

UNIVERSIDADE FEDERAL DO RIO GRANDE
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

Dissertação de Mestrado

Um Sistema de Detecção de Fogo Baseado em Vídeo

Cristiano Rafael Steffens

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande, como requisito parcial para a obtenção do grau de Mestre em Engenharia de Computação

Orientador: Prof. Dr. Ricardo Nagel Rodrigues

Rio Grande, 2015

Dados de catalogação na fonte:
Bibliotecário Clériston Ribeiro Ramos – CRB- 10/1989
Biblioteca Central – FURG

A999a Steffens, Cristiano Rafael

Um Sistema de Detecção de Fogo Baseado em Vídeo / Cristiano Rafael Steffens. – Rio Grande, 2015. – 100 f: gráf. – Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande. Centro de Ciências Computacionais. Rio Grande, 2015. – Orientador Ricardo Nagel Rodrigues.

1. Detecção de fogo. 2. Aprendizado de máquina. 3. Visão computacional. I. Rodrigues, Ricardo Nagel. II. Título.

CDD: 999.9

Banca examinadora:

Prof. Dr. Ricardo Matsumura de Araújo

Prof^a. Dr^a. Silvia Silva da Costa Botelho

Prof. Dr. Ricardo Nagel Rodrigues

AGRADECIMENTOS

Aos meus pais, por sempre terem colocado a minha educação em primeiro lugar e por sempre terem se esforçado para que eu pudesse realizar os meus objetivos. Aos meus irmão pelo apoio, carinho, suporte e incentivo.

Aos colegas e amigos do Centro de Ciências Computacionais, por terem ajudado a colocar em prática as ideias e auxiliado de alguma forma na realização do presente trabalho, dedico o meu apreço. Em especial, agradeço ao companheirismo e colaboração dos colegas do grupo de Sistemas Computacionais e Tecnologias para Percepção.

Ao meu orientador, Prof. Dr. Ricardo Nagel Rodrigues, pela confiança em mim depositada, pelos ensinamentos, pelo apoio e ajuda no desenvolvimento deste trabalho. Por não ter medido esforços para, com dedicação, permitir uma argumentação clara e concisa, tornando o trabalho mais técnico e estimulante.

Aos professores Dr^a. Sílvia Silva da Costa Botelho (FURG) e Dr. Ricardo Matsumura de Araújo (UFPEL), pela disponibilidade e participação na banca de avaliação

Ao coordenador, aos professores e técnicos do Centro de Ciências Computacionais - C3 - FURG.

À Migrate Company Sistemas de Informação LTDA por flexibilizar meu horário de trabalho, permitindo-me conciliar o trabalho com o curso de pós-graduação. À CAPES pelo apoio financeiro que permitiu que eu me dedicasse integralmente às atividades de pesquisa.

*Everything that I can spy
Through the circle of my eye.
Everything that I can see
Has been woven out of me.
I have sown the stars ; I threw
Clouds of morn and noon and eve
In the deeps and steeps of blue ;
And each thing that I perceive.
Sun and sea and mountain high.
Are made and moulded by my eye :
Closing it, I do but find
Darkness, and a little wind.*

— JAMES STEPHENS

RESUMO

STEFFENS, Cristiano Rafael. **Um Sistema de Detecção de Fogo Baseado em Vídeo**. 2015. 100 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande, Rio Grande.

A detecção automatizada de fogo é foco de pesquisa em visão computacional desde o início da década de 90. Neste trabalho, um novo método para a identificação de fogo baseado em vídeo é proposto e analisado. Algoritmos de detecção de fogo utilizando visão computacional são normalmente aplicados em cenários que utilizam câmeras estáticas, tais como circuitos fechados de TV utilizados em sistemas de vigilância. O método proposto, não se restringe apenas a estas condições, mas também possibilita a detecção automática de fogo em vídeos obtidos a partir de câmeras móveis. A utilização de câmeras móveis adiciona uma complexidade significativa ao incluir variação no *background*, movimento em três eixos, trepidação e iluminação não controlada. O método proposto utiliza características estatísticas do fogo e *Random Forests* para detectar potenciais regiões com fogo, suprimindo a dependência da dinâmica do movimento. Estas características incluem a cor, o formato, a dispersão e assimetria da intensidade. Para complementar a classificação atingida a partir das propriedades espaciais, o sistema utiliza-se também das propriedades de oscilação e pulsação ao longo de um curto período de tempo. Estas características temporais são fatores discriminantes, largamente empregados na bibliografia consultada, resultando em classificadores robustos. Os resultados apresentados demonstram a aplicabilidade do método, superando outros métodos anteriormente propostos.

Palavras-chave: Detecção de fogo, aprendizado de máquina, visão computacional.

ABSTRACT

STEFFENS, Cristiano Rafael. **A Video Based Fire Detection System**. 2015. 100 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Computação. Universidade Federal do Rio Grande, Rio Grande.

Automated fire detection is a research focus in computer vision since the early 90s. In this paper, we introduce and evaluate a new Video Fire Detection method that aims to produce good results on steady and mobile cameras. Usually, computer vision based fire detection algorithms are applied in scenarios with static cameras, such as closed-circuit TV used in surveillance systems. The method we are proposing here, in contrast, was developed so that it can be used not only in these conditions, but also for automatic fire detection in videos obtained from mobile cameras. The use of mobile cameras adds a lot of complexity by bringing variation in background, movement in three axes, camera vibration and uncontrolled lighting conditions. The proposed method uses statistical characteristics of fire like regions as input data for a Random Forests classification model, frame-by-frame, in order to detect fire regions. These characteristics are the color, the shape, the color skewness and dispersion measures considering pixel value. In order to increase results given by the spatial properties, our method also uses the flickering over a short period of time. The flickering properties are widely used as a discriminant in the related bibliography, resulting in robust classifiers. The presented results illustrate the applicability of the method, significantly increasing the precision and recall metrics.

Keywords: fire detection, probabilistic fire detection, active-vision, hand-held camera.

LISTA DE FIGURAS

Figura 1	Reflexão da luz	20
Figura 2	Passo a passo da segmentação por Split and Merge	30
Figura 3	Filtro passa-banda dois estágios	42
Figura 4	Método de detecção de MUELLER et al. (2013)	45
Figura 5	Histograma normalizado das direções do fluxo óptico utilizando NSD	45
Figura 6	Anotações de <i>ground truth</i> no <i>dataset</i>	57
Figura 7	Segmentação manual das regiões de chama análise estatística	63
Figura 8	Histogramas de regiões de fogo RGB	64
Figura 9	<i>Boxplot</i> de regiões de fogo RGB	65
Figura 10	Histogramas de regiões de fogo HSV	67
Figura 11	<i>Boxplot</i> de regiões de fogo RGB	68
Figura 12	Histogramas de regiões de fogo Lab	69
Figura 13	<i>Boxplot</i> de regiões de fogo Lab	70
Figura 14	Histogramas de regiões de fogo YCbCr	71
Figura 15	<i>Boxplot</i> de regiões de fogo YCbCr	72
Figura 16	Histogramas de oscilação do valor de píxel no tempo	74
Figura 17	Fluxograma do método proposto	75
Figura 18	Distribuição das regiões de chama	78
Figura 19	Comparativo de segmentação de cor em imagens sem fogo	81
Figura 20	Comparativo de segmentação de cor em imagens com fogo	82
Figura 21	Método de avaliação do classificador proposto	83

LISTA DE TABELAS

Tabela 1	Vídeos base do Dataset	55
Tabela 2	Estatísticas de cor nas regiões de fogo	66
Tabela 3	Estatísticas de cor nas regiões de não-fogo	66
Tabela 4	Oscilação dos píxels no tempo	72
Tabela 5	Parâmetros da <i>Random Forest</i> para classificação de cor	77
Tabela 6	Parâmetros da <i>Random Forest</i> para classificação de regiões	79
Tabela 7	Resultados do método proposto na avaliação <i>frame a frame</i>	85
Tabela 8	Resultados do método proposto considerando localização	86
Tabela 9	Resultados do método proposto na avaliação <i>frame a frame</i> para o caso de teste II	87
Tabela 10	Resultados do método proposto considerando localização para o caso de teste II	87

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ANN	Artificial Neural Network
CC	Creative Commons License
CCD	Charge-coupled Device
CUDA	Compute Unified Device Architecture
OpenCV	Open Computer Vision library
PCA	Análise de Componentes Principais
HMM	Hidden Markov Models
MCMC	Markov Chain Monte Carlo
LBP	Local Binary Pattern
NP	Non-Deterministic Polynomial time
FFT	Fast Fourier Transform
FLANN	Fast Library for Approximate Nearest Neighbors
HMM	Hidden Markov Model
HSV	Hue Saturation Value colorspace
FDDB	Face Detection Database
FDRB	Função de Distribuição de Refletância Bidirecional
OMT	Optimal Mass Transport
RGB	Red Blue Green Colorspace
ROI	Region of Interest
SIMD	Single Instruction Multiple Data
SIFT	Scale Invariant Feature Transform
SURF	Speeded Up Robust Features
SVM	Support Vector Machine
SSLT	Subsurface Light Transport
VFD	Video Fire Detection

SUMÁRIO

1	INTRODUÇÃO	13
1.1	Objetivo	14
1.2	Recursos	14
1.3	Estrutura Do Texto	17
2	AQUISIÇÃO E PROCESSAMENTO DE IMAGENS	19
2.1	Fenômenos Ópticos	19
2.2	Processamento de imagens	21
2.2.1	Detecção de bordas, linhas e círculos	21
2.2.2	Detectores baseados em <i>templates</i>	23
2.2.3	Detector de Marr-Hildreth	24
2.2.4	Detector de Canny	25
2.3	Segmentação	28
2.3.1	Thresholding	28
2.3.2	Segmentação baseada em região	29
2.4	Detecção de Movimento	31
3	TRABALHOS RELACIONADOS	34
3.1	Detecção de Cor	35
3.2	Detecção de Movimento	40
3.2.1	Análise da Oscilação no Domínio da Frequência	41
3.2.2	Análise da Oscilação Utilizando Fluxo Óptico	43
3.2.3	Análise da Oscilação Utilizando Mineração de Dados	46
3.2.4	Análise da Oscilação com Base no Formato da Chama	47
3.3	Detecção Baseada em Textura	49
4	DATASET E MÉTRICAS DE AVALIAÇÃO	52
4.1	Dataset	53
4.1.1	Anotações de <i>Ground Truth</i>	56
4.1.2	Artefatos de Software	56
4.2	Metodologia de Avaliação	57
5	AVALIAÇÃO ESTATÍSTICA	61
5.1	Classificação De Cor	61
5.2	Classificação da Oscilação	72
6	MÉTODO PROPOSTO	75

7	RESULTADOS	80
7.1	Casos de teste	81
8	TRABALHOS FUTUROS	89
9	CONCLUSÃO	91
	REFERÊNCIAS	93

1 INTRODUÇÃO

A detecção de fogo a partir de imagens é algo recente quando comparada a outras aplicações dentro da visão computacional. Ao passo que técnicas de reconhecimento de faces e objetos foram minuciosamente estudadas permitindo que grande parte das aplicações possam ser criadas apenas fazendo pequenas adaptações e o uso de algoritmos já validados, a detecção de fogo acaba caindo em uma categoria específica de problemas da visão computacional. Os motivos são claros quando faz-se uma rápida avaliação do que é indicativo de fogo em uma imagem e como as características do fogo se diferenciam das de outros objetos.

Considerando o problema dentro de um contexto maior, pode-se rapidamente perceber a importância de uma detecção rápida de incêndios e outros incidentes envolvendo fogo não controlado. Os incêndios afetam a vida cotidiana em todo o mundo, colocando pessoas em risco e gerando danos materiais. É intuitivo afirmar que quanto antes o fogo é detectado, maiores serão as chances de evitar danos e em casos mais graves até mesmo assegurar a sobrevivência. Informações como a localização inicial do fogo, área afetada, direção de propagação da fumaça e a taxa de crescimento do fogo são parâmetros importantes na análise de segurança e combate a incêndios, bem como a redução dos impactos decorrentes.

Recentemente as técnicas de detecção de incêndio através de vídeo (em inglês *Video Fire Detection* - VFD) têm se mostrado úteis para complementar as técnicas de detecção de incêndio tradicionais, como as baseadas em sensores de fumaça, que são geralmente limitados a ambientes fechados, não sendo aplicáveis em grandes espaços abertos. Os sensores tradicionais exigem uma proximidade com o fogo e a maioria deles não é capaz de fornecer informações adicionais sobre a localização do fogo, dimensão, etc. Outra limitação dos sistemas de alarme de incêndio convencionais é o tempo de resposta alto, visto que são acionados pelas partículas de carbono e fumaça para que chegam até o detector.

1.1 Objetivo

Propôr um detector de chammas com bons resultados no processamento de vídeos obtidos a partir de câmeras móveis e capaz de realizar detecção em tempo real minimizando as taxas de falsos positivos e falsos negativos. No contexto, entende-se por detecção o conjunto das atividades de detectar a existência de chama na cena e apontar o local onde ocorre. Ao final espera-se que os resultados do trabalho possam ser integrados ao *Framework CvWorks*, um *Framework* para Visão Computacional de Alto Nível que é desenvolvido pela Universidade Federal do Rio Grande (FURG). De forma mais detalhada, o trabalho enfatiza as seguintes questões:

- Estudar os detectores utilizados em sistemas baseados em câmeras estacionárias;
- Criar um *dataset* para detecção de fogo em vídeos com anotações de *ground truth*;
- Implementar alguns detectores de chama obtidos na bibliografia e verificar os resultados obtidos quando usados para detecção de imagens em vídeos de câmera móvel;
- Apresentar uma alternativa funcional para detecção e localização de fogo a partir de câmeras estáticas;
- Apresentar uma alternativa funcional para detecção e localização de fogo a partir de câmeras móveis;

1.2 Recursos

No desenvolvimento deste trabalho fez-se o uso da linguagem de programação C++ acompanhado das bibliotecas OpenCV e Boost. Os testes foram executados utilizando um *dataset* composto por vídeos obtidos a partir de diversas fontes.

1.2.0.1 Dataset

No decorrer deste trabalho não possível encontrar *datasets* para fazer a validação de *ground truth* e comparação automatizada do método proposto com os métodos de detecção de chammas apresentados nos trabalhos relacionados. Com a finalidade de conseguir uma validação dos resultados atingidos, verificando a acurácia, a taxa de positivos verdadeiros, a taxa de falsos positivos e falsos negativos, criou-se vídeos obtidos no site de compartilhamento *YouTube.com* que estão licenciados pela *Creative Commons License*. A criação do *dataset* e as métricas de avaliação utilizadas são detalhadas no Cap. 4.

Nas fases iniciais do trabalho, foram utilizados também os vídeos disponibilizados na *Bilkent VisiFire Sample Fire and Smoke Video Clips* e *KMU Fire and Smoke Dataset*. Estes *datasets* foram amplamente utilizados para avaliar

sistemas de detecção de fogo baseados em câmeras de vídeo publicados anteriormente. *Bilkent VisiFire Sample Video Clips*, disponível através do link <http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SampleClips.html>, foi utilizado nos trabalhos publicados por TOREYIN; CETIN (2009), TOREYIN et al. (2005), ÇETIN et al. (2013), TOREYIN; DEDEOGLU; CETIN (2006) e GÜNAY et al. (2010). O *dataset* contém 30 arquivos, distribuídos em quatro categorias:

1. Vídeos de fogo;
2. Vídeos de fumaça;
3. Vídeos de fumaça em floresta;
4. Vídeos sem incidência de fogo nem fumaça visível.

Já a *KMU Fire & Smoke database* (<http://cvpr.kmu.ac.kr/Dataset/Dataset.htm>), utilizado em KO; HAM; NAM (2011), PARK et al. (2013) e SHIDIK et al. (2014), é composto por 38 vídeos de resolução 340 x 240, com taxa de quadros por segundo variando entre 15 e 30 Herz. Os vídeos são distribuídos nas seguintes categorias:

1. Fogo em ambientes abertos e fechados;
2. Fumaça em ambientes abertos e fechados;
3. Objetos com comportamento semelhante ao do fogo.

Tanto a *KMU Fire & Smoke database* quanto a *Bilkent VisiFire Sample Video Clips* são compostos por vídeos de câmeras estacionárias. A câmera encontra-se sempre em um lugar fixo e a trepidação da imagem é insignificante. Um ponto que merece ser ressaltado é que estes *datasets* não fornecem nenhuma anotação sobre a localização da chama na imagem, limitando a avaliação dos algoritmos propostos. Dado que objetiva-se construir um sistema capaz de funcionar também com sistemas de câmeras não estacionárias, obteve-se pela criação de um *dataset* próprio composto por vídeos de câmeras móveis e onde as regiões de chama são anotadas individualmente.

1.2.0.2 Ferramentas de desenvolvimento

A prototipagem e desenvolvimento do trabalho é feita, em sua maior parte, utilizando a linguagem de programação C++ em conjunto com as bibliotecas Boost (BOOST C++ LIBRARIES, 2014) e OpenCV (BRADSKI, 2000). Boost é uma coleção de bibliotecas que estendem a funcionalidade da linguagem de programação C++ e sua licença permite o uso em qualquer projeto. A Boost inclui funcionalidades para diversas áreas dentre programação concorrente, estrutura de dados, análise sintática, grafos e funções para entrada e saída de dados.

A OpenCV é uma biblioteca *Open Source* de processamento de imagens com implementações em C, C++ e Python. Ela implementa várias funções de detecção de bordas, reconhecimento de objetos, análise de vídeos, fluxo óptico, descritores SIFT e SURF e algoritmos de aprendizagem de máquina. Dentre as técnicas de aprendizagem de máquina pode-se destacar *Support Vector Machine*, Redes Neurais com *Back-Propagation*, *K-means*, *K-Nearest Neighbors*, *Classificador Bayesiano* e árvores de classificação.

Em algumas etapas, para descoberta de conhecimento utilizou-se também a ferramenta de mineração de dados WEKA (HALL et al., 2009). A WEKA foi utilizado para seleção de melhores atributos e classificação utilizando árvores. A principal vantagem de utilizar a WEKA é que esta permite utilizar os métodos de aprendizagem no *dataset* e analisar os resultados obtidos com um tempo de *setup* mínimo. Utilizando a ferramenta é possível também encontrar a correlação entre os atributos. Em conjunto com a WEKA (HALL et al., 2009) utilizou-se também editores de planilha e texto para pré-processamento.

Conforme citado na Sec. 1.1 os resultados deste trabalho devem ser integrados no *framework* CvWorks. Trata-se de um *framework* de visão computacional desenvolvido no Centro de Ciências Computacionais da Universidade Federal do Rio Grande (FURG) que traz implementações genéricas, sendo voltado para uma ampla gama de aplicações independentes da representação dos objetos. O objetivo principal deste *framework* é implementar código reutilizável, incluindo interface gráfica, algoritmos de detecção de faces, tags de realidade aumentada, placas, instrumentos industriais, fogo e rastreamento de objetos. O *framework* utiliza orientação a objetos e *templates* com o objetivo de se tornar mais genérico e compatível com outras bibliotecas de visão computacional.

