-Y. (2010). Early fire detection algorithm based on irregular patterns of flames and hierarchical bayesian networks. *Fire Safety Journal*, 45(4):262 – 270.

Kubat, M. e Matwin, S. (1997). Addressing the curse of imbalanced training sets: one-sided selection. Em *Proc. 14th International Conference on Machine Learning*, pgs. 179–186. Morgan Kaufmann.

Lavorato, M. L. A. (2002). A importância da educação ambiental para o brasil e o mundo. *Jornal Ambiente Brasil*, pg. 4.

Lima, M. G. (2010). Avaliação da sensibilidade ao contraste espacial de cores em humanos. Master's thesis, Univerisdade Federal do Pará.

Liu, Z., Hadjisophocleous, G., Ding, G., e Lim, C. (2012). Study of a video image fire detection system for protection of large industrial applications and atria. *Fire Technology*, 48:459–492. 10.1007/s10694-011-0237-6.

Marbach, G., Loepfe, M., e Brupbacher, T. (2006). An image processing technique for fire detection in video images. *Fire Safety Journal*, 41:285–289.

Margotto, P. R. (2010). Curva roc - como fazer e interpretar no spss. Technical report, Escola Superior de Ciências da Saúde.

Metz, C. E. (1978). Basic principles of roc analysis. *Seminars in nuclear medicine*, 8(4):283–298.

Motta, D. S. (2008). Identificação dos fatores que influenciam no comportamento do fogo em incêndios florestais. Master's thesis, UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO.

National Wildfire Coordinating Group (NWCG) (2009). Glossary of wildland fire terminology. Technical report, National Wildfire Coordinating Group.

OpenCV (2013). Opencv documentation. http://docs.opencv.org/.

Patwary, M. M. A., Blair, J., e Manne, F. (2010). Experiments on union-find algorithms for the disjoint-set data structure. Em *Proceedings of the 9th international conference on Experimental Algorithms*, SEA'10, pgs. 411–423, Berlin, Heidelberg. Springer-Verlag.

Phung, S. L., Bouzerdoum, A., e H., N. G. (2009). Learning pattern classification tasks with imbalanced data sets. Em *Pattern Recognition*.

Piccardi, M. (2004). Background subtraction techniques: a review. Em *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 4, pgs. 3099–3104 vol.4.

Rauta, L. R. P. e Fernandes, A. M. R. (2012). Algoritmos de subtração básica de fundo. Em *Simpósio de Excelência em Gestão e Tecnologia*.

Ruas, G. I. S. (2009). Sistema visual de segmentação de objetos robusto a iluminação para monitoramento automático de tráfego de veículos terrestres. Master's thesis, Universidade de Brasília.

Sasaki, Y. (2007). The truth of the f-measure. Technical report, University of Manchester.

Sóbiologia (2013). Morfofisiologia vegetal. http://www.sobiologia.com.br/.

Schultze, T., Kempka, T., e Willms, I. (2006). Audiovideo fire-detection of open fires. *Fire Safety Journal*, 41(4):311 – 314.

Singh, A. K. e Singh, H. (2012). Forest fire detection through wireless sensor network using type-2 fuzzy system. *International Journal of Computer Applications*, 52(9):19–23. Published by Foundation of Computer Science, New York, USA.

Sonaglio, S. e Moecke, M. (2010). Detecção de movimento para sistema automático de vigilância por vídeo. Em *Caderno de publicações acadêmicas 3*.

Stauffer, C. e Grimson, W. E. L. (1999). Adaptive background mixture models for real-time tracking. Em *Second IEEE Workshop on Visual Surveillance*, pgs. 246–252.

Stirling Consolidated Boiler Company (1905). *Stirling, a book on steam for engineers*. The Stirling Company.

Summit Counts (2013). Category: Fire. http://www.summitcounts.com/category/fire/.

Toreyin, B. U. (2013). Computer vision based fire detection software. http://signal.ee.bilkent.edu.tr/VisiFire/.

Toreyin, B. U., Cinbis, R. G., Dedeoglu, Y., e Çetin, A. E. (2007). Fire detection in infrared video using wavelet analysis. *Optical Engineering*, 46:067204–067204–9.

Toreyin, B. U., Dedeoglu, Y., U., G., e Çetin, A. E. (2006). Computer vision based method for real-time fire and flame detection. *Pattern Recognition Letters*, 27(1):49 – 58.

Toreyin, B. U. e Çetin, A. E. (2005). Flame detection in video using hidden markov models. Em *In Proceedings of the IEEE International Conference on Image Processing (ICIP*, pgs. 1230–1233.

Universidade de São Paulo - USP (2013). Você sabe o que é uma câmera digital. http://educar.sc.usp.br/licenciatura/trabalhos/camera.htm.

Verstockt, S., Hoecke, S. V., Beji, T., Merci, B., Gouverneur, B., Cetin, A. E., Potter, P. D., e de Walle, R. V. (2012). A multi-modal video analysis approach for car park fire detection. *Fire Safety Journal*.

Wang, L., Ye, M., e Zhu, Y. (2010). Hybrid fire detection using hidden markov model and luminance map. Em *Medical Image Analysis and Clinical Applications (MIACA)*.

Wen, Y.-C., Yu, F.-X., Zhou, X.-L., e Lu, Z.-M. (2010). A vector quantization based automatic fire detection system. *Information Technology Journal*, 9:758–765.

WWF (2002). *Living Planet Report 2002*. World Wildlife Fund.

WWF (2012). *Living Planet Report 2012*. World Wildlife Fund.

Yuan, F. (2010). An integrated fire detection and suppression system based on widely available video surveillance. *Machine Vision and Applications*, 21:941–948.

Yuen, W. e Chow, W. (2005). A monte carlo approach for the layout design of thermal fire detection system. *Fire Technology*, 41:93–104.

Zhou, X.-L., Yu, F.-X., Wen, Y.-C., Lu, Z.-M., e Song, G.-H. (2010). Early fire detection based on flame contours in video. *Information Technology Journal*, 9:899–908.