Em nível de aplicação o CvWorks é dividido em três partes assim definidas na arquitetura:

- *VisionGUI*: Oferece uma interface gráfica de uso geral que permite a visualização da saída do algoritmos de detecção e rastreamento.
- *C3 Scout*: Voltado para aplicação em robôs móveis utilizáveis em ambientes externos e internos. O objetivo é que o robô possa fazer uma ronda contínua, restrito à um ambiente pré-determinado, identificando ameaças e irregularidades. Nesta aplicação incluem-se funcionalidades de reconhecimento e rastreamento de pessoas, detecção de armas, fogo e placas de automóveis.
- *Field Vision*: Voltado para a extração de informação de plantas industriais a partir de câmeras estáticas e disponibilização das mesmas através de protocolos industriais. Esta aplicação inclui algoritmos para detecção e leitura de medidores analógicos, status de registradores, leitura de nível de tanque e detecção de fogo.

É importante citar que como este não é um *software* fechado ou voltado para uma finalidade específica vários fatores foram levados em consideração e impactaram nas decisões sobre o método de detecção de fogo proposto. Quando trabalha-se com robôs móveis e visão ativa, como é o caso da aplicação *C3 Scout*, torna-se impossível fazer uma validação utilizando como informação a variação no pixel ou na região ao longo de uma sequência de *frames*. Tanto em câmeras acopladas a robôs móveis quanto em câmeras estáticas fazendo a supervisão de uma planta industrial, o sistema irá se deparar com *backgrounds* complexos e uma variedade de cenários, com interação de pessoas, balançar de folhas, objetos deformáveis, movimentação de equipamentos, atuadores e sinalizadores industriais. As condições de iluminação não são controladas, o que aumenta o grau de dificuldade, criando diferentes entradas dependendo de parâmetros da câmera como balanceamento de brancos e exposição.

1.3 Estrutura Do Texto

O capítulo 1 introduz o problema da detecção de chamas baseada em vídeo, elucidando as principais dificuldades, objetivos e recursos utilizados, contextualizando assim, o problema tratado neste trabalho.

No capítulo 2 apresenta a evolução dos sistemas de reconhecimento de objetos ao longo dos anos e concebe uma visão geral das abordagens propostas. Faz-se um levantamento amplo das soluções adotadas na tentativa de resolver o problema de visão, considerando as subáreas como representação de imagens, filtragem, detecção de bordas, segmentação, análise de texturas, reconhecimento de formas e objetos e da modelagem matemática dos principais descritores e *matchers* utilizados.

Na sequência, o Capítulo 3 faz a definição de termos e apresenta as técnicas que representam o estado da arte em sistemas VFD (*Video Fire Detection*). É feita uma revisão geral sobre o que é o fogo e suas propriedades. Além disso, são apresentadas as técnicas utilizadas na tarefa de detectar fogo usando apenas dados de vídeo. São apresentados os trabalhos relacionados e discute-se seus resultados.

No capítulo 4 apresenta-se o *dataset* utilizado para treinamento e avaliação dos detectores. Discute-se o processo de criação do *ground truth*, apresentando as características, limitações e convenções utilizadas. Na parte final, apresentam-se as métricas pelas quais os detectores são comparados.

A seguir, no capítulo 5 faz-se uma detalhada avaliação sobre as propriedades e comportamento do fogo utilizando estatística descritiva. Busca-se levantar dados para fundamentar a proposta de detector trazida no Capítulo 5. Faz-se um estudo das características que permitem que uma região de fogo seja identificada através de um único *frame*, avaliando apenas os dados espaciais e resultando em poucos falsos positivos.

O Capítulo 6 apresenta a metodologia adotada neste trabalho, as técnicas utilizadas

para segmentação, tratamento e extração de características de cor e oscilação. Além disso, detalha-se e apresenta-se os resultados atingidos pelo sistema de detecção de fogo baseado em vídeo proposto neste trabalho.

Continuando, o Capítulo 7 apresenta os resultados do trabalho através de uma perspectiva do que foi proposto, estudado e os resultados obtidos. Através de uma análise quantitativa e qualitativa busca-se identificar pontos de melhoria para trabalhos futuros.

Os capítulos 8 e 9 trazem respectivamente as propostas para a continuação do trabalho de pesquisa e as conclusões com base nos objetivos do trabalho. Apresenta-se uma revisão das principais dificuldades encontradas e possíveis soluções para superá-las. Da mesma forma são discutidas possíveis melhorias de desempenho da solução apresentada.

2 AQUISIÇÃO E PROCESSAMENTO DE IMAGENS

A visão depende fundamentalmente da luz. Nesta seção busca-se descrever como a iluminação, as propriedades da superfície dos objetos e a óptica da câmera interagem de modo a produzir os valores de cor que se enquadram no sensor de imagem. Descreve-se como imagens coloridas contínuas são transformadas em amostras digitais discretas dentro do sensor de imagem e as deficiências decorrentes deste processo.

As imagens são formadas pelas cores e pela intensidade. A fotometria trata da medição da luz, em termos de como seu brilho é percebido pelo olho humano. Já a radiometria, é um conjunto de técnicas de medição da radiação eletromagnética que mede a luz em termos de sua potência absoluta no espaço, desconsiderando sua interação com o olho humano. Dado que as propriedades da luz e os fenômenos ópticos como reflexão, refração, difusão, fluorescência, fosforescência, *Subsurface Scattering* e inter-reflexão impactam diretamente na aquisição das imagens estes fenômenos serão rapidamente estudados neste capítulo.

A luz é um fenômeno de natureza ondulatória. É uma radiação eletromagnética que se propaga através de diferentes meios, como o ar, a água ou mesmo o vácuo. A luz é constituída de fótons, minúsculas partículas elementares que apresentam simultaneamente natureza de onda e partícula. Um fóton surge quando ocorre a transição de um elétron de um átomo entre dois estados de energia diferentes. Quando excitado o elétron passa para um nível de maior energia e, ao retornar para o estado inicial, emite a energia correspondente à diferença entre os dois níveis.

2.1 Fenômenos Ópticos

Os fenômenos ópticos estão diretamente relacionados à forma como percebemos a imagem. Pelo fato de o fogo representar uma fonte de luz estes fenômenos estão diretamente associados aos resultados atingidos. A reflexão é um dos fenômenos ópticos mais importantes para a fotografia (TRIGO, 1998). Na reflexão o ângulo de incidência i é igual ao ângulo de reflexão r medido em relação a reta normal N . Estão no mesmo plano os raios incidente e refletido, conforme pode ser observado na figura 1.

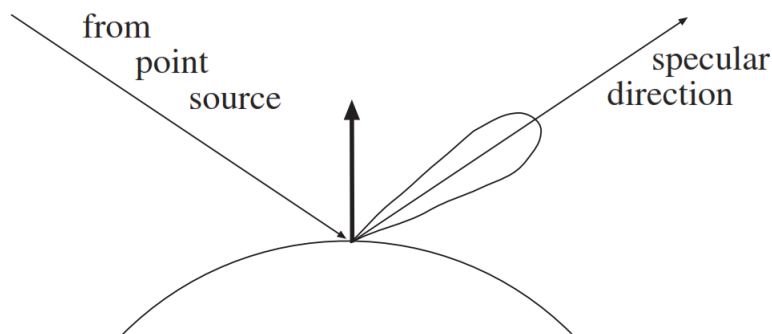


Figura 1: Reflexão da luz (FORSYTH; PONCE, 2002)

O modelo geral para reflexão local é dado pela Função de Distribuição de Refletância Bidirecional (FDRB) que modela a distribuição da refletância em função da posição da fonte de iluminação e do observador. Assim, a refletância em uma determinada direção é dada pelo material que reflete, pelo comprimento da onda e pela geometria de iluminação e de observação (FORSYTH; PONCE, 2002). A reflexão pode ser classificada em três tipos:

- Especular ou Regular: Quando a reflexão ocorre em uma superfície lisa. Especular advém da palavra grega *speculo* que significa espelho.
- Difusa: O feixe de raios paralelos incide na superfície do objeto, os raios refletidos perdem o paralelismo e são espalhados em todas as direções.
- Lambertiana: A reflexão não está ligada ao ângulo de incidência dos raios provenientes da fonte luminosa. Uma superfície Lambertiana dispersa a luz igualmente em todas as direções. Alguns materiais com este tipo de reflexão são roupas de algodão, carpetes e papel fosco.

Outro fenômeno óptico é a refração. A refração ocorre quando a luz passa de um material óptico para outro, sofrendo uma mudança na velocidade de propagação. Um exemplo clássico desse fenômeno é visto colocando-se um lápis em um copo com água. Existe uma tendência de que a luz se aproxime da normal ao mudar de meio de propagação.

Alguns materiais têm a capacidade de fazer o *Subsurface Scattering* (também conhecido como o *Subsurface Light Transport* (SSLT)) onde a luz penetra na superfície de um objeto translúcido, sendo dispersada por interação com o material, e sai da superfície de um ponto diferente. A luz penetra na superfície e é refletida algumas vezes em ângulos irregulares no interior do material, antes de sair do material em um ângulo diferente daquele que teria se tivesse sido refletida diretamente para fora da superfície. Como resultado desse efeito, os fótons penetram na superfície e depois da dispersão da luz, tem-se então áreas dessa mesma superfície com altos níveis de saturação. Isso faz com que a

superfície pareça iluminada diretamente do seu interior. É um fenômeno que ocorre por exemplo quando a luz atravessa uma orelha humana.

Por fim, tem-se a inter-reflexão, fenômeno em que a luz refletida de um objeto atinge outros objetos na área circundante, iluminando-os. A inter-reflexão difusa trata especificamente da luz refletida a partir de objetos que não são brilhantes ou especulares, ou seja, onde que a luz é refletida por superfícies não-brilhantes, como o chão, paredes ou tecido, para chegar a áreas que não tem visada direta da fonte de luz. Se a superfície difusa é colorida, a luz refletida também é colorida, resultando em coloração semelhante nos objetos próximos.

2.2 Processamento de imagens

Nesta secção são apresentados os conceitos fundamentais sobre imagens digitais, espaços de cores, contraste, brilho, suavização, realce e segmentação.

2.2.1 Detecção de bordas, linhas e círculos

A detecção de bordas é uma das operações empregadas com maior frequência na análise de imagens e é provavelmente o tema mais maturado dentro desta área de pesquisa. O principal motivo de haver um grande número de pesquisas que trabalham a detecção de bordas é que, genericamente, pode-se afirmar que a melhor representação de um objeto é dada pelas suas bordas (PARKER, 2010).

Entende-se por borda um limite que separa o objeto do restante da cena contida na imagem bem como indicar uma sobreposição de um objeto em relação a outro (ZIOU; TABBONE et al., 1998). Em outras palavras pode-se afirmar que se as bordas de um objeto podem ser identificadas precisamente então todos os objetos podem ser localizados e suas propriedades podem ser mensuradas. Sabendo que a visão computacional está assentada sobre o conceito de objeto é fácil notar que o seu reconhecimento é essencial para os passos seguintes do processo.

Pode-se considerar a detecção de bordas como uma parte do processo de segmentação de imagens. O objetivo da segmentação de imagens é identificar regiões dentro de uma imagem. Essas regiões podem ser então isoladas e posteriormente a aplicação de outros algoritmos pode definir que tipo de objeto cada segmento isolado representa. A detecção de bordas usualmente requer suavização e diferenciação das imagens. Uma borda pode ser caracterizada por variações significativas na refletância, iluminação, orientação e profundidade da cena. É difícil criar um algoritmo detector de bordas genérico que ofereça bons resultados em diferentes condições e aplicações. Isso obrigou os pesquisadores a criarem detectores, variando entre abordagens matemáticas e algorítmicas (ZIOU; TABBONE et al., 1998).

Existem três conceitos principais associados à bordas (PARKER, 2010):

- Detecção: é o processo de localização de píxeis de borda.
- Realce: é o processo de aumentar o contraste entre as bordas e a cena de forma a deixá-las mais visíveis.
- Rastreamento: é o processo de seguir as bordas, geralmente criando listas de píxeis podendo ser utilizada para identificação do modelo ou objeto.

A detecção de bordas em imagens digitais precisa contornar algumas dificuldades e limitações ao lidar com imagens digitais. A primeira delas é diretamente decorrente do processo de digitalização que obriga a imagem a ser amostrada de forma que todos os píxeis assumam um valor, normalmente entre 0 a 255 quando trabalhamos com o espaço de cor RGB. Isso significa que uma mudança gradual (analógica) da cor foi transformada em uma sequência de números inteiros que se aproximam dos valores reais (PARKER, 2010). Neste caso, fica fácil perceber que existe uma perda de informações que estavam presentes na imagem original. Existe ainda o fator de uma borda ser distribuída em vários píxeis na imagem capturada.

A segunda limitação é o ruído, que pode ser decorrente de diversos fatores como iluminação, tipo de lente, câmera utilizada para a captura da imagem, movimento, poeira ou mesmo efeitos atmosféricos. O efeito deste ruído é alterar valores de píxeis de forma aleatória. Os efeitos podem ser minimizados utilizando métodos estatísticos como a distribuição probabilística, mas não existem garantias de que os dados se mantenham fiéis aos dados originais. Devido à sua natureza aleatória, o ruído não pode ser previsto e é difícil mensurá-lo adequadamente. Para a visão computacional, considerando que a maior parte dos algoritmos usa escala de cinza, o ruído pode ser classificado como independente de sinal ou dependente de sinal (PARKER, 2010).

O ruído independente de sinal ocorre quando uma imagem é transferida eletronicamente de um lugar para outro e caracteriza-se como um conjunto de píxeis em escala de cinza que não tem uma ligação com os conjunto de dados originais são adicionados a imagem original. Matematicamente, pode se definir este ruído como $B = A + R$, onde A é uma imagem perfeita, R é o ruído que ocorre durante a transmissão e B é a imagem resultante. A e R não tem qualquer relação entre si, sendo que o ruído R pode ter propriedades estatísticas quaisquer, mas usualmente assume-se que segue uma distribuição gaussiana normal com média 0 e um desvio padrão mensurado ou presumido PARKER (2010).

Já o ruído dependente de sinal leva este nome porque cada píxel de ruído na imagem pode ser calculado em função dos píxeis vizinhos. Este é o tipo de ruído que pode ser observado em algumas fotografias e é mais complicado de ser trabalhado segundo PARKER (2010). Por sua vez, geralmente tem um impacto menor nos resultados e é gerenciável quando a imagem é amostrada apropriadamente.

Bons detectores de bordas devem ser capazes de tratar estes tipos de ruído fornecendo resultados confiáveis. Cabe salientar que tanto as bordas quanto o ruído independente de sinal apresentam variações locais bruscas na intensidade dos píxeis e por este motivo a borda é definida pela conexão entre vários píxeis que poderiam definir uma borda. Genericamente, todo detector de borda é composto pelas etapas de diferenciação para avaliar as derivadas desejadas da imagem, suavização da imagem para remoção de ruído e etiquetagem para encontrar as bordas e aumentar a tolerância ao ruído, suprimindo falsas bordas.

Em relação aos tipos, a borda degrau (ou *step edge*) é definida como um ponto onde ocorre uma descontinuidade na escala de cinza. Para fazer sua detecção precisa-se fazer um tratamento do ruído presente na imagem capturada e obter a derivada de primeira ordem (onde as bordas serão indicadas pela máxima positiva ou pela mínima negativa) ou obter a derivada de segunda ordem (onde as bordas são dados pelos pontos onde existe passagem por zero). Já para bordas linha, que são resultado da iluminação mútua entre objetos que estão em contato, ou objetos pequenos em frente a objetos maiores, bordas correspondem a extremos locais na imagem. Eles são localizados como cruzamentos em zero da primeira derivada, máximas locais da laplaciana, ou ainda como máximas locais da variação dos níveis de cinza ZIOU; TABBONE et al. (1998).

Por fim, uma junção é um tipo de borda obtida quando duas ou mais bordas se encontram e pode ser localizada normalmente como pontos com alta curvatura, pontos com grande variação na direção do gradiente ou ainda como pontos de cruzamento em zero da laplaciana que apresentem uma alta curvatura.

Desde seu surgimento o número de detectores de bordas vem crescendo continuamente fazendo com que seja praticamente impossível fazer um levantamento completo dos algoritmos existentes. O que diferencia um detector de outro são os filtros de suavização utilizados, os operadores de diferenciação, o processo de etiquetagem, a complexidade computacional e o modelo matemático utilizado.

2.2.2 Detectores baseados em *templates*

Os primeiros detectores de bordas baseados em gradiente ou em operadores Laplacianos, propostos a mais de 40 anos, se limitavam a operação de diferenciação. Os filtros Sobel (SOBEL; FELDMAN, 1968) e Prewitt (PREWITT, 1970) estimam o vetor de gradiente com base na operação de convolução utilizando a máscara

$$\Delta_x = \begin{vmatrix} -1 & 0 & 1 \\ -a & 0 & a \\ -1 & 0 & 1 \end{vmatrix}$$

para obtenção das bordas em x , tendo a como um número real positivo (1 no caso de Prewitt e 2 para Sobel). Já para detecção das bordas na direção y é utilizado o seguinte

kernel de convolução:

$$\Delta_y = \begin{vmatrix} -1 & -a & -1 \\ 0 & 0 & 0 \\ 1 & a & 1 \end{vmatrix}$$

Estas mascaras de convolução resultam na primeira derivada parcial de uma área de 3 x 3 píxeis. Para a estimativa Laplaciana utiliza-se a seguinte máscara:

$$\nabla = \begin{vmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{vmatrix}$$

Usualmente estas máscaras tem tamanho fixado em 3 x 3. Cabe salientar que a os resultados destas operações pioram quando existe muito ruído, dado que, ambas assumem que o ruído é aditivo e a imagem é linear. ROSENFELD; THURSTON (1971) apresentam a ideia de suprimir os problemas com ruído aplicando um filtro de suavização e assim melhorar o processo de diferenciação. Essa suavização é obtida alterando o valor de cada píxel para um valor que represente a média dos píxeis naquela região. Estes filtros contudo, tem como contraponto o fato de não ser adaptáveis ao tamanho da imagem e mesmo depois de aplicá-los não existe garantia de que o ruído foi removido. Precisa-se deixar claro que a redução de ruído traz como efeito negativo a perda de informação.

Estes filtros são uma abstração daquilo que se imagina como uma borda. Note que os pesos dos elementos na diagonal são menores que os pesos na horizontal e vertical. Cabe salientar que esta abordagem utiliza convolução de um *template* sobre a imagem. A convolução é um processo de fácil implementação mas que tem um custo computacional alto. Nesta categoria de detectores de borda baseado em *template* enquadra-se também o filtro proposto por KIRSCH (1971).

2.2.3 Detector de Marr-Hildreth

O detector de Marr-Hildreth é uma tentativa de unir o conhecimento da época no que se refere a visão biológica em um modelo que pudesse ser utilizado na visão computacional. O objetivo era produzir uma primitiva, mas rica descrição de imagens que pudesse ser utilizada para determinar a refletência e a iluminância de uma superfície, bem como determinar sua orientação e sua distância com relação ao observador/câmera (MARR; HILDRETH, 1980). Nesta pesquisa, os autores identificaram cinco pontos principais do sistema visual dos mamíferos. São eles:

- Em imagens naturais as características de interesse ocorrem em diversas escalas e não é possível criar um único operador para trabalhar em todas elas. Assim, é necessário que sejam construídos vários operadores e seus resultados sejam combinados para obter resultados satisfatórios.

- Uma cena natural não parece consistir de padrões de difração ou outros efeitos de onda, e por este motivo podem ser necessárias técnicas de suavização.
- O filtro de suavização que mais se aproxima da visão dos mamíferos é o gaussiano.
- Uma borda (mudança brusca de intensidade) é representada pelos extremos da primeira derivada, ou ao que corresponde aos pontos de passagem por zero na segunda derivada.
- O operador invariante à rotação de mais baixa ordem é o Laplaciano.

Tendo por base estes pontos Marr e Hildreth quebram o processo de detecção de bordas nas etapas de convoluir a imagem com uma função gaussiana bi-dimensional (horizontal e vertical), calcular a laplaciana da imagem resultante e por fim encontrar os píxeis que representam bordas (onde existe passagem por 0). Para PARKER (2010) este não é um algoritmo difícil de implementar, mas é, no entanto, mais difícil que os que usam template simples.

Matematicamente a convolução de uma imagem com uma função Gaussiana G é dada pela Eq. 1:

$$I * G(i, j) = \sum_n \sum_m I(n, m) G(i - n, j - m) \quad (1)$$

onde a função G é dada pela Eq. 2:

$$G_\sigma(x, y) = \sigma^2 e^{-\left(\frac{x^2 + y^2}{2\sigma^2}\right)} \quad (2)$$

Depois disso aplica-se o operador Laplaciano. Na prática, como a ordem das operações não altera o resultado, costuma-se computar o Laplaciano da Gaussiana e depois convoluí-lo com a imagem original. O Laplaciano da Gaussiana, representado na literatura pela sigla LoG (*log of gaussian*) é definido pela Eq. 3, onde $r = \sqrt{x^2 + y^2}$. Conforme apresentado em PARKER (2010) os resultados são significativamente afetados variando-se os parâmetros de tamanho e desvio padrão da função gaussiana. Para atender ao ponto listado no início desta secção que afirma que a visão do mamíferos ocorre em várias escalas, este método utiliza dois ou mais filtros LoG, combinando-os de modo que somente sejam aceitas as bordas detectadas em mais de uma escala.

$$\nabla^2 G_\sigma = \frac{r^2 - 2\sigma^2}{\sigma^4} e^{-\frac{r^2}{2\sigma^2}} \quad (3)$$

2.2.4 Detector de Canny

O detector de bordas de Canny é provavelmente um dos mais conhecidos. Trata-se de um detector pensado para atender a três metas principais:

- Taxa de erros: o detector deve encontrar somente bordas e deve encontrar todas elas. Nenhuma borda deve ser ignorada.
- Localização: a distância entre a borda detectada e a borda real deve ser a menor possível.
- Resultado: o detector não deve identificar vários píxeis de borda quando apenas uma borda existe.

Assume-se que as bordas degrau estão sujeitas a ruído branco Gaussiano. O detector de bordas é uma máscara de convolução que suaviza o ruído e localiza a borda (CANNY, 1986). O problema maior é identificar um filtro que otimiza estes três critérios previamente estabelecidos. Segundo PARKER (2010) usualmente faz-se uma aproximação da melhor relação sinal-ruído utilizando a primeira derivada da Gaussiana. Tendo que a Gaussiana é dada pela Eq. 4 a primeira derivada é dada pela Eq. 5. A gaussiana bidimensional é dada pela Eq. 2.

$$G(x) = e^{-\frac{x^2}{2\sigma^2}} \quad (4)$$

$$G'(x) = \left(-\frac{x}{\sigma^2}\right)e^{-\left(\frac{x^2}{2\sigma^2}\right)} \quad (5)$$

Para PARKER (2010) a melhor aproximação para o filtro ótimo de Canny é G' e assim, ao convoluir a imagem de entrada com G' obtém-se uma imagem E com as bordas realçadas, mesmo que a original apresente ruído. A convolução em duas dimensões pode, no entanto, ter uma complexidade alta e por este motivo computa-se cada uma das dimensões separadamente. Por fim, o filtro de Canny apresenta os seguintes passos:

1. Ler uma imagem de entrada para o processamento I .
2. Criar uma máscara gaussiana uni-dimensional G para convoluir com I . O desvio padrão desta Gaussiana é um parâmetro para o detector de bordas.
3. Criar uma máscara uni-direcional para a primeira derivada da Gaussiana nas direções x e y . Respectivamente G_x e G_y . O mesmo desvio padrão do passo dois é utilizado.
4. convoluir G com I passando por todas as linhas obtendo I_x e colunas obtendo I_y .
5. convoluir I_x com G_x para obter I'_x e I_y com G_y para obter I'_y .
6. Calcular a magnitude da resposta de borda combinando os componentes em x e y . Esta magnitude no ponto pode ser computada como

$$M(x, y) = \sqrt{I'_x(x, y)^2 + I'_y(x, y)^2}$$

Cada píxel da imagem filtrada é processado nesta maneira, a grandeza de gradiente é calculada para duas posições, uma de cada lado do píxel, e a magnitude no píxel deve ser maior que seus vizinhos. Segundo PARKER (2010) geralmente, existem oito casos importantes para verificar. Alguns atalhos que podem ser utilizados para aumentar a eficiência, mas o método acima é, essencialmente, o que é usado na maioria das implementações do detector de bordas Canny.

Por fim, depois de calcular a amplitude em cada píxel, faz-se uma supressão dos píxels que não são máximas locais e uma limiarização (*thresholding*) utilizando dois filtros distintos. O primeiro filtro obtém somente os pontos com amplitude elevada. O segundo aceita amplitudes menores se, e somente se, estiverem próximos aos píxels de borda identificados pelo filtro anterior.

Como já introduzido anteriormente, linhas correspondem a extremos locais da imagem em escala de cinza e são de grande valia para a identificação de algumas características em imagens, sendo utilizadas, por exemplo, para a identificação de rios e estradas no sensoriamento remoto. A maior parte dos processos de detecção de linhas consiste em algoritmos de afinamento criados para imagens binárias que, em alguns casos, podem ser utilizados também em escala de cinza. O principal problema destes detectores é que eles normalmente não são precisos e não fornecem bons resultados em imagens complexas (ZIOU; TABBONE et al., 1998).

A maioria dos trabalhos na área de detecção de fogo focam exclusivamente na detecção de bordas em imagens em escala de cinza utilizando Canny. OSKOEI; HU (2010) menciona que a detecção de bordas em imagens coloridas não vem recebendo a mesma atenção. Para eles, a diferença fundamental está exatamente no fato de um píxel em uma imagem colorida ser normalmente representado por um vetor (de 3 a 4 componentes nos espaços de cor usualmente empregados) e em decorrência disto as funções precisam trabalhar com vetores ao invés de valores escalares.

A detecção de bordas em imagens coloridas pode ser dividida de acordo com duas técnicas (KOSCHAN; ABIDI, 2005):

- Técnicas monocromática tratando as informações dos canais de cores individualmente (separadamente cada componente do vetor de cor) e depois combinando os resultados obtidos.
- Técnicas que operam no espaço vetorial tratando a informação de cor como vetores em um espaço vetorial e executando funções vetoriais.

A adaptação de técnicas monocromáticas é a utilizada com maior frequência, obtendo, na maioria das vezes, resultados melhores que os obtidos quando as imagens são trabalhadas em escala de cinza. Contudo, KOSCHAN; ABIDI (2005) trabalham com um píxel diretamente no espaço vetorial. As técnicas anteriores buscavam extrair os contornos em

imagens coloridas quebrando as imagens em canais e aplicando o filtro de Canny sobre o canal da luminosidade (que tinha resultados pífios pois algumas bordas existentes não são detectadas operando só em um canal), aplicando Canny em todos os canais individualmente e combinando os resultados (o que resultava em bordas imprecisas) ou ainda em uma técnica que calculava o gradiente de cada canal, somava os valores resultantes e aplicava uma limiarização para definir a presença das bordas (KOSCHAN; ABIDI, 2005).

Entre os detectores que apresentam melhores resultados na detecção de bordas em imagens coloridas está o filtro proposto em CUMANI (1991). Este filtro sugere procedimentos com base nas segundas derivadas parciais das funções de imagem. Esta técnica pode em teoria ser aplicada em imagens com um número qualquer de canais. Neste contexto, assume-se sempre que uma métrica Euclidiana existe para o espaço vetor n -dimensional. Os resultados podem variar conforme a imagem que está sendo analisada e nem sempre é possível fazer uma generalização sobre os resultados obtidos (KOSCHAN; ABIDI, 2005).

2.3 Segmentação

A segmentação de imagens pode ser entendida como uma matéria complementar da detecção de bordas. Na detecção de bordas faz-se uma segmentação da imagem com base nos limites ou bordas de cada objeto. Na segmentação busca-se agrupar píxeis que são similares de acordo com as propriedades de uma região. Na segmentação de uma imagem, segundo FORSYTH; PONCE (2002), busca-se obter sub-imagens agrupando, geralmente, píxeis com valores próximos na escala de cinza. No entanto, esta não é única abordagem possível para a segmentação.

2.3.1 Thresholding

A distribuição dos níveis de cinza pode ser utilizada para definir os níveis de *threshold* a serem utilizados para a imagem binária. Um histograma mapeia o número de píxeis em uma imagem que apresentam determinado valor de cinza como uma função dos níveis de cinza da imagem (FORSYTH; PONCE, 2002). Imagens reais raramente contêm bordas claras entre o *foreground* e o *background*, mas apresentam variações graduais na escala de cinza, resultando em picos e depressões no histograma. A técnica de segmentação utilizando *thresholding* é uma das mais antigas, simples e frequentemente utilizadas (PAL; PAL, 1993).

Objetos com valores de cinza próximos formam uma classe. O histograma de uma imagem apresenta picos para cada um dos objetos encontrados e provavelmente um pico maior que indica o *background*. Para diferenciar entre k classes de objetos é necessário definir k *thresholds*, cada um definido entre dois picos. FORSYTH; PONCE (2002) re-

forçam ainda que é importante ter claro que histogramas conseguem diferenciar apenas as classes de objetos, ignorando completamente qualquer informação espacial, o que indica que um tabuleiro de xadrez, que tem uma distribuição característica de blocos brancos e pretos terá o mesmo histograma que uma imagem gerada aleatoriamente com cores brancas e pretas.

2.3.2 Segmentação baseada em região

São facilmente perceptíveis as limitações de um procedimento de segmentação através de um *thresholding* da imagem com um histograma global. Por este motivo, desenvolveram-se algoritmos de segmentação que consideram os componentes conectados. O uso de componentes conectados adiciona a informação espacial aos dados extraídos. Essa conexão entre píxeis com níveis de cinza semelhantes é feita por um algoritmo que encontra grupos de píxeis conectados através dos seguintes passos:

1. Varrer a imagem binária da esquerda para a direita, de cima a baixo.
2. Se existe um píxel não marcado com o valor '1' ele deve receber uma marcação.
3. Varrer recursivamente os píxeis vizinhos deste e marcá-los com o mesmo valor, caso não estejam marcados com '1'.
4. O processo termina quando todos os píxeis de valor '1' são marcados.

ZHANG (1996), PAL; PAL (1993), PARKER (2010), FORSYTH; PONCE (2002) apresentam variações deste algoritmo buscando obter melhores resultados em termos de custo computacional e precisão da segmentação, mas a ideia essencial é mantida, embora busque-se fugir do uso de programas recursivos e utilizar algoritmos sequenciais. Duas abordagens principais podem ser destacadas (HOROWITZ; PAVLIDIS, 1976):

- Mesclando ou bottom-up: A imagem é dividida em várias pequenas regiões (eventualmente coincidindo com um único píxel), que são então fundidas para formar regiões maiores. Enquadram nesta categoria o trabalho de BRICE; FENNEMA (1970) e outros que buscam digitalizar sequencialmente, da esquerda para a direita, a forma de onda para determinar o intervalo mais longo de tal forma que uma aproximação é feita de acordo com uma tolerância previamente estabelecida.
- Dividindo ou top-down: A imagem é dividida sucessivamente em regiões cada vez menores até que certos critérios estão preenchidos. Na época em que HOROWITZ; PAVLIDIS (1976) foi publicado, poucos algoritmos utilizavam esta abordagem.

O algoritmo de *Split and Merge* é o mais simples para crescimento de regiões, ou, em inglês *region growing*. Uma apresentação é feita na figura 2. Neste algoritmo as regiões

são sequencialmente divididas e mescladas de acordo com certos predicados até chegar a uma situação em que nenhuma nova alteração possa ser feita seguindo esses predicados. Um exemplo de predicado simples pode ser: o valor dos píxeis em escala de cinza é o mesmo. Para começar, toda imagem é considerada uma região. Como nem todos os píxeis da imagem tem a mesma intensidade na escala de cinza, divide-se a imagem em 4 quadrantes. Em seguida faz-se uma tentativa de mesclar duas regiões adjacentes, mas nenhum dos quadrantes satisfaz este predicado, motivo pelo qual todos os quadrantes, exceto o esquerdo superior, são novamente divididos em 4 regiões menores. Faz-se uma nova tentativa de combinar algumas divisões e o processo é repetido até que todas as partes da imagem tenham sido unidas.

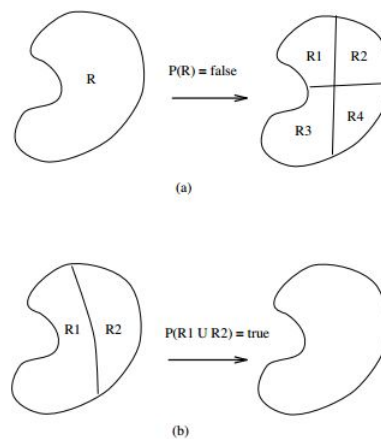


Figura 2: Passo a passo da segmentação por Split and Merge (MUBARAK, 1997)

Outro algoritmo bastante conhecido é como *Phagocyte Algorithm* ou *boundary melting*. Neste algoritmo a ideia central é relaxar os limites fracos entre duas regiões permitindo que elas sejam mescladas. Primeiro, pares de pontos vizinhos são ligados para formar um conjunto se tiverem o mesmo valor de cinza, ou assim chamadas, regiões atômicas. Ou seja, se os píxeis têm a mesma intensidade e são 4-conectados, então podem ser unidos formando regiões conexas. Na próxima etapa, heurística da fagocitose, regiões menores são englobadas por regiões maiores segundo o seguinte critério: Sejam P_1 e P_2 os comprimentos das bordas de duas regiões R_1 e R_2 , respectivamente. Seja $|I|$ o comprimento da fronteira I entre R_1 e R_2 . Seja $|W|$ o comprimento das partes "fracas" da borda, por exemplo, o comprimento de $W \subset I$ onde a diferença absoluta entre os níveis de cinza de R_1 e R_2 é menor do que um certo ϵ_1 . Por este algoritmo, pode-se afirmar que duas regiões, R_1 e R_2 , são ligadas entre si se

$$\frac{|W|}{\min(|P_1|, |P_2|)} > \epsilon_2$$

onde ϵ_2 é usualmente igual a $\frac{1}{2}$. Na terceira etapa, regiões adjacentes cuja fronteira é fraca são ligadas segundo o seguinte critério:

$$\frac{|W|}{|I|} > \epsilon_3$$

onde usualmente ϵ_3 é igual a $\frac{3}{4}$ (BRICE; FENNEMA, 1970).

De acordo com MUBARAK (1997), este método produz uma segmentação aceitável para imagens simples com poucos objetos e pouca textura. Duas heurísticas são utilizadas: *weakness* e *phagocyte*. Se a relação entre o número de pontos em bordas fracas e o número total de pontos na borda que separa uma região da outra satisfaz a um *threshold* pré-determinado, então essas regiões são mescladas. Contudo essa heurística tende a sobre-mesclar áreas e por este motivo utiliza-se uma heurística que faça um controle levando em conta o modelo da região resultante.

A semelhança LH entre as duas regiões é dada por

$$LH = \frac{P(H_1)}{P(H_2)}$$

e as regiões são mescladas se forem maiores que um *threshold* T .

2.4 Detecção de Movimento

Um dos principais objetivos da visão artificial é emular características do sistema visual humano, como a habilidade de reconhecer objetos e rastrear seu movimento em ambientes complexos. O primeiro passo para os softwares de rastreamento é detectar os objetos que se movem, classificar os pixels que o representam agrupando-os em regiões, conhecidos no campo da visão computacional pelo termo *blob*. Através deste agrupamento diminui-se a complexidade do problema, dada uma percepção global da cena.

A maior parte dos trabalhos publicados é voltada para aplicação em vídeos obtidos a partir de câmeras perfeitamente estáticas. Duas abordagens principais para o problema são apresentados na literatura científica: *image difference* e subtração de *background*. A primeira consiste em uma diferença limiarizada entre o *frame* no tempo t e o tempo $t - 1$, se mostrando extremamente performático em termos de custo computacional e fornecendo uma detecção do movimento entre dois quadros. Este método apresenta, no entanto, duas desvantagens bem conhecidas causadas pela taxa de quadros por segundo e a velocidade do objeto se movendo na cena: abertura do primeiro plano e detecções fantasmas (COLLINS et al., 2000). Com a intenção de resolver este problema KAMEDA; MINOH (1996) propõe uma variação do método utilizando 3 quadros fazendo a diferença entre os *frames* no tempo t e $t - 1$ e no tempo $t - 1$ e $t - 2$. Ao final as duas saídas são combinadas através do operador lógico "AND". MIGLIORE; MATTEUCCI; NACCARI

(2006) fazem uma ressalva à esta abordagem afirmando que se os objetos não tiverem textura suficiente ela pode não permitir uma detecção precisa do movimento e a posição do objeto não é estimada em tempo real. Para tratar estes problemas, foi proposta uma técnica que considera a diferença entre os quadros t e $t - 1$ e t e $t - 2$ de modo a evitar o surgimento de fantasmas que é apresentado pelas técnicas anteriores (COLLINS et al., 2000). Este sistema é amplamente utilizado em ambientes externos com imagens de baixa profundidade de campo.

Por outro lado, a detecção de movimento utilizando *background subtraction* utiliza-se de um modelo do plano de fundo para compará-lo com o quadro atual. Estes algoritmos resolvem os problemas apresentados pelos métodos que utilizam *frame difference* e usualmente gera bons resultados na segmentação dos objetos que apresentam movimento. Um tratamento especial é necessário para lidar com os variações na luminância e os objetos que iniciam ou terminam seu movimento durante a sequência de quadros. Isto inclui por exemplo pessoas andando e automóveis que param no meio da cena. Para resolver este problema várias técnicas foram propostas, dentre as quais pode-se destacar os métodos baseados em média e variância como *Weighted Moving Mean*, *Weighted Moving Variance*, *Adaptive Background Learning*, *Temporal Mean*, *Adaptive Median* apresentado em MCFARLANE; SCHOFIELD (1995) e *Temporal Median* proposto em STAUFFER; GRIMSON (1999). Entre os métodos baseados em mistura de gaussianas pode-se destacar as publicações STAUFFER; GRIMSON (1999) e BOUWMANS; EL BAF; VACHON (2008). Na literatura pode-se ainda encontrar diversas propostas que utilizam lógica Fuzzy, redes neurais ou modelos de cor e textura para separação entre o primeiro e segundo plano.

As mais simples fazem uma composição do *background* utilizando uma composição convexa dos píxeis no segundo plano no tempo $t - 1$ e os píxeis da imagem no tempo t . Assume-se geralmente que o segundo plano é menos dinâmico do que o objeto a ser rastreado e que este é distribuído de acordo com uma Gaussiana unimodal. É comum que esta modelagem do segundo plano falhe ao lidar com mudanças bruscas ou casos em que o fundo não possa ser dado por uma função Gaussiana unimodal. Esta premissa é considerada no método Pfinder apresentado em WREN et al. (1997).

Uma das técnicas mais populares de *background subtraction* e amplamente conhecida pela sigla MoG foi proposta em STAUFFER; GRIMSON (1999). A modelagem do *background* utilizando mistura de gaussianas em nível de píxel tornou-se muito popular devido à sua eficiência em modelar distribuições multi-modais do segundo plano, o que ocorre por exemplo com árvores balançando, ondas no oceano e reflexão da luz. Uma segunda característica importante é que ela se adapta bem às mudanças no fundo (MIGLIORE; MATTEUCCI; NACCARI, 2006).

Cada píxel é modelado como uma mistura de gaussianas e o modelo é gerado através de uma aproximação. As distribuições gaussianas são então avaliadas para determinar

quais apresentam a maior probabilidade de serem do *background*. Cada píxel é classificado como parte do plano de fundo de acordo com a distribuição gaussiana que o representa, permitindo a obtenção de resultados estáveis em rastreadores *outdoor*, onde a confiança depende das variações na iluminação, movimentos desordenados repetitivos e variações na cena STAUFFER; GRIMSON (1999).

Para este trabalho foram testados alguns destes modelos com o objetivo de suprimir os falsos positivos remanescentes das técnicas de classificação de fogo, utilizando segmentação por cor, frequência de oscilação e formato do contorno. Como a pesquisa se propõe a suprimir as dificuldades com segundo plano dinâmico, condições distintas de iluminação, variação de posição e oscilação das imagens obtidas através de câmeras não estacionárias, espera-se que o método tenha baixa sensibilidade aos parâmetros da câmera e que possa ser facilmente incorporado em outras soluções. Considerações mais detalhadas a respeito dos resultados atingidos quando as técnicas de detecção de movimento foram aplicadas sobre o *dataset* introduzido na Sec. 1.2.0.1 são apresentados no Cap. 7.

3 TRABALHOS RELACIONADOS

As técnicas de processamento de vídeo para detecção e análise de fogo são relativamente novas quando comparadas com outras técnicas dentro da visão computacional. Considerando-se o problema dentro de um contexto maior de aplicações, seja dentro da indústria ou mesmo em sistemas de vigilância, é fácil perceber a importância de uma detecção rápida de incêndios e outros incidentes envolvendo fogo não controlado. Existe uma necessidade de análise inteligente de conteúdo em vídeo para auxiliar vigilantes a identificar os eventos logo que eles ocorram. Quanto antes o fogo é detectado, maiores são as chances de evitar que este se alastre, causando danos e colocando a vida de pessoas em risco.

Nos últimos anos, técnicas de detecção de fogo através de vídeo (em inglês *Video Fire Detection* - VFD) tem sido alternativas viáveis ou complementos para os sistemas de detecção de incêndio convencionais, como os baseadas em sensores de fumaça, e têm se mostrado úteis para resolver vários problemas relacionados com os sensores tradicionalmente utilizados. Sensores convencionais são geralmente limitados a ambientes fechados por exigirem uma proximidade com o fogo, o que os torna inadequados para ambientes como centros comerciais, aeroportos, estacionamentos ou mesmo florestas. Outra limitação dos sistemas de alarme de incêndio convencionais é que pode levar um longo tempo para partículas de carbono e fumaça acionem o detector, o que os autores costumam chamar de atraso do transporte. Já quando são utilizadas câmeras de vídeo para a detecção, este atraso não ocorre pois a câmera pode identificar as chamas imediatamente (ÇETIN et al., 2013).

Os primeiros estudos neste campo de pesquisa ocorreram no final da década de 90. Várias abordagens, de maior ou menor complexidade de implementação, foram propostas com a finalidade de criar um bom detector de fogo. Inicialmente existia uma ênfase maior na detecção da chama, mas atualmente vários trabalhos vem abordando também a fumaça. A maioria dos detectores atuais funciona combinando detectores de cor e características de borda com uma análise temporal do formato da chama e o seu comportamento. ÇETIN et al. (2013) ressalta que, no entanto, devido a variedade de formas e comportamentos que o fogo e a fumaça podem assumir (cor, transparência, temperatura ou cintilação)

muitos dos detectores considerados como estado da arte continuam apresentando falsos positivos. Devido a estas características, aliadas a outros fatores como o ruído, sombras e mudanças de iluminação, a detecção de fogo permanece sendo um grande desafio para os pesquisadores.

Apesar de ser um assunto relativamente novo, a detecção de fogo baseada em vídeo já produziu resultados promissores. No entanto não é um problema completamente resolvido, assim como a maior parte dos problemas de visão computacional. O comportamento da fumaça e das chamas de um incêndio não controlado varia com a distância e iluminação. Além disso, diferentes câmeras podem produzir diferentes imagens e vídeos para a mesma cena por causa de suas configurações internas e algoritmos (ÇETIN et al., 2013).

3.1 Detecção de Cor

A grande maioria dos detectores de chama tem na cor uma informação primordial para seu funcionamento. A maioria dos filtros utiliza o esquema de cor RGB, e em alguns casos, ele é combinado com HSI/HSV, como por exemplo em CHEN; HE; WANG (2010), CHEN; WU; CHIOU (2004), GÜNAY et al. (2010). Isto se deve, segundo ÇETIN et al. (2013) ao fato de a maioria das câmeras trabalharem com RGB e existir uma associação espectral óbvia com este espaço de cor. Da mesma forma trabalha-se com a ideia de que áreas de fumaça apresentam níveis próximos em cada um dos canais de cor quando trabalhando com o esquema de cor RGB.

A cor da chama vista pelo olho humano é geralmente algo entre vermelho e amarelo. A maioria das técnicas de classificação de fogo com base na cor partem deste pressuposto. CHEN; WU; CHIOU (2004) propõem um modelo de classificação utilizando o espaço de cor RGB, que aplica avaliação da crominância e medidas de desordem de forma a extrair os pixels que representam regiões de chama ou fumaça. O método utiliza valores de *threshold* fixados empiricamente. Um pixel pode ser considerado fogo se atender as condições da Eq. 6 onde S_T e R_T são os limiares de saturação e vermelho. A mesma equação foi utilizada posteriormente em CHENEVERT; BRECKON; GASZCZAK (2011) para separar segmentar as regiões pela cor e então extrair dados da textura.

$$P(x, y) = \begin{cases} R > R_T \\ R \geq G > B \\ S \geq ((255 - R) \times S_T) / R_T \end{cases} \quad (6)$$

PHILLIPS III; SHAH; VITORIA LOBO (2002) utiliza a informação do espaço de cores RGB e a variação temporal dos pixels para a detecção do fogo. O processo é realizado em 3 etapas:

1. Criação de pares de imagens em que cada par consiste em uma imagem colorida

contendo chamas e uma máscara binária que aponta onde o fogo está presente. Dez imagens são utilizadas.

2. Construção de um histograma de cor onde para cada píxel marcado como chama uma distribuição gaussiana ($\sigma = 2$) das cores é adicionada. Para os píxels marcados como não fogo uma distribuição gaussiana menor ($\sigma = 1$) é subtraída do histograma.
3. *Threshold* do histograma de forma a resultar em uma função que recebendo os valores RGB de um píxel retorna um valor booleano que indica se o mesmo é fogo ou não.

Como o fogo pode ser translúcido e, em alguns casos, dispersar-se de forma a não ser detectável através da técnica proposta, uma média das estimativas de um curto período de tempo é utilizada. De forma resumida, somam-se as saídas do píxel durante n frames da sequência e divide-se por n . Através do método de pesquisa experimental mostrou-se possível utilizar um n entre 3 e 7 com vídeos gravados a 30 FPS. A média deve ser maior que $\frac{1}{2}$ para que o píxel seja considerado como fogo (PHILLIPS III; SHAH; VITORIA LOBO, 2002).

HORNG; PENG; CHEN (2005), por sua vez, faz uma conversão de RGB para HSI utilizando uma equação própria, mais conveniente para o processamento, utilizando uma escala de 1 à 360 para cada canal HSI. Em seguida, é aplicado um método de segmentação pela cor para classificar regiões com aparência semelhante à fogo, removendo posteriormente os píxels saturados ou com pouca intensidade, tais como fumaça e regiões com reflexão da luz. Como os autores não fornecem maiores detalhes sobre como os valores de *threshold* foram definidos acredita-se que o fizeram de forma empírica. Depois desta classificação, espera-se que as regiões tenham sido propriamente classificadas. Contudo, faz-se um tratamento de falsos positivos considerando principalmente dois fatores:

1. Devido à reflexão das chamas alguns objetos tem a sua aparência alterada de forma que as características de crominância ou luminosidade (*color shift*).
2. Objetos com cores semelhantes ao fogo são erroneamente classificados (falsos positivos).

Mesmo após a classificação através da cor e o uso de *frame subtraction* para remover objetos estáticos com cor de fogo muitas regiões continuam sendo classificadas erroneamente. Isto pode não ser um problema se o objetivo for apenas detectar a presença ou não de chamas, contudo, se o objetivo for extrair dados sobre a localização das chamas, então estes falsos positivos precisam ser eliminados. Ao examinar cuidadosamente o fenômeno de *color shift* foi constatado que, dado que a cor alaranjada é um reflexo da cor do fogo

na fonte, sua intensidade e saturação devem ser inferiores às aquelas encontradas no fogo verdadeiro (HORNG; PENG; CHEN, 2005).

Continuando com as técnicas de classificação baseadas em cor, (CELIK et al., 2007) propõem um modelo estatístico de cor combinado com subtração de *background* para fazer a segmentação das áreas de fogo. O plano de fundo é modelado como uma gaussiana unimodal. Para gerar o modelo estatístico de cor, os autores coletaram 150 imagens da internet e as segmentaram em regiões de fogo e não-fogo. Cinco regras são aplicadas comparando o canal de cor vermelho (R) em cada píxel com a média na imagem e com os outros canais de cor. O modelo proposto utiliza ainda um limiar no canal vermelho para eliminar falsos positivos e tratar efeitos decorrentes da variação na iluminação. Um píxel é considerado como fogo se atender todas as condições apresentadas na Eq. 7:

$$P(x, y) = \begin{cases} R > R_m \\ R > G > B \\ 0.25 \leq G/(R + 1) \leq 0.65 \\ 0.05 \leq B/(R + 1) \leq 0.45 \\ 0.20 \leq B/(G + 1) \leq 0.60 \end{cases} \quad (7)$$

CELIK; DEMIREL (2009) propõem o uso do *color space YCbCr* para detecção de fogo a partir da cor. Este método contrói um modelo genérico de classificação através de crominância para os píxels contendo chama. Cada píxel é normalizado conforme o valor da intensidade. Os autores argumentam que o algoritmo proposto utiliza o espaço de cor YCbCr para separar a luminância da crominância de forma mais eficaz que outros espaços de cor. O desempenho do algoritmo proposto é testado em dois conjuntos de imagens, um dos quais contém fogo. Além disso, o modelo de cores proposto pode ser utilizado para a detecção em tempo real de fogo em vídeos. O modelo proposto apresenta uma taxa de falsos positivos de 31,5%.

Apesar do espaço de cor RGB poder ser usado para a classificação, ele apresenta a desvantagem de depender da iluminação. Isso significa que, se a iluminação da imagem mudar, a classificação do píxel provavelmente falhará. Além disso, não é possível separar o valor de um píxel em intensidade e crominância. A crominância pode ser utilizada em modelos de cor de fogo em vez de modelar a sua intensidade. Isso dá a representação mais robusta para píxels que representam chamas. Por isso é necessário converter a imagem RGB para um espaço de cor onde a separação entre a intensidade e crominância é mais discriminante. Por causa da conversão linear entre RGB e YCbCr utilizou-se YCbCr para modelar os píxels de fogo (CELIK; DEMIREL, 2009). A conversão de RGB para YCbCr é obtida através da Eq. 8 onde Y é a luminância e Cb e Cr são respectivamente a crominância azul e a crominância vermelha:

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2568 & 0.5041 & 0.0979 \\ 0.1482 & 0.2910 & 0.4392 \\ 0.4392 & 0.3678 & 0.0714 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \end{bmatrix} + \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} \quad (8)$$

O passo seguinte é a obtenção das médias para cada canal no *frame* e a validação da cor de acordo com as regras apresentadas na Eq. 9.

$$P(x, y) = \begin{cases} Y > Cb \\ Cr > Cb \\ Y > Y_{mean} \\ Cb < Cb_{mean} \\ Cr > Cr_{mean} \\ |Cb - Cr| > \tau \end{cases} \quad (9)$$

O valor de τ é uma constante determinada através da análise de uma curva ROC sobre um *dataset* de 1000 imagens, cuja segmentação do *ground truth* foi feita manualmente. Para cada valor de τ a taxa de falsos positivos e falsos negativos foi calculada. Considerando o custo do falso negativo maior que o do falso positivo os autores obtaram por utilizar $\tau = 40$, o que garantiu uma taxa de detecção de 90% e uma taxa de falsos positivos de 40%. Todas as imagens utilizadas eram diurnas, de ambientes externos e foram obtidas por câmeras instaladas em helicóptero. (CELIK; DEMIREL, 2009)

Um modelo de segmentação por cor utilizando o *color space Lab* é proposto em CELIK (2010). O *color space Lab* tem cores uniformemente distribuídas (cromaticidade uniforme). A sua distribuição é perceptualmente uniforme, a luminosidade é separada da cromaticidade e é muitas vezes apresentada como a representação dos contrastes cromáticos naturais do ser humano. A letra L na sigla corresponde à luminosidade (*lightness*) e a e b correspondem aos componentes de cor, baseados no modelo não linear de coordenadas CIE XYZ, sendo que a vai de verde a vermelho e b vai de azul até amarelo onde as duas dimensões são combinadas originando novas cores. As regras de classificação de um píxel como fogo com base na cor conforme apresentadas por CELIK (2010) são dadas pela Eq. 10:

$$P(x, y) = \begin{cases} L > L_m \\ a > a_m \\ b > b_m \\ b > a \end{cases} \quad (10)$$

onde L_m , a_m e b_m representam os valores correspondentes à média de cada canal. Um píxel é considerado como fogo somente se este satisfizer todas as regras apresentadas na equação.

Um bom modelo de cor de fogo e uma boa segmentação dos pixels em movimento são essenciais devido ao seu papel fundamental nos sistemas de detecção de incêndio baseados na visão computacional. O CIE Lab foi escolhido porque é um espaço de cor uniforme perceptivamente, tornando assim possível representar informações de cor de fogo melhor do que outros espaços de cor. Os pixels em movimento são detectados através da aplicação de um algoritmo de *background subtraction*. Os pixels que apresentam movimento são analisados em quadros consecutivos para levantar um alarme de fogo. A fim de encontrar a correlação entre a L , a e b um conjunto de 500 imagens RGB foram coletados a partir da Internet. Em seguida, cada imagem foi manualmente segmentada para identificar todas as regiões de chamas. Regiões de incêndio segmentadas são convertidos para o *color space* Lab. Um histograma dos pixels de fogo é criado para cada um dos três planos de cor (CELIK, 2010).

Em um processo que CELIK (2010) chama de quantificação, estes histogramas são agrupados em níveis de probabilidade. Através de experimentação foi definido que uma quantização em 24 níveis oferecia bons resultados. A tabela de consulta é criada para cada par de 24 níveis quantificados para descrever a probabilidade de que qualquer par de L , a e b pertença a um incêndio. Por fim o pixel é definido como fogo ou não é definido pela Eq. 11, onde α é um limiar definido manualmente com base em uma curva ROC. O *dataset* utilizado para validação dos resultados é o mesmo utilizado em CELIK; DEMIREL (2009).

$$P(x, y) = \begin{cases} 1, & \text{if } P(Lab) \geq \alpha, \\ 0, & \text{else} \end{cases} \quad (11)$$

Já NGUYEN-TI; NGUYEN-PHUC; DO-HONG (2013) combinam nove regras utilizando tanto o espaço de cor RGB quanto o espaço de cor YCbCr com o objetivo de minimizar falsos positivos. As regras visam eliminar regiões de pele humana, definir como regiões de fogo apenas as regiões que são assim detectadas em ambos os espaços de cor. Estas regiões se referem à pele humana e amarelo puro, ao mesmo tempo que não excluem regiões brancas que geralmente se formam no centro das chamas. Depois disso é aplicado um algoritmo de *Region Growing* para tolerar pixels próximos que tenham aparência parecida mas não tenham sido classificados como fogo dentro de alguma das regras do filtro de cor. Embora não sejam fornecidas estatísticas NGUYEN-TI; NGUYEN-PHUC; DO-HONG (2013) afirmam que mesmo combinando as nove regras ainda é provável que muitas regiões classificadas como fogo não correspondam exatamente a fogo. Por este motivo são combinadas técnicas de *Frame Subtraction* e *Region Growing* e acompanham a oscilação com base na variação horizontal e vertical das regiões classificadas.

Além disso, MARBACH; LOEPFE; BRUPBACHER (2006) trabalha a ideia de que o fogo é geralmente a maior fonte de luz, motivo pelo qual a luminância dos pixels perto do fogo tendem para o valor máximo permitido pela câmera (usualmente 255), atingindo, na maioria dos casos, o nível de saturação. As duas propriedades conhecidas do fogo, oscilar entre 1 e 10 Hz e atingir a luminância máxima, são usadas para modelar o algoritmo, que explora a luminância através do esquema de cores YUV. No YUV, o Y representa a luminância enquanto U e V representam a crominância. A primeira parte do processo inclui acumular a diferença entre os *frames* de forma a encontrar uma região de interesse que defina as partes que apresentam oscilação. Depois de extrair as ROIs uma classificação por cor é utilizada sobre os valores acumulados no tempo.

A maioria dos trabalhos publicados utiliza o *color space* RGB. Em alguns trabalhos utiliza-se também soluções baseadas em HSI/HSV como é o caso de CHEN; HE; WANG (2010); CHEN; WU; CHIOU (2004) e GÜNAY et al. (2010). De acordo com ÇETIN et al. (2013) isto se deve ao fato de a maioria das câmeras trabalharem por padrão com este sistema de cor e pelo fato de RGB apresentar uma associação espectral clara com a visão humana. Alguns autores propõe abordagens diferentes, como PHILLIPS III; SHAH; VITORIA LOBO (2002) que faz uso de histogramas e valores de cor que correspondem a uma gaussiana da região da chama.

3.2 Detecção de Movimento

É seguro afirmar que a cor não pode, por si só, classificar uma determinada região da imagem como fogo ou não, pois existe uma enorme variação na cor, densidade, quantidade de luz gerada pelo fogo, condições de iluminação do ambiente ou mesmo elementos do background que podem influenciar na cor capturada pela câmera e nos resultados da conversão de um *color space* para outro. Por este motivo, a maioria dos VFDs utiliza também outras características como o formato da chama, textura, movimento e relações espaciais.

Para determinar se o movimento é decorrente de uma chama ou fumaça é necessária uma análise de uma sequência de quadros. Usualmente, considerando que trabalha-se com câmeras estacionárias, utilizam-se técnicas *background subtraction*. Este é o caso de GÜNAY et al. (2010), CHEN; HE; WANG (2010), citettoreyin2005wavelet e JIAO; WEIR; YAN (2011). Outros trabalhos como GÓMEZ-RODRÍGUEZ et al. (2002), GOMEZ-RODRIGUEZ; ARRUE; OLLERO (2003) e KOLESOV et al. (2010) buscam abordar o problema do movimento utilizando fluxo óptico ou análise temporal como em LEE; HAN (2007).

ÇETIN et al. (2013) afirma que muitas das abordagens iniciais baseavam-se apenas na detecção de cor e movimento, mas esta abordagem usualmente resulta em falsos positivos. Uma análise mais aprofundada é necessária para atingir resultados mais acurados.

Uma destas análises é baseada na cintilação. HAMINS; YANG; KASHIWAGI (1992) apresenta um estudo detalhado do comportamento do fogo. De acordo com o estudo, um incêndio tem a propriedade de cintilação, aumentando e diminuindo a intensidade da luz emitida. Do ponto de vista de uma câmera, esta oscilação faz com que ocorra o aumento e a diminuição da luminância das imagens no decorrer do vídeo. A frequência de oscilação típica do fogo está na faixa de 1-10 Hz.

3.2.1 Análise da Oscilação no Domínio da Frequência

É amplamente aceita a ideia de que o fogo oscila entre 1 e 10 Hertz e que isto depende do material que está queimando. Isto que permite que filtros no domínio da frequência sejam utilizados para distinguir fogo de objetos em movimento presentes na cena. Os métodos baseados nesta característica utilizam o comportamento do píxel ao longo do tempo. Esta premissa não é válida para os primeiros instantes, quando o fogo está iniciando e apresenta um comportamento completamente aleatório. Neste estágio inicial é impossível perceber uma oscilação constante na frequência. Para a fumaça também não é possível obter resultados precisos baseados apenas na frequência (ÇETIN et al., 2013).

HAMINS; YANG; KASHIWAGI (1992) apresenta um estudo experimental realizado com o objetivo de estudar a liberação de calor, diâmetro das chamas e a frequência de pulsação do fogo considerando vários combustíveis gasosos e líquidos. No processo foram usados vídeos gravados a 30FPS. A frequência de pulsação foi determinada a partir da contagem do número de *frames* necessários para que o vórtex completasse 50 ciclos. Este processo foi repetido em diversas partes do vídeo para estimar a média e o desvio padrão da frequência de pulsação. O experimento demonstrou ainda que a frequência é fracamente correlacionada com a velocidade de liberação do combustível, variando no máximo 1,5 Hz.

A frequência de oscilação é utilizada para diferenciar chamas de outras fontes de luz em movimento através da realização de uma análise de frequência. No entanto, de acordo com o teorema da amostragem de Nyquist–Shannon, a frequência de amostragem de um sinal deve ser maior do que duas vezes a frequência do sinal. Infelizmente, a taxa de captura de vídeo é, em geral, a 25 Hz (25FPS) na maioria dos sistemas de vigilância. Esta amostragem só pode revelar variações de até 12,5 Hz, e não pode cumprir a exigência de amostragem. Além disso, a transformação de sinais do domínio de tempo para o domínio da frequência é custosa e afeta a eficiência do algoritmo de detecção (CHEN; HE; WANG, 2010).

TÖREYIN et al. (2006) utilizam a análise no domínio da frequência para acompanhar o comportamento do valor no píxel dentro da região que foi inicialmente classificada como chama com base na cor e movimento. Neste caso, os autores fazem uma ressalva afirmando que para funcionar adequadamente o vídeo deve ser capturado em no mínimo

20 FPS. Sabe-se contudo que quando ocorre *undersampling* existe um efeito de *aliasing*, o que sugere que com um devido tratamento ainda seja possível fazer a filtragem no domínio da frequência. Para a filtragem aplicou-se o filtro sobre o canal vermelho do *color scheme* RGB. Um filtro de dois estágios que filtra as altas frequências. Se o píxel pertencer ao contorno da chama então seu valor oscilará várias vezes. Se for apenas um objeto qualquer então espera-se que apenas um pico seja gerado. O número de passagens por zero durante alguns segundos (não especificado no *paper*) é utilizado para classificar os píxels utilizando um *threshold* qualquer.

É natural que exista uma associação entre fogo e fumaça pois a presença do segundo é geralmente um indicativo de que o primeiro também está presente, mesmo que em uma região fora do campo de visão da câmera. TOREYIN et al. (2005) afirmam que a oscilação/pulsação do fogo induz nas bordas da fumaça uma oscilação entre 1 e 3 Hertz. Nos testes foram utilizados vídeos gravados 5 *frames* por segundo e utiliza-se um banco de filtros de frequência sobre os canais de cor YUV que permite pegar frequências entre 0.625 e 2.5Hz.

A figura 3 apresenta o filtro de dois estágios utilizado. $X_n[k, l]$ representa a variação dos descritores de Fourier entre o *frame* n e $n + 1$. A decomposição em dois canais constitui-se de um filtro passa-alta (-0.25, 0.5, -0.25) e um filtro passa baixa (0.25, 0.5, 0.25). Se o píxel apresenta alta frequência d_e e e_n devem ser diferentes de zero. Pelo contrário, se o *frame* n permanece estável com relação ao *frame* $n + 1$ então estes sinais devem ser próximos à zero. A taxa de cruzamentos por zero de d_n e e_n durante alguns segundos é utilizada para caracterizar as regiões de fogo. Se este número for maior que um *threshold* predeterminado (TOREYIN et al., 2005).

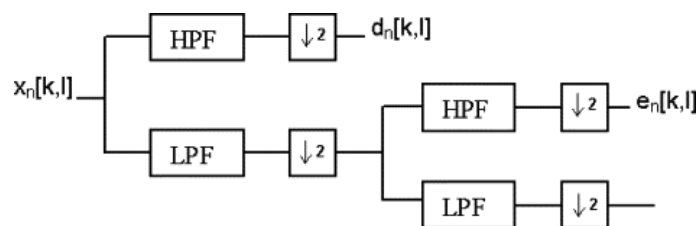


Figura 3: Filtro passa banda de dois estágios utilizado em ZHANG et al. (2008) e TÖREYIN et al. (2006). HPF e LPF representam respectivamente filtros passa-alta e passa-baixa.

Na última etapa, a convexidade com o formato de regiões de fumaça é verificada. A fumaça de um fogo não controlado se expande no tempo resultando em regiões com bordas convexas. Os limites das regiões que contêm "píxels candidatos" são verificados quanto à sua convexidade ao longo de linhas verticais e horizontais. Na implementação foram utilizadas cinco linhas horizontais e cinco linhas verticais. A análise consiste em verificar se os píxels em cada uma das linhas pertencem à região em movimento ou não. Pelo menos três píxels consecutivos nas linhas que se cruzam nas regiões movimento

devem pertencer ao *background* a fim de verificar se a região em movimento viola a condição de convexidade. Se ao longo de qualquer uma das linhas, a regra de convexidade não for cumprida, os pixels de fumaça na região são descartados. Essas pistas são então combinadas para dar uma decisão final. Se todos os critérios acima mencionados são atendidos para um pixel, o pixel é determinado como fumaça (TÖREYIN et al., 2006).

Já ZHANG et al. (2008) apresenta um sistema de detecção de fogo e fumaça voltado especificamente para detecção de fogo em florestas. A análise temporal no domínio da frequência segue um modelo semelhante ao apresentado por TÖREYIN et al. (2006) e apresentado na Fig. 3. Os valores dos pixels são extraídos do canal R no espaço de cor RGB. A amostragem utilizada é de 30 Hz (30 FPS) e os resultados apontam que um limiar de 3 Hz é adequado para detecção de fogo em vídeos.

3.2.2 Análise da Oscilação Utilizando Fluxo Óptico

Fluxo óptico refere-se ao problema de estimar um vetor de distância que mostre o deslocamento de um ponto em uma sequência de imagens. Na formulação do problema de fluxo óptico assume-se que um ponto, definido por (x_1, y_1) de uma imagem obtida no instante t deve ter um correspondente no instante $t + 1$. O deslocamento local, assumidamente pequeno, permite estimar o deslocamento do objeto observado nos *frames* subsequentes (CYGANER; SIEBERT, 2011).

As duas abordagens mais usuais para o problema do fluxo óptico são dadas por HORN; SCHUNCK (1981) e LUCAS; KANADE et al. (1981). O algoritmo de Horn-Schunck é o mais simples e embasa-se na ideia de que o brilho de um objeto é sempre constante (*Brightness Constancy Assumption*), fazendo com que seu movimento possa ser identificado diretamente pelo valor do pixel. Assim, o gradiente de intensidade é utilizado para encontrar a melhor correspondência entre dois pixels. Já o método de Lucas-Kanade utiliza descritores Eigenvalue para encontrar a correspondência entre características da imagem e determinar o deslocamento.

Para detecção de fogo, KOLESOV et al. (2010) propõe o uso de fluxo óptico OMT (Optimal Mass Transport) como um descritor de baixo nível. O processo de detecção é tratado como um problema de classificação Bayesiana onde as entradas são compostas pelas velocidades OMT e pela cor no espaço RGB. O classificador é implementado como uma rede neural de 1 camada oculta, sendo que os resultados apresentam a possibilidade de cada pixel ser de fogo ou fumaça.

O problema de transporte de massa foi inicialmente formulado por Gaspar Monge em 1781, e focado em encontrar um caminho ótimo, no sentido de reduzir o custo do transporte, de transportar um volume fixo de terra de um ponto para outro. Uma formulação moderna para este problema foi proposta no trabalho de Kantorovich em 1947, sendo por este motivo também conhecido como problema de Monge–Kantorovich

[...] A métrica de Kantorovich–Wasserstein define a distância entre duas densidades ao computar o menor custo de transporte de um domínio para outro (KOLESOV et al., 2010).

As abordagens que partem do pressuposto de que o brilho dos objetos é constante entre os quadros não são adequadas para a detecção de fogo e fumaça. Este pressuposto é válido para objetos rígidos com uma superfície Lambertiana, mas falha para materiais líquidos ou gasosos. Na visão computacional, estes materiais devem ser modelados por texturas dinâmicas. As texturas dinâmicas típicas de fumaça e fogo possuem dinâmica intrínseca e, portanto, não podem ser capturadas de forma confiável utilizando fluxo óptico na sua forma tradicional. Deve-se levar em conta também que a região de fogo/fumaça tende a apresentar um fluxo muito mais rápido do que a área em torno dele (KOLESOV et al., 2010).

KOLESOV et al. (2010) buscam criar um modelo de fluxo que represente melhor o comportamento de fogo e fumaça baseando-se nos atributos físicos deste processo:

Um atributo simples é que fogo e fumaça tendem a conservar intensidade tomada como uma massa generalizada e mover a massa de uma forma otimizada. Assim, uma restrição matemática apropriada não é a preservação da intensidade, mas sim da conservação de massa ou conservação brilho (KOLESOV et al., 2010).

A soma da energia total da imagem com a preservação da massa é utilizada como entrada em uma rede neural para ser classificada como fogo ou não-fogo. Sem fornecer maiores detalhes, os autores afirmam que apenas seis *frames* de um vídeo, segmentados manualmente, foram utilizados para o treinamento da rede neural. Os testes foram realizados no mesmo vídeo do qual os *frames* foram extraídos. Embora os autores assegurem que os resultados são promissores e apresentem imagens de classificações corretas, não são fornecidas quaisquer estatísticas sobre taxas de detecção, falsos positivos e falsos negativos.

Outra proposta utilizando fluxo óptico é apresentada em MUELLER et al. (2013), que trabalham considerando dois métodos de fluxo óptico desenhados especificamente para detecção de chamas: OMT para extração de textura dinâmica nas regiões onde o fogo apresenta variação de intensidade e uma abordagem clássica de fluxo óptico que busca explorar o fluxo óptico das áreas saturadas que os autores tratam por NSD (Non-Smoothed Data). A partir disso, características relacionadas com a magnitude e direção são extraídas para discriminar entre regiões de fogo e não fogo. Uma visão geral do método é apresentada na Fig. 4.

Sobre a variação direcional das áreas saturadas, MUELLER et al. (2013) defendem que o método proposto é capaz de diferenciar o movimento das bordas da chama das bordas de objetos rígidos presentes na cena e que apresentem movimento, ao quantificar a variância da direção do fluxo dos pixels em movimento. Conforme pode ser observado

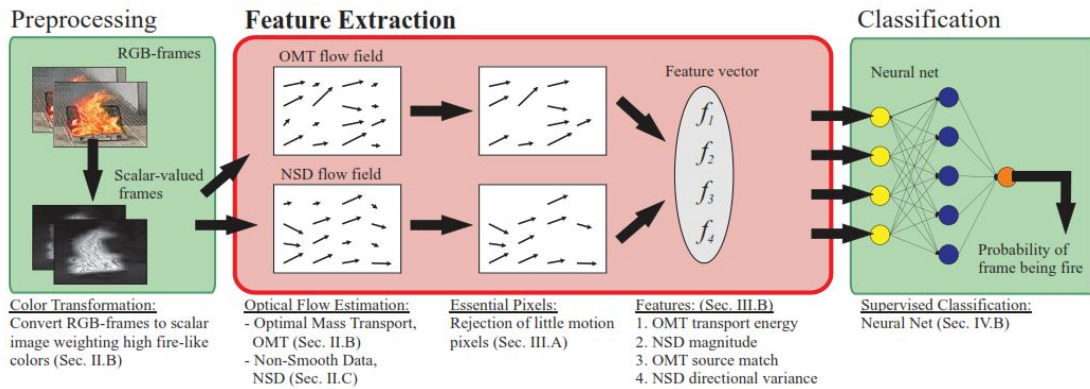


Figura 4: Método de detecção de fogo proposto por MUELLER et al. (2013). A ênfase do trabalho está na extração de características combinando transporte ótimo de massa e magnitude e variação na direção das áreas saturadas, criando um vetor 4D que é utilizado como entrada em uma rede neural.

na Fig. 5 uma variância alta sugere que movimento em várias direções (fogo), ao passo que uma variância pequena induz à um movimento uni-direcional (objeto rígido).

Sobre a variação direcional das áreas saturadas, MUELLER et al. (2013) defendem que o método proposto é capaz de diferenciar o movimento das bordas da chama das bordas de objetos rígidos presentes na cena e que apresentem movimento, ao quantificar a variância da direção do fluxo dos pixels em movimento. Conforme pode ser observado na Fig. 5, uma variância alta sugere que existe movimento em várias direções (fogo), ao passo que uma variância pequena induz à um movimento uni-direcional (objeto rígido).

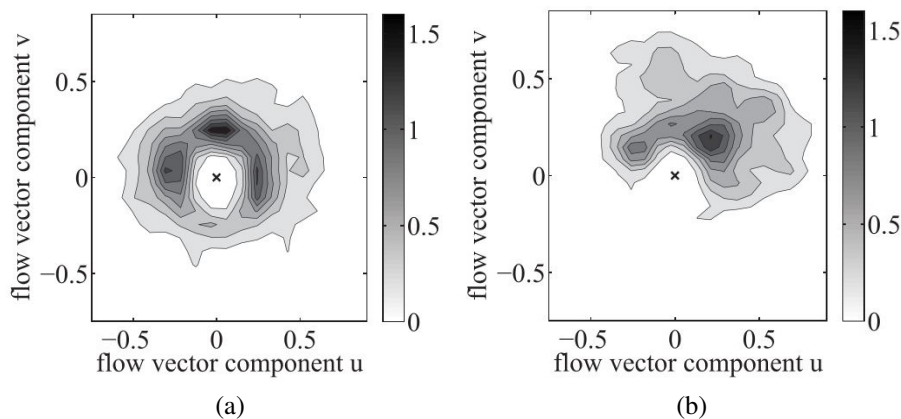


Figura 5: Histograma normalizado das direções do fluxo óptico utilizando NSD conforme apresentadas em MUELLER et al. (2013). Os dados foram obtidos a partir de dois vídeos, ambos apresentando regiões onde o brilho do objeto provoca saturação da imagem. Naturalmente o componente u corresponde ao deslocamento no eixo x e o componente v corresponde ao deslocamento do píxel no ponto y.

Os vídeos utilizados pra treinamento e testes são gravados a 30 FPS e

com resolução 240x360 píxels. De cada um dos 263 cenários (169 contendo fogo e 94 sem fogo), 10 quadros consecutivos são rotulados como *ground truth* fornecendo um banco de dados de teste de 2630 *frames*. Note-se que os cenários de fogo são escolhidos de forma a serem prováveis falsos positivos, ou seja, apresentam movimento e/ou objetos com cor de fogo, como carros, pessoas, folhas vermelhas e luzes. Uma rede neural foi treinada em quadros de 20 desses vídeos que não são utilizados nos testes. O resultado do teste mostra que fogo é detectado de forma confiável usando fluxo óptico. A taxa de falsos positivos é de 3,19% e a taxa de falsos negativos é de 3,55%. Visto que apenas as características do fluxo óptico são exploradas considera-se que o método é viável para aplicação em grandes bases de dados (MUELLER et al., 2013).

Em uma análise qualitativa, 4 cenários geraram erros de classificação produzindo falsos negativos (MUELLER et al., 2013):

- Ruído estruturado no vídeo, apresentando linhas horizontais faz com que o movimento do fogo não seja capturado de forma adequada;
- Distorção da cor do fogo pela câmera faz com que a estimativa de transporte de massa falhe.
- Áreas de chama muito pequena não geram movimento suficiente para serem detectadas através do método.
- Oclusão parcial do fogo, onde algum dos lados do fogo está oculto faz com que a variação das direções do movimento não possam ser estimadas, levando o método a produzir falsos negativos.

Embora apresentem algumas limitações claras, pode-se considerar que a detecção da oscilação do fogo utilizando fluxo óptico representa uma alternativa promissora. Esta abordagem é relativamente nova e portanto existem poucos trabalhos publicados. Além dos trabalhos de MUELLER et al. (2013) e KOLESOV et al. (2010), que focam especificamente em detecção de chamas, CHUNYU et al. (2010) propõem o uso de fluxo óptico conforme apresentado em LUCAS; KANADE et al. (1981) para a detecção de regiões de fumaça. Em linhas gerais, o método propõem encontrar as regiões de movimento utilizando fluxo óptico, fazer uma classificação baseada em oscilação e cor no *color space* RGB utilizando uma rede neural e disparar um alarme caso fumaça seja detectada.

3.2.3 Análise da Oscilação Utilizando Mineração de Dados

Também utilizando as características de oscilação do fogo LI; WU (2011) combinam um modelo estatístico de cor e mineração de padrões sequenciais para estabelecer uma

classificação dos pixels como fogo ou não-fogo. A mineração de dados é um processo de inferência de conhecimento de grande quantidade de dados e uma de suas áreas é a descoberta de padrões que ocorrem frequentemente em dados sequenciais. AGRAWAL; SRIKANT (1995) definem o problema de mineração de padrões sequências da seguinte forma:

Dado um conjunto de sequências, em que cada sequência é constituída por uma lista de elementos e cada elemento é constituído de um conjunto de itens, e dado um limiar de suporte mínimo especificado pelo utilizador, o objetivo da mineração de padrões sequenciais é encontrar todas as subsequências frequentes, isto é, as subsequências cuja frequência de ocorrência no conjunto de sequências é maior que o suporte mínimo. (AGRAWAL; SRIKANT, 1995).

LI; WU (2012) e LI; WU (2011) utilizam o algoritmo de mineração de padrões sequenciais *PrefixSpan* para identificar as sequências mais comuns que pixels de fogo assumem durante 8 *frames*. Para criar o que os autores chamam de "*fire rules*", os autores coletaram diversos vídeos da internet, segmentando manualmente as regiões que apresentam fogo. No sentido de reduzir as possíveis regras e diminuir o espaço de busca o valor de cada pixel é dividido por 8, fazendo com que os possíveis valores para cada canal de cor RGB passem de 2^8 para 2^5 , antes de ser utilizado para a mineração de sequências comuns áreas de chamas. Mais detalhes sobre o funcionamento do algoritmo *PrefixSpan* podem ser obtidos em PEI et al. (2001).

Assumindo que existem sequências que ocorrem com frequência um histórico de cada pixel é gerado e verifica-se se ele satisfaz as "*fire rules*" que foram anteriormente mineradas. Cabe salientar que uma vez minerados os padrões o algoritmo *PrefixSpan* não é mais utilizado. Embora os dados estatísticos apresentados sejam incompletos e não apresentem por exemplo a taxa de falsos positivos e falsos negativos, LI; WU (2012) asseguram que os resultados são melhores que os atingidos pelo detector proposto por CELIK et al. (2007). É impossível reproduzir o experimento apresentado visto que os resultados podem variar dependendo dos dados utilizados para a mineração das sequências, os quais não são disponibilizados publicamente.

3.2.4 Análise da Oscilação com Base no Formato da Chama

Após uma validação dos pixels com coloração de chama é possível extrair regiões e determinar as características destas regiões de forma a verificar se elas representam ou não fogo. YAMAGISHI; YAMAGUCHI (2000) transformam a imagem de entrada de RGB para HSV e verificam se a cor satisfaz as cores de chama. Os valores de H, S e V são definidos através do método experimental. Os dados de regiões que apresentam cor de fogo são então projetados em um plano HS (Matiz e Saturação) e a variância de H na região é calculada. A região que apresentar a maior variância é utilizada e uma elipse que

melhor acomoda a região é criada. Utilizando como base o maior eixo da elipse a região é normalizada para um tamanho previamente estipulado. Argumenta-se que isso permite detectar chamas mesmo que o tamanho da região em chamas varie significativamente de uma fonte para outra.

Quando uma região apresenta coloração de fogo durante vários quadros consecutivos é utilizada para encontrar o centro de massa das chamas e então desenhar um círculo utilizando com o raio o maior eixo da elipse. As regiões com coloração de fogo conectadas a este círculo são então extraídas, sendo sua área utilizada como parâmetro para análise da oscilação. A distância entre cada um dos pontos do contorno para o centro de massa é extraída durante vários *frames* consecutivos gerando uma matriz que é então convertida para o domínio da frequência através da transformada de Fourier, obtendo dados de amplitude e fase. Desta forma, consegue-se separar fogo de outras fontes artificiais de luz (YAMAGISHI; YAMAGUCHI, 2000).

Em caso de fogo, o componente de frequência tende a se concentrar no canto superior esquerdo (área de baixa frequência). Já em caso de luz artificial existe uma tendência de a distribuição ser uniforme. (YAMAGISHI; YAMAGUCHI, 2000)

No entanto, devido a uma série de condições que podem afetar diretamente o comportamento do fogo (YAMAGISHI; YAMAGUCHI, 2000) afirmam que é impossível através deste método fazer uma classificação precisa, optando então por treinar e classificar as regiões utilizando uma rede neural de três camadas, que tem como entrada um vetor de tamanho $n \times n$ que correspondem a cortes de 10 x 10 na matriz bidimensional produzida com a flutuação dos contornos. As possíveis saídas do classificador são: fogo e não-fogo.

Outra proposta de detecção de fogo baseada no formato da região classificada como apresentando cor de fogo é apresentada em ZHOU et al. (2010). O primeiro passo neste caso é a classificação de cor no *color space* RGB utilizando limiares estabelecidos a partir de resultados experimentais. Para remover regiões pequenas são utilizados os operadores de erosão e dilatação. A partir disso, a classificação pode ser definida em quatro etapas:

- Converter a imagem de entrada anteriormente classificada para uma em tons de cinza e outra binária;
- Executar operações de dilatação e erosão para suavizar as imagens;
- Remover regiões pequenas preenchendo as com os mesmos valores da região maior onde estão englobados;
- Classificar as regiões como chama ou não chama com base nos contornos obtidos através do detector de Canny utilizando os últimos 20 *frames*.

A partir dos dados do contorno a região é classificada como fogo quando satisfizer a regra 12 (ZHOU et al., 2010):

$$F = \begin{cases} P_S > P_{ST} \\ P_C > P_{CT} \\ P_R > P_{RT} \end{cases} \quad (12)$$

onde P_S , P_C e P_R são os percentuais de *frames* que apresentam aumento na área de fogo, *frames* que apresentam aumento no perímetro do fogo e a média de circularidade nos últimos 20 quadros processados. Neste caso P_{ST} , P_{CT} e P_{RT} são definidos através de resultados experimentais. Faz-se a ressalva de que esta abordagem é confiável apenas para a detecção dos estágios iniciais do fogo, quando existe um crescimento da região em chamas, não podendo ser utilizada em fogo já estabilizado. Outra limitação da abordagem proposta diz respeito a distância entre a câmera e a chama, que pode afetar os resultados devido a dependência de parâmetros estáticos (ZHOU et al., 2010).

3.3 Detecção Baseada em Textura

Além das abordagens que utilizam o movimento para melhorar a classificação dada pela cor, trabalhos recentes como CHENEBERT; BRECKON; GASZCZAK (2011) e BORGES; IZQUIERDO (2010) passaram a utilizar também informações de textura para conseguir uma classificação utilizando apenas um *frame* do vídeo, ou uma imagem estática. Este tipo de abordagem mostra-se mais adequado para cenários onde câmeras não estacionárias são utilizadas.

BORGES; IZQUIERDO (2010), trabalharam em um detector para a classificação de vídeos de telejornais e noticiários, que foram previamente manipulados. Embora estabilizados, estes vídeos criam uma série de dificuldades que não existem em vídeos estacionários, como o movimento da câmera, cortes de cena e trepidação. Considerando que o único objetivo é detectar a presença ou não de fogo no vídeo como um todo, os autores afirmam ter atingido uma taxa de 0,68% de falsos positivos e nenhum falso negativo. Para atingir este resultado, utilizam uma solução que extrai dados de cor, posição do fogo na imagem, formato da chama e textura, e um algoritmo de aprendizagem de máquina:

O método proposto analisa as mudanças quadro a quadro utilizando características de baixo nível para descrever as regiões de fogo. Estas características são a cor, a área, a granularidade da superfície, a rugosidade do contorno e a assimetria das regiões de fogo. Devido à oscilação e as características aleatórias do fogo estas características são discriminantes importantes. O comportamento de cada uma destas características é avaliado e os resultados são então combinados de acordo com um classificador Bayesiano de forma a obter uma detecção robusta. Além disso, informações previamente conhecidas do fogo são utilizadas para melhorar os resultados da classificação. (BORGES; IZQUIERDO, 2010)

No sentido de fugir de uma análise temporal utilizando uma avaliação do contorno no domínio da frequência da forma como é feita em YAMAGISHI; YAMAGUCHI (2000), que sabidamente tem um custo computacional alto, BORGES; IZQUIERDO (2010) apresentam uma nova proposta: as características de rugosidade do contorno são computadas como a relação entre o contorno e o fecho convexo (Convex Hull). Assim, a rugosidade da borda é definida por BORGES; IZQUIERDO (2010) através da Eq. 13 onde P_S é o perímetro da borda e P_{CHS} é o perímetro do fecho convexo:

$$BR = P_S/P_{CHS} \quad (13)$$

Para calcular o perímetro usa-se a contagem de píxels conectados horizontal e verticalmente, mais $\sqrt{2}$ vezes o número de píxels na diagonal. Utiliza-se um *threshold* para decidir entre fogo e não fogo. Os experimentos demonstram que os resultados são parecidos com aqueles atingidos utilizando a transformada de Fourier (BORGES; IZQUIERDO, 2010).

Na mesma linha, CHENEBERT; BRECKON; GASZCZAK (2011) buscam utilizar informações de textura para a classificação de regiões de fogo e não fogo. O método apresentado é dividido em duas etapas. A primeira etapa consiste na classificação de cada um dos píxels individualmente, utilizando a Eq. 6 proposta por *chen2004early*, que leva em consideração limiares fixos no *color space* RGB. Na segunda etapa, para a classificação das regiões são extraídos histogramas de 10 *bins* para os canais H e S do *color space* HSV resultando em 20 atributos. Ainda para a classificação das texturas os autores utilizam GLCM, proposto inicialmente por HARALICK; SHANMUGAM; DINSTEN (1973), extraindo energia, entropia, contraste, homogeneidade e correlação dos valores de píxel em H e S.

Cada região é então classificada com base nestes 30 atributos que são testados em um classificador de árvore e em uma rede neural. Os resultados obtidos mostram que o algoritmo CART (Árvores de Classificação e Regressão), proposto inicialmente em BREIMAN et al. (1984) é o que fornece os melhores resultados. Os autores tentaram ainda reduzir o número de atributos utilizando Análise dos Componentes Principais, de forma a encontrar os mais discriminantes, mas os resultados mostraram que não houve uma melhora significativa. Os resultados obtidos na classificação quadro a quadro em um *dataset* de 1194 imagens igualmente distribuídas entre fogo e não-fogo, utilizando validação cruzada com os dados divididos em 8 subconjuntos, mostraram um desempenho de 87,83% com uma taxa de falsos positivos de aproximadamente 7%.

Precisa-se no entanto considerar que todas as técnicas apresentadas até aqui, com exceção de BORGES; IZQUIERDO (2010) e CHENEBERT; BRECKON; GASZCZAK (2011), buscam resolver o problema de detecção utilizando imagens obtidas a partir de câmeras perfeitamente estacionárias tipicamente utilizadas em sistemas de vigilância. O

tipo de câmera utilizada e os objetivos específicos do trabalho afetam diretamente as técnicas propostas. Pode-se ainda afirmar que a evolução dos algoritmos de aprendizagem de máquina permitiu que novas soluções explorassem uma quantidade maior de características do fogo. Por outro lado, fica claro que a parte de geração de descritores foi pouco explorada ou não foi capaz de gerar bons resultados.

4 DATASET E MÉTRICAS DE AVALIAÇÃO

A existência de um *ground truth* e de métricas padronizadas é fundamental para permitir a avaliação dos algoritmos e a evolução das técnicas de visão computacional. Apesar de haver uma quantidade significativa de trabalhos publicados na área, é difícil comparar e reproduzir os resultados de diferentes detectores de fogo. Neste sentido, fez-se necessária a criação de um *dataset* contemplando vídeos e anotações, e também um esquema de avaliação que se adaptassem aos trabalhos que são considerados o estado da arte neste campo de pesquisa. Além disto, acredita-se que a existência de um *dataset* permita uma evolução natural dos algoritmos VFD representando para esta área o mesmo que os *datasets* de reconhecimento de face, como os apresentados por BERG et al. (2004) e JAIN; LEARNED-MILLER (2010), e os *datasets* de segmentação e reconecimento de objetos, como os apresentados em RUSSELL et al. (2008), XIAO et al. (2010), FEI-FEI; FERGUS; PERONA (2007), DENG et al. (2009) e SCHROFF; CRIMINISI; ZISSERMAN (2011), representaram para as suas respectivas áreas.

A utilização de *datasets* proprietários representa uma barreira à avaliação independente dos algoritmos. Em um cenário ideal na ciência, as partes interessadas precisam ser capazes de reproduzir os resultados alegados, validando-os com o *ground truth* e avaliando os resultados. No entanto, durante o decorrer deste trabalho, não foi possível encontrar nenhum *dataset* público voltado para a avaliação de sistemas de detecção de fogo baseados em vídeos de câmeras não-estacionárias.

Os vídeos utilizados em BORGES; IZQUIERDO (2010), pertencentes à rede de emissoras de televisão alemã Deutsche Welle GmbH., não são disponibilizados publicamente. Já CHENEVERT; BRECKON; GASZCZAK (2011) não fazem qualquer referência aos vídeos que foram utilizados para os testes. Quanto aos *datasets* utilizados para avaliação de sistemas estacionários como os apresentados em ÇELIK; OZKARAMANLI; DEMIREL (2007), CHEN; HE; WANG (2010), CELIK (2010), KOLESOV et al. (2010), LABATI et al. (2013) e ZHOU et al. (2010) são incompletos no sentido de não trazerem as informações de *ground truth*, caracterizando-se assim, como uma coleção de vídeos sem qualquer informação adicional. A falta de informações essenciais, como a contagem de *frames* que representam fogo e não fogo, faz com que não seja possível calcular métricas

como a acurácia, especificidade, *fall-out*, taxa de falsos negativos, taxa de falsos positivos e outras estatísticas relevantes.

De fato, a inexistência de dados de *ground truth* e a falta de padronização para avaliar a saída dos algoritmos de detecção de chamas resulta em uma situação onde os resultados são muito dependentes daquilo que os autores consideram como uma detecção correta. Neste sentido, uma das contribuições deste trabalho é a criação de um *dataset* novo, composto por 20 vídeos e suas respectivas anotações. Os vídeos utilizados apresentam uma variedade de dificuldades, incluindo oclusão, variação de escala, vibração de câmera e diferentes características de brilho e contraste, de forma a se aproximar das condições que os sistemas de detecção de fogo encontrarão em situações reais. Uma segunda contribuição neste sentido é a apresentação de um esquema de avaliação completo, composto dos seguintes artefatos:

- Um algoritmo para encontrar a correspondência entre a saída do detector de fogo e as anotações do *ground truth*.
- Dois métodos para a avaliação dos resultados atingidos pelo detector quando utilizado no *dataset*: um para verificar somente a classificação *frame-a-frame* e outro que leva em consideração a posição da chama na imagem;
- O código-fonte e bibliotecas que implementam os métodos propostos.

Além de oferecer uma maneira de comparar diretamente as saídas de diferentes algoritmos, acredita-se que a existência de um *dataset* aberto possa, em um futuro próximo, permitir que outros pesquisadores utilizem abordagens baseadas em aprendizagem de máquina. Uma vez que o processo de avaliação é automatizado e os artefatos de software liberados simplificam a criação de dados de treinamento, os pesquisadores podem direcionar seus esforços no desenvolvimento de técnicas e algoritmos mais avançados para resolver o problema da detecção de fogo baseado em vídeo.

4.1 Dataset

No sentido de desenvolver os objetivos e entender as necessidades que o *dataset* e as métricas propostas deveriam satisfazer, buscou-se inicialmente entender como autores de trabalhos anteriores na área haviam avaliado seus resultados. O problema da detecção de fogo utilizando vídeos não-estacionários foi apresentado pela primeira vez em BORGES; IZQUIERDO (2010) que apresenta, conforme descrito no Cap.3, um método para classificar arquivos gravados de noticiário de televisão. O método proposto apresenta dois estágios, onde o primeiro é a classificação individual dos pixels utilizando a cor e o segundo é extração de texturas e categorização utilizando um classificador Bayesiano. Neste caso, avaliam seus resultados considerando apenas a classificação atribuída

ao vídeo como um todo, desconsiderando qualquer informação referente ao momento da detecção ou a localização do fogo na imagem.

Posteriormente, CHENEVERT; BRECKON; GASZCZAK (2011) propõem também uma abordagem que combina a classificação da cor através da equação proposta anteriormente por CHEN; WU; CHIOU (2004) com descritores de textura GLCM e algoritmos de aprendizagem de máquina como redes neurais e árvores de regressão. Utilizando esta abordagem, os autores afirmam que conseguiram processar 12 *frames* por segundo atingindo uma precisão de 87.83%. Estes resultados são dados considerando apenas se os *frames* foram classificados como contendo ou não regiões de fogo. Cabe reforçar que os autores não detalham os dados de teste utilizados e deixam de apresentar por exemplo o *recall* e acurácia.

Considerando as possíveis aplicações que o *dataset* proposto possa vir a ter, optou-se por disponibilizá-lo como um pacote único, permitindo que os desenvolvedores possam utilizar os dados da forma que lhes for mais conveniente. Enquanto nas abordagens baseadas em aprendizagem de máquina é necessário dividir os dados em conjuntos de treinamento, testes e avaliação, as abordagens tradicionais podem utilizar todos os dados somente na avaliação da solução proposta. Assim, os artefatos de software que acompanham o *dataset* foram desenvolvidos permitindo o acesso a *frames* específicos de forma simplificada, permitindo que desenvolvedores implementem sem muito esforço técnicas de subamostragem como validação cruzada e *bootstrapping*. Outro fator que influenciou na decisão de distribuir os dados um pacote único é a preexistência de técnicas que utilizam uma sequência de *frames* para extrair informações de pulsação do fogo e fluxo óptico.

Os vídeos utilizados na criação do *ground truth* são distribuídos sob a licença *Creative Commons By 3.0*. O CC BY 3.0, permite copiar e redistribuir o material em qualquer formato. Permite também a remixagem e compartilhamento livre com a menção da autoria e a licença. Inicialmente foram gerados 20 vídeos de 30 segundos, mantendo somente as partes que mostram a filmagem da câmera. Os vídeos utilizados como base são listados na Tabela 1.

Ao todo, os 20 vídeos produzidos somam 24.372 *frames*. Destes, 11.779 (48,33%) apresentam fogo e 12.593 (51,66%) não apresentam. Assumindo que alguns *frames* podem conter mais de uma região de fogo, o *dataset* apresenta 12.203 anotações (3,59% maior que o número de *frames* contendo fogo). Para criar um conjunto de vídeos que permita avaliar os sistemas de detecção de fogo em vídeos não estacionário, em condições próximas às que eles precisarão atuar em campo, buscou-se utilizar vídeos que apresentem as seguintes propriedades:

- Variedade de fontes: diferentes combustíveis (líquidos ou sólidos) produzem diferentes chamas. Embora, HAMINS; YANG; KASHIWAGI (1992) afirmem que o

Tabela 1: Vídeos base utilizados para a criação do dataset

Nome do vídeo	URL
Mobile Robot Roams Show Floor	http://youtu.be/v6QVTIAUsf0
Drone-Ground Footage Working Fire - Coventry, RI	http://youtu.be/IQu_C7MRPUU
STS-31 Crew Training - Firefighting, EVA Prep and Post	http://youtu.be/8DZWKaWEB6o
Blowing Up House!!!	http://youtu.be/XnbVBMwP9fM
Burning Down The House	http://youtu.be/aIGTyBCMFb4
Car catches fire after flipping near Christchurch	http://youtu.be/3C9t4J0Bwa4
Dachstuhl im vollbrand, Burning house RC Fire Fighters	http://youtu.be/RhGYj9asHhI
My car burning to the ground after doing donuts	http://youtu.be/WY89eGi834Y
Raw Video - Burning Car Lifted Off Trapped Man	http://youtu.be/aIGTyANMFb4
Winnebago RV explodes into flames and hits Firetruck	http://youtu.be/6uL1Y2W0jcl
Fire Hazard of Lithium-ion Batteries in Warehouse Storage	http://youtu.be/MitnyQ4d_4g
A Ride on CoolerBot	http://youtu.be/G43DTnX8o0U
Robots Invade ICRA Show Floor	http://youtu.be/q6MZV88AvXo
6-Wheel Rover Initial Tests	http://youtu.be/Ij9EYxLW1Vs
Samsung Navibot SR8845 cleaning robot	http://youtu.be/pnOIP3f7Rww
Fire on 49 street	https://youtu.be/UOHFTMLJ0TY
House Fire in Clarksburg, MD on 4/28/2012	https://youtu.be/GVP6EQ4E7a8
IU-MIEN hiking in Laos	https://youtu.be/5MQ_gD4MwEw

combustível e o comburente tenham uma influência insignificante na oscilação e pulsação do fogo, é inegável que afetem a cor e forma do fogo.

- Iluminação não-controlada: os vídeos foram gravados em várias condições de iluminação. Cabe mencionar também que o fogo, sendo uma fonte de luz, acaba afetando a iluminação e coloração na cena observada. O *dataset* apresenta também uma ampla variação nas características de contraste e brilho.
- Movimento de câmera: os vídeos foram gravados utilizando câmeras portáteis ou fixadas em robôs móveis, apresentando movimento em 3 eixos e rotação.
- Precisão de cor: Várias câmeras diferentes foram utilizadas, fazendo com que exista uma variedade de parâmetros intrínsecos afetando as características dos vídeos.
- Ruído: existe ruído na imagem, incluindo as linhas verticais e horizontais causadas pelo uso de sensores e *encoders* de baixa qualidade.
- Oclusão parcial e desordem na cena: a oclusão parcial das regiões de chama foi considerado um dos maiores desafios, ainda por resolver, dos detectores de fogo anteriormente propostos. A oclusão é especialmente difícil de tratar em abordagens que utilizam a forma da chama ou o fluxo-óptico.

- *Motion blur*: o borramento da imagem é um fenômeno que ocorre frequentemente como consequência da trepidação e do movimento da câmera enquanto o obturador está aberto.
- Escala e projeção: o fogo não tem características bem definida de tamanho, escala ou ponto de vista. Contudo, procurou-se utilizar videos onde o fogo está entre 1 e 15 metros da câmera, utilizando uma visão que enquadre a origem das chamas.
- Reflexão e outros fenômenos ópticos: Visto que o fogo é uma fonte de luz, podendo afetar os objetos que o circundam, as cenas gravadas podem conter imagens podem conter reflexão especular, *subsurface scattering*, inter-reflexão ou refração, entre outros fenômenos discutidos no Cap. 2 deste trabalho.

4.1.1 Anotações de *Ground Truth*

O fogo pode, muitas vezes, assumir diferentes características de cor, transparência e forma. Estas propriedades impactam diretamente nas áreas que os detectores são capazes de encontrar. Em alguns casos, quando combinado com imagens de baixa resolução, definir onde uma região de fogo começa ou acaba é uma tarefa ambígua. Pela falta de um critério objetivo para incluir ou excluir uma determinada região nas anotações de fogo, recorreu-se ao julgamento de um anotador humano.

Ocasionalmente a oscilação decorrente do fogo resulta em pequenas regiões de chamas que ficam completamente separadas do foco principal do fogo. Dado que o fogo pulsa entre 2 e 10 Herz é praticamente impossível anotar o contorno exato das chamas para cada *frame*. Assim sendo, optou-se por fazer as anotações na forma de um retângulo que engloba toda a região de fogo na imagem. Em casos onde um foco do fogo está afastado suficientemente de outro para ser anotado sem que os retângulos apresentassem intersecção, as regiões foram anotadas como regiões distintas. Desta forma, manter os dados puros para que possam ser utilizados em técnicas baseadas em aprendizagem de máquina. Para cada um dos *frames* que apresentam fogo, uma ou mais anotações foram feitas. Um exemplo de anotação é apresentado na Fig. 6.

Estas anotações são arquivadas em formato XML. A linguagem de marcação XML foi escolhida por ser um padrão definido pela W3C, endossado pela industria de software e que pode ser facilmente lido e entendido por um humano. Cada vídeo é liberado com um arquivo XML contendo as anotações correspondentes, fazendo com que seja simples estender o *dataset* ou separá-lo em partes menores para treinamento, testes e avaliação.

4.1.2 Artefatos de Software

Durante a criação do *dataset* também foram construídos alguns artefatos de software que se destinam a manipular os vídeos e arquivos XML. Uma biblioteca de manipulação



Figura 6: Anotações de *ground truth* no *dataset*. A ferramenta de anotações desenvolvidas permite delimitar as regiões de fogo utilizando retângulos e salvá-las em um arquivo XML para que possam ser utilizadas posteriormente.

de arquivos, baseada na OpenCV (BRADSKI, 2000), é fornecida para que outros pesquisadores possam ter um acesso simplificado ao conteúdo do *dataset* utilizando métodos já implementados. Da mesma forma, são disponibilizados o software e documentação para permitir que as métricas de avaliação apresentadas em Sec. 4.2 sejam facilmente calculadas, evitando retrabalho com a implementação.

Uma ferramenta de visualização do *ground truth* é fornecida, permitindo aos usuários do *dataset* carregar um arquivo de vídeo e o XML correspondente, mostrando as anotações sobrepostas ao vídeo. A ferramenta permite também que os usuários possam obter dados pertinentes, como estatísticas da imagem e histogramas.

Para as ocasiões em que seja interessante adicionar novos vídeos ao *dataset* ou modificar os dados existentes. Por este motivo, um editor de *ground truth* é fornecido, permitindo que os usuários possam fazê-lo utilizando uma interface gráfica. Dado que as anotações são liberadas no padrão de XML, os arquivos podem ser editados diretamente, utilizando um editor de texto qualquer.

4.2 Metodologia de Avaliação

Um dos principais desafios ao comparar diferentes sistemas de detecção de fogo baseados em vídeo é a falta de um acordo sobre a saída esperada. Os resultados apresentados em diferentes trabalhos são, neste sentido, muito dependentes da definição de “detecção correta” que os autores consideram. Por este motivo, são propostas três diferentes formas

de avaliação:

1. *Frame-a-frame*: onde o problema da detecção torna-se um análogo do problema de classificação binária, e cada *frame* possui uma classe que indica a existência, ou não, de fogo.
2. Baseada em localização: onde são levadas em conta informações de localização da chama na imagem, de forma semelhante ao que ocorre atualmente com os detectores de face e objetos.
3. Baseada em tempo: onde leva-se em consideração o tempo que o detector demora para detectar a existência de fogo.

Na avaliação *frame-a-frame*, cada *frame* é uma instância. Esta é a abordagem utilizada com maior frequência para avaliar os sistemas de detecção de chamadas. Embora os autores não explorem todo o potencial desta abordagem, pode-se afirmar que a mesma já foi utilizada anteriormente em CELIK; DEMIREL (2009), ÇELIK; OZKARAMANLI; DEMIREL (2007) and CHENEBERT; BRECKON; GASZCZAK (2011), onde o *Recall* foi renomeado para taxa de detecção e a precisão foi renomeada para taxa de alarme falso. Ao considerar um *frame* inteiro como fogo ou não-fogo, é possível calcular todas as métricas que seriam calculadas para um classificador binário. Estas métricas incluem a taxa de positivos verdadeiros TPR ¹, taxa de negativos verdadeiros SPC ², valor preditivo positivo PPV ³, valor preditivo negativo NPV ⁴, taxa de falsos positivos FPR ⁵, taxa de falsa descoberta FDR ⁶ e a taxa de falsos negativos FNR ⁷. As equações correspondentes são apresentadas nas Eq. 14, 15, 16, 17, 18, 19 e 20, em que P e N representam respectivamente a contagem de instâncias positivas e negativas presentes no *dataset*, T_P e T_N representam o número de verdadeiros positivos e verdadeiros negativos e F_P e F_N representam os falsos positivos e falsos negativos.

$$TPR = \frac{T_P}{P} \quad (14)$$

$$SPC = \frac{T_N}{N} \quad (15)$$

$$PPV = \frac{T_P}{T_P + F_P} \quad (16)$$

¹TPR do inglês *True Positive Rate*, também conhecida como *Recall*, *Sensibilidade* ou *Hit Rate*.

²TNR do inglês *True Negative Rate*, também conhecida como *Especificidade*.

³PPV do inglês *Positive Predictive Value*, também conhecida como *Precisão*.

⁴NPV do inglês *Negative Prediction Value*.

⁵FPR do inglês *False Positive Rate*, também conhecida como *Fall-Out*.

⁶FDR do inglês *False Discovery Rate*.

⁷FNR do inglês *False Negative Rate*, também conhecida como *Miss Rate*.

$$NPV = \frac{T_N}{T_N + F_N} \quad (17)$$

$$FPR = \frac{F_P}{N} \quad (18)$$

$$FDR = \frac{F_P}{F_P + T_P} \quad (19)$$

$$FNR = \frac{F_N}{F_N + T_P} \quad (20)$$

Ao avaliar o problema da detecção de fogo como uma classificação binária *frame-a-frame*, é possível também calcular a acurácia *ACC* (Eq. 21), *F₁Score* (Eq. 22) e o coeficiente de correlação de Matthews *MCC* (Eq. 23). Embora não tenha sido encontrado um caso anterior onde estas métricas tenham sido utilizadas, acredita-se que possam vir a melhorar a avaliação da qualidade dos detectores, visto que, levam em conta tanto a precisão quanto o *recall*, balanceando as taxas de falsos positivos e negativos.

$$ACC = \frac{T_P + T_N}{P + N} \quad (21)$$

$$F_1 = \frac{2 \times PPV \times TPR}{PPV + TPR} \quad (22)$$

$$MCC = \frac{T_P T_N - F_P F_N}{\sqrt{(T_P + F_P)(T_P + F_N)(T_N + F_P)(T_N + F_N)}} \quad (23)$$

Enquanto a abordagem baseada apenas na classificação individual dos *frames* parece ser adequada para a maioria dos casos, pensando especialmente em sistemas de visão ativa e robôs móveis, é importante também verificar a localização do fogo na imagem. Neste sentido, levou-se em consideração as métricas de avaliação comparativa já utilizadas em sistemas de detecção de face como JAIN; LEARNED-MILLER (2010) e BERG et al. (2004). No entanto, dado que o fogo não tem bordas exatas e as chamas podem se dividir em áreas menores, aceita-se que possa existir mais de uma detecção associada à uma determinada anotação de *ground truth*. Neste caso, utiliza-se uma abordagem 1-N, removendo a etapa de busca por combinações ótimas.

As detecções são consideradas como positivos verdadeiros ou negativos verdadeiros levando em conta sua área de intersecção com os retângulos do *ground truth*. A função de similaridade *S* é apresentada na Eq. 24, em que d_i é uma saída do detector e gt_i é uma anotação. Uma detecção é considerada correta se $S > 0.5$, funcionando com qualquer saída do detector, incluindo formas geométricas ou mesmo o contorno da chama em um vetor de pontos.

$$S_i = \frac{d_i \cap gt_i}{d_i} \quad (24)$$

Uma vez que a abordagem baseada em localização não permite a contagem dos valores negativos no *ground truth*, somente podem ser calculadas as métricas que independem dos dados negativos. Utilizando apenas os dados positivos e considerando que, por definição, a precisão é a razão entre o total de instâncias recuperadas corretamente em relação ao total de instâncias recuperadas, é possível calculá-la através da equação Eq. 25. Já o *recall*, que indica a quantidade de instâncias relevantes que são recuperadas, pode ser calculada, nesta abordagem, através da Eq. 26. D representa o número total de detecções, enquanto G representa o total de anotações presentes na parte do *dataset* que foi utilizada para a avaliação. Por fim, pode-se calcular a similaridade média \bar{S} através da Eq. 27 que fornece um índice não discretizado da exatidão que o detector atinge ao apontar a localização do fogo na imagem.

$$L_{PPV} = \frac{T_P}{D} \quad (25)$$

$$L_{TPR} = \frac{GT - F_N}{GT} \quad (26)$$

$$\bar{S} = \sum \frac{S_i}{n} \quad (27)$$

Outra informação fundamental para avaliar a qualidade de um sistema detector de fogo é o intervalo de tempo entre a detecção do fogo e a primeira vez que este aparece nas anotações do *ground truth*. A latência Lt é dada pela Eq. 28, em que $\min(d_i)$ é o índice do *frame* em que a primeira detecção ocorre e $\min(gt_j)$ é o índice do *frame* em que aparece no *ground truth*. Usualmente, os detectores de fogo apresentam uma latência intrínseca devido ao uso de informações de pulsação do fogo. Esta métrica foi proposta em YIN; MAKRIS; VELASTIN (2007) para avaliação de rastreadores de objetos, onde os autores preferem o termo Critério de Sobreposição Temporal.

$$Lt = \min(d_i) - \min(gt_j) \quad (28)$$

Através do *dataset* e das métricas de avaliação propostas, juntamente com os artefatos de software disponibilizados em <https://github.com/steffensbola/furg-fire-dataset>, permite-se que os desenvolvedores possam avaliar de forma precisa os métodos de detecção de fogo. Espera-se desta forma que os vídeos e anotações de *ground truth* possam representar para a detecção de fogo o que outros *datasets* representaram para áreas como detecção de face, detecção de objetos e segmentação, levando os seus resultados a níveis em que possam ser aplicados na vida real.

5 AVALIAÇÃO ESTATÍSTICA

A avaliação estatística permite criar uma descrição quantitativa dos dados observados. Através da tabulação dos dados, sumarização, observação de medidas estatísticas e gráficos, busca-se entender as principais propriedades que caracterizam o fogo, bem como, encontrar suporte para a proposição de um novo sistema de detecção de chamas. Segundo GIL (2010), embora não se deva considerar os resultados estatísticos como uma verdade absoluta, o método estatístico se caracteriza por ter um razoável grau de precisão e probabilidade de acerto, fazendo-se adequada para os objetivos deste trabalho.

A maioria dos sistemas VFD propostos anteriormente são estritamente voltados para a detecção de chamas utilizando câmeras estacionárias. Muitos deles, como GÓMEZ-RODRÍGUEZ et al. (2002), GOMEZ-RODRIGUEZ; ARRUE; OLLERO (2003), ZHANG et al. (2008), CELIK; DEMIREL (2009), GÜNAY et al. (2009), TOREYIN; CETIN (2009), KRSTINIĆ; STIPANIČEV; JAKOVČEVIĆ (2009), CHUNYU et al. (2010), CELIK (2010), GÜNAY et al. (2010), ZHAO et al. (2011), LABATI et al. (2013) são claramente voltados para a detecção de incêndio em florestas e grandes ambientes abertos. Os trabalhos que mais se aproximam dos objetivos deste são BORGES; IZQUIERDO (2010) e CHENEBERT; BRECKON; GASZCZAK (2011). Por este motivo, a maior parte das técnicas que são apresentadas é voltada à detecção de fumaça, deixando a detecção de chamas como uma tarefa secundária sob a alegação de que a presença de fumaça sempre antecede a presença de fogo, devido a oclusão gerada pela copa das árvores. Também por serem voltadas à ambientes abertos, a presença do sol, a tonalidade do céu e outros elementos naturais afetam como a cor é percebida na câmera, levando possivelmente a desvios da tonalidade real da chama.

5.1 Classificação De Cor

Considerando que grande parte da detecção do fogo passa necessariamente pela classificação do píxel com base na cor, buscou-se encontrar o melhor método ao utilizar-se tanto de referências apresentadas no Cap. 3, quanto ao apresentar-se uma nova proposta utilizando como estatística descritiva e algoritmos de escolha de atributos, aplicada em

dados extraídos do *dataset* apresentado na Sec. 1.2.0.1.

Visando encontrar padrões que possam ser úteis para diferenciar um píxel que represente fogo de píxels que não representam fogo, foram selecionados aleatoriamente vários *frames* do *dataset* para extrair valores de píxel de regiões de fogo e não fogo. Selecionar essas regiões é uma tarefa manual e trabalhosa, visto que o fogo, em alguns casos, apresenta bordas transparentes e uma variação de um quadro para outro. Para cada uma destas anotações os seguintes dados foram extraídos os seguintes dados de cada pixel.

- Valor de cada canal no *color space* RGB.
- Valor de cada canal no *color space* Lab.
- Valor de cada canal no *color space* HSV.
- Valor de cada canal no *color space* YCbCr.

A segmentação das regiões que representam chamas para extração de dados é uma tarefa manual. Pelo fato de muitas vezes o fogo ser translúcido e trazer a coloração do *background* nas regiões de borda, é difícil definir com exatidão onde o fogo inicia ou termina. A Fig. 7 mostra como foram definidas as regiões de chama para extração dos dados que foram utilizados na análise estatística, na seleção de atributos e no treinamento dos classificadores. Todos os píxels contidos nas regiões em azul são considerados como fogo e todos os píxels marcados nas regiões em amarelo são consideradas como não fogo. As regiões que não pertencem a nenhum retângulo não são utilizadas. Cabe salientar que este método foi utilizado apenas para extração dos dados de cor. Já para a extração dos dados contidos nas outras etapas foram utilizadas outras marcações.

Utilizando esta abordagem, foram coletados dados de 1.264.168 pixels, dos quais 279.883 correspondem a regiões consideradas como fogo e 984.285 correspondem a regiões que foram consideradas como não fogo. Como o fogo não apresenta contornos fixos e bem definidos, em alguns casos, a região anotada como fogo inclui também píxels que não são do fogo em si, mas de algum material próximo ou até mesmo do material que está sendo queimado. Acredita-se que, o fato da anotação da região de fogo não ser idêntica ao formato do mesmo, não impacte significativamente nas estatísticas obtidas.

Dado que cada um dos autores citados na Sec. 3 defende um *color space* em específico e não existe um consenso sobre qual deles oferece a melhor separação entre regiões de chama e não chama, a primeira etapa do estudo consistiu na avaliação dos dados obtidos. A tabela 2 apresentam as principais medidas estatísticas com relação às regiões de chama, incluindo também medidas de dispersão. As Figuras 8, 10, 12 e 14 trazem o histograma de cada um dos canais de cor da amostra extraída. No lado esquerdo são apresentados os dados das regiões que representam chamas e no lado direito é apresentado o histograma que corresponde aos dados de regiões que foram consideradas como não fogo.

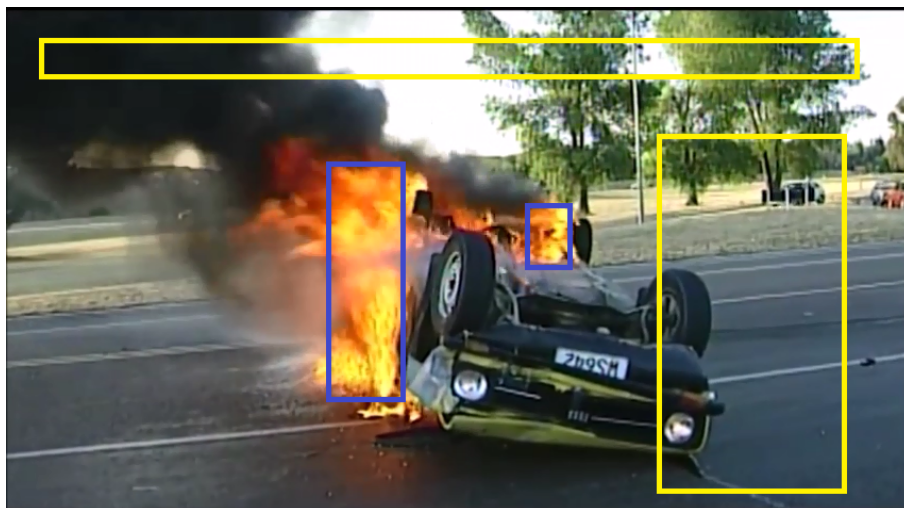


Figura 7: Segmentação manual das regiões de chama para mineração de dados e análise estatística. Todos os pixels contidos nas regiões em azul são considerados como fogo e todos os pixels marcados nas regiões em amarelo são consideradas como não fogo. As regiões que não pertencem a nenhum retângulo não são utilizadas.

Através das estatísticas é possível obter um resumo das características e fazer algumas inferências sobre os dados. Combinando as informações apresentadas na Fig. 8 com as tabelas 2 e 3 pode-se verificar que, em regiões de chama, o canal azul B tende a assumir valores baixos, ao passo que, o canal verde G tende a acompanhar em menor grau o crescimento do canal vermelho R . Pode-se atribuir esta característica ao associar este comportamento diretamente com o espectro de frequência, dado que B representa ondas curtas, G representa ondas médias e R representa ondas longas. A maior parte das regiões de chama apresenta vermelho com intensidade alta, sendo que os valores abaixo de 120 não são representativos. Uma baixa intensidade da cor vermelha está associada a regiões de não fogo.

Assim como ocorre com o canal R , é comum também a saturação do canal G nas regiões que apresentam chama, conforme observado na Fig. 8.c. A saturação do canal B vista na Fig. 8.e, embora menos comum, está relacionada com a cor branca que ocorre no centro da chama, onde o brilho é mais alto. A maior parte dos pixels apresenta valores baixos de azul. Já nas regiões onde não ocorre chama, Fig. 8.e, o azul está mais bem distribuído. A Tab. 2 mostra que as médias para R, G e B são respectivamente 93,15, 165,69 e 237,08 enquanto as medianas são 78, 161 e 252 fazendo com que exista uma assimetria negativa no canal R e uma assimetria próxima a zero no canal G . Observando a Fig. 8.a e 8.c pode-se afirmar que isto se deve, em grande parte, aos valores próximos aos níveis de saturação nestes canais, que elevam a média e fazem com que a cauda esquerda seja mais longa.

Considerando que o *color space* RGB pode ser representado por um cubo, onde cada um dos seus componentes representa um eixo, buscou-se encontrar a distância entre a

cor que representa o fogo e a cor que não representa fogo. No canal B ocorre a menor diferença entre as duas médias, 19,18. Já para G a diferença entre as médias apresentada é 51,76, aumentando para 127,92 no canal R . A distância euclidiana entre regiões de chama e regiões ordinárias é de 139,32. Finalizando a análise estatística do espaço de cor RGB, o desvio-padrão do canal R é um pouco mais da metade do canal G . A medida do desvio padrão neste caso indica a dispersão das intensidades dos pixels amostrados. A partir destes dados, pode-se concluir que o fator discriminante na classificação por cor é o vermelho.

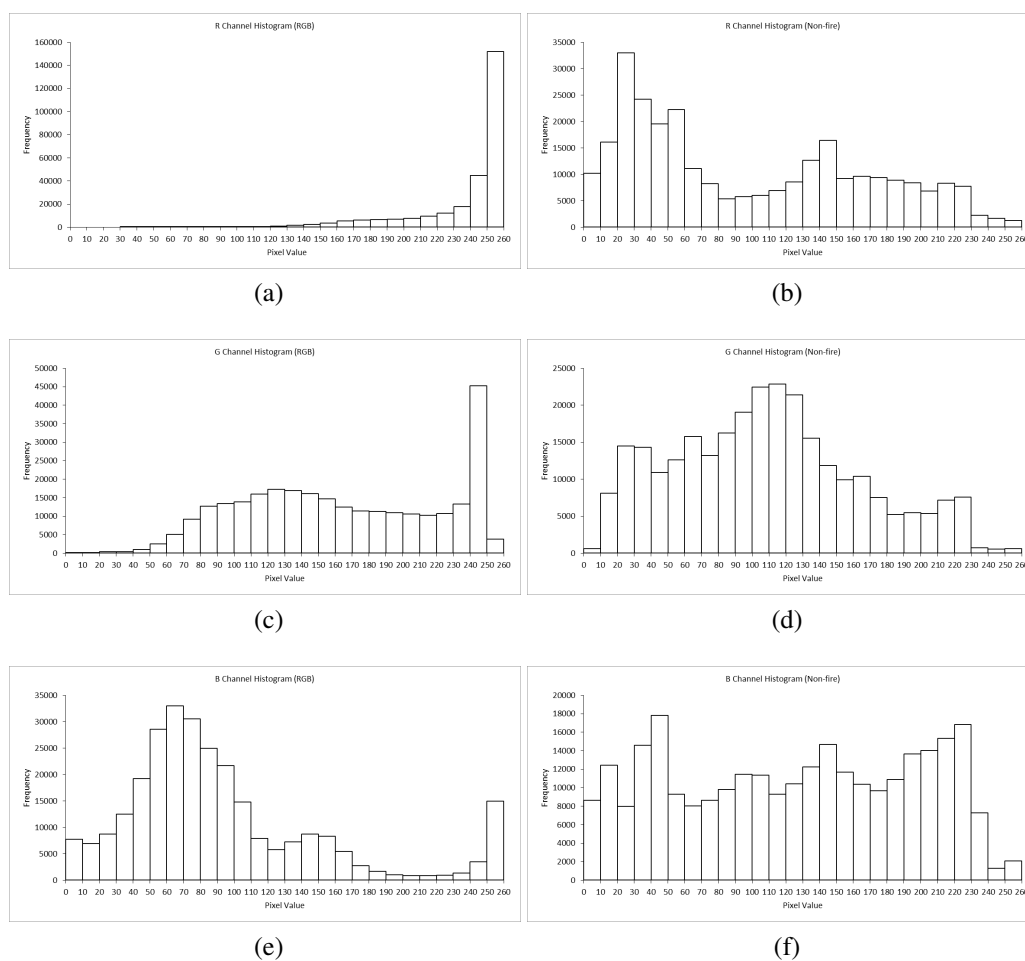


Figura 8: Histogramas de regiões de fogo para o *color space* RGB. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

Analisando as estatísticas obtidas com relação ao *color space* HSV pode-se notar claramente que a média e a mediana no canal H são muito próximas, além de apresentar um desvio padrão pequeno. Observando os histogramas apresentados na Fig. 10a e 10b e os gráficos das Fig. 11a e 11b é possível verificar que grande parte dos pixels podem ser classificados como chama ou não-chama utilizando como limiar superior o valor 50 no canal H . Embora existam instâncias que apresentem valores acima disso, pode-se

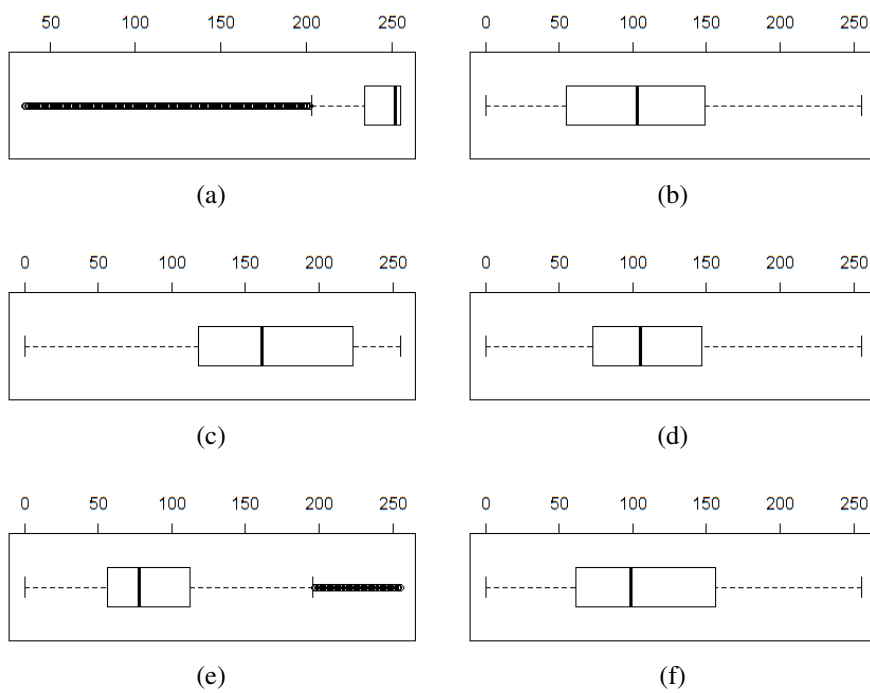


Figura 9: *Boxplot* de regiões de fogo para o *color space* RGB. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

Tabela 2: Estatísticas de cor nas regiões de fogo. Os cabeçalhos correspondem respectivamente aos *color spaces* RGB, HSV, Lab e YCbCr.

	B	G	R	H	S	V	L	a	b	Y	Cb	Cr
Média	93,15	165,69	237,08	32,39	162,44	200,78	156,53	147,38	170,47	146,63	166,02	85,74
Desvio Padrão	59,31	58,78	30,50	24,59	38,86	66,92	60,68	20,25	18,99	59,98	21,25	21,87
Mediana	78	161	252	26	162	233	161	151	174	148	169	85
Distorção ou obliquidade	1,25	-0,03	-2,22	3,07	0,16	-1,22	-0,33	-0,29	-0,11	-0,19	-0,37	-0,21
Assimetria de Pearson	0,77	0,24	-1,47	0,78	0,03	-1,44	-0,22	-0,54	-0,56	-0,07	-0,42	0,10
Mínimo	0	0	35	0	32	15	10	109	114	14	117	30
Máximo	255	255	255	358	255	255	250	188	212	246	213	139
Amplitude	255	255	220	358	223	240	240	79	98	232	96	109
Desvio Médio	44,09	51,22	22,08	18,77	31,14	54,87	51,36	17,81	15,94	51,24	17,91	18,24
Amplitude Inter-quartil	56	105	21	34	54	91	95	39	31	97	33	35

Tabela 3: Estatísticas de cor nas regiões consideradas como não fogo. Os cabeçalhos correspondem respectivamente aos *color spaces* RGB, HSV, Lab e YCbCr.

	B	G	R	H	S	V	L	a	b	Y	Cb	Cr
Média	112,33	113,92	109,16	130,18	76,20	131,05	120,48	127,87	128,73	112,56	125,54	128,18
Desvio Padrão	63,59	57,39	65,11	97,62	59,31	61,98	58,14	9,12	19,91	57,96	14,91	20,20
Mediana	99	105	103	100	59	122	115	128	132	103	127	125
Distorção	0,59	0,65	0,54	0,41	1,13	0,29	0,44	0,49	-1,22	0,67	-1,18	1,50
Assimetria de Pearson	0,63	0,47	0,28	0,93	0,87	0,44	0,28	-0,04	-0,49	0,49	-0,29	0,47
Mínimo	0	0	0	0	0	6	1	98	65	2	72	53
Máximo	255	255	255	358	255	255	255	185	198	255	200	194
Amplitude	255	255	255	358	255	249	254	87	133	253	128	141
Desvio Médio	52,81	45,54	53,57	87,52	46,38	51,79	46,28	6,38	13,37	46,10	10,08	13,15
Amplitude Inter-quartil	95	74	94	172	70	95	76	9	16	75	13	14

ver claramente que é uma quantidade insignificante em comparação com o tamanho da amostra. É importante ressaltar que a escala apresentada vai de 0 até 179 e não de 0 até 360 como na especificação do *color space*, visto que esta é uma simplificação feita na biblioteca OpenCV.

O canal *S*, que indica a pureza, também permite a aplicação direta de um *threshold* para a classificação dos pixels. Isto pode ser observado na Fig. 10.c. Quanto menor o nível de pureza mais cinza o píxel é. Neste caso, avaliando pelas médias apresentadas na tabelas 3 e 2, pode-se afirmar que regiões de fogo são mais puras que a média das regiões que não representam fogo. No que se refere ao brilho (canal *V*) pode-se afirmar que as regiões que apresentam chamas são em média muito mais brilhantes que o restante da imagem. O desvio padrão, no entanto, é alto, fazendo com que abordagens que utilizem valores de *threshold* fixo, como as apresentadas em YAMAGISHI; YAMAGUCHI (2000), CHEN; HE; WANG (2010); CHEN; WU; CHIOU (2004) e GÜNAY et al. (2010) venham, provavelmente, a produzir resultados ruins quando utilizados com este *dataset*.

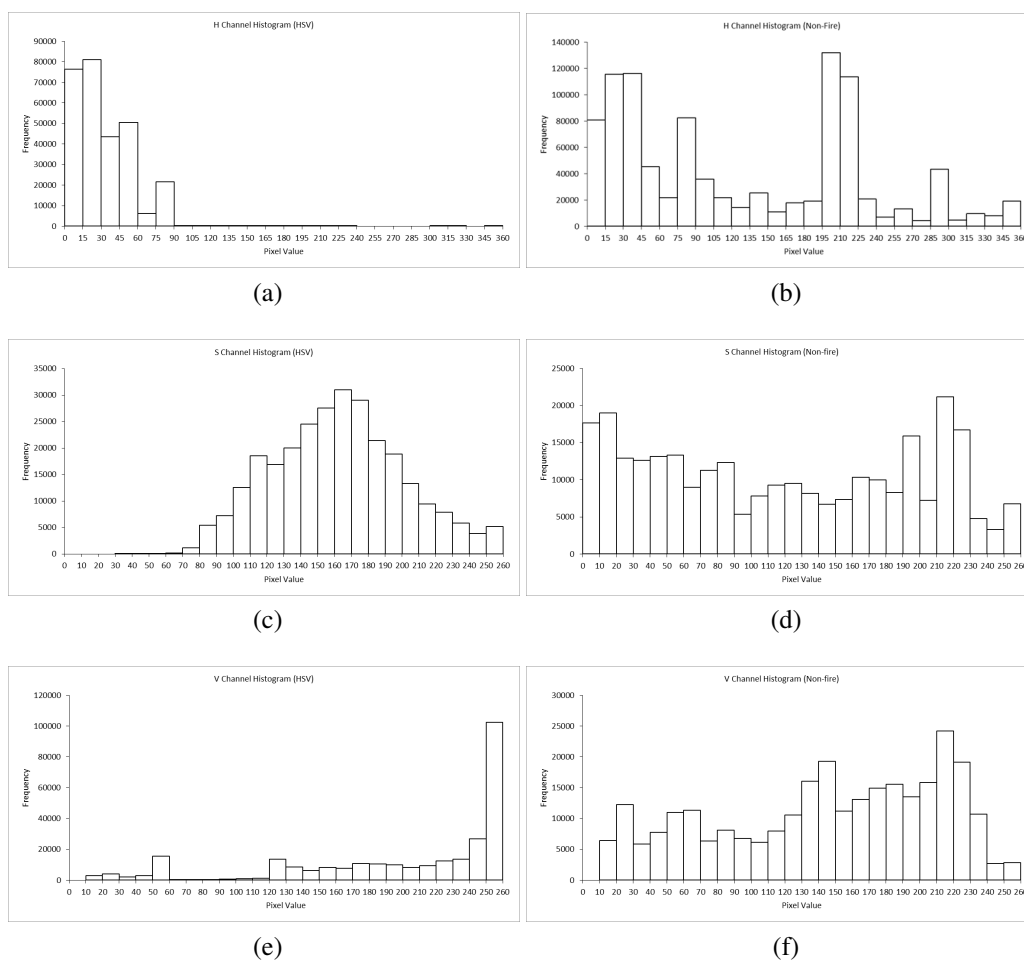


Figura 10: Histogramas de regiões de fogo para o *color space* HSV. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

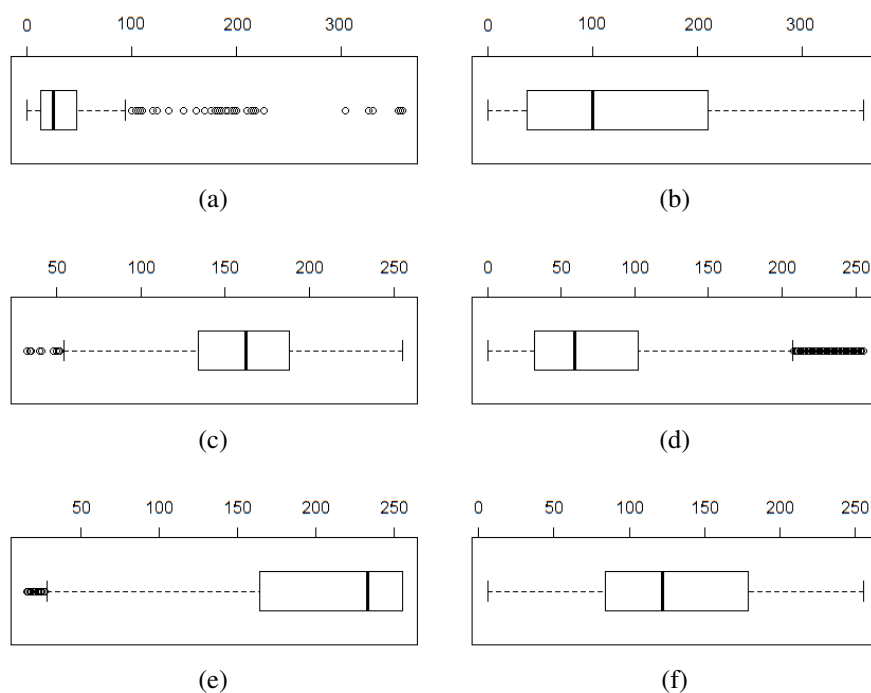


Figura 11: *Boxplot* de regiões de fogo para o *color space* HSV. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

O *color space* Lab tem cores uniformemente distribuídas (cromaticidade uniforme). A sua distribuição é visualmente uniforme, a luminosidade é separada da cromaticidade e é muitas vezes apresentada como a representação dos contrastes cromáticos naturais do ser humano. L corresponde à luminosidade enquanto a e b correspondem aos componentes de cor, baseados no modelo não linear de coordenadas CIE XYZ, sendo que a vai de verde a vermelho e b vai de azul até amarelo, em que as duas dimensões são combinadas de forma a representar as outras cores (CELIK, 2010).

Considerando que as médias entre regiões de fogo e não fogo para o canal L são próximas e que o desvio padrão é alto, pode-se concluir que a luminosidade não é um bom discriminante. Os histogramas apresentados nas figuras 12.a e 12.b corroboram para esta afirmação. Já os contrastes cromáticos representados, neste *color space*, por a e b estabelecem uma distinção melhor, embora a intersecção entre os valores que representam fogo e não fogo ainda seja grande. Pode ser observado ainda que neste *color space* existe um abrandamento da assimetria, fazendo com que os valores de intensidade em cada um dos canais de cor fiquem muito próximos.

Assim como Lab, o *color space* YCbCr também parece não fornecer os melhores discriminantes para regiões de fogo e não fogo. As estatísticas mostram, por exemplo, que o valor médio do canal Y para os pixels que representam regiões de fogo é de 146,63 enquanto regiões que não representam fogo apresentam um valor médio de 112,56, o que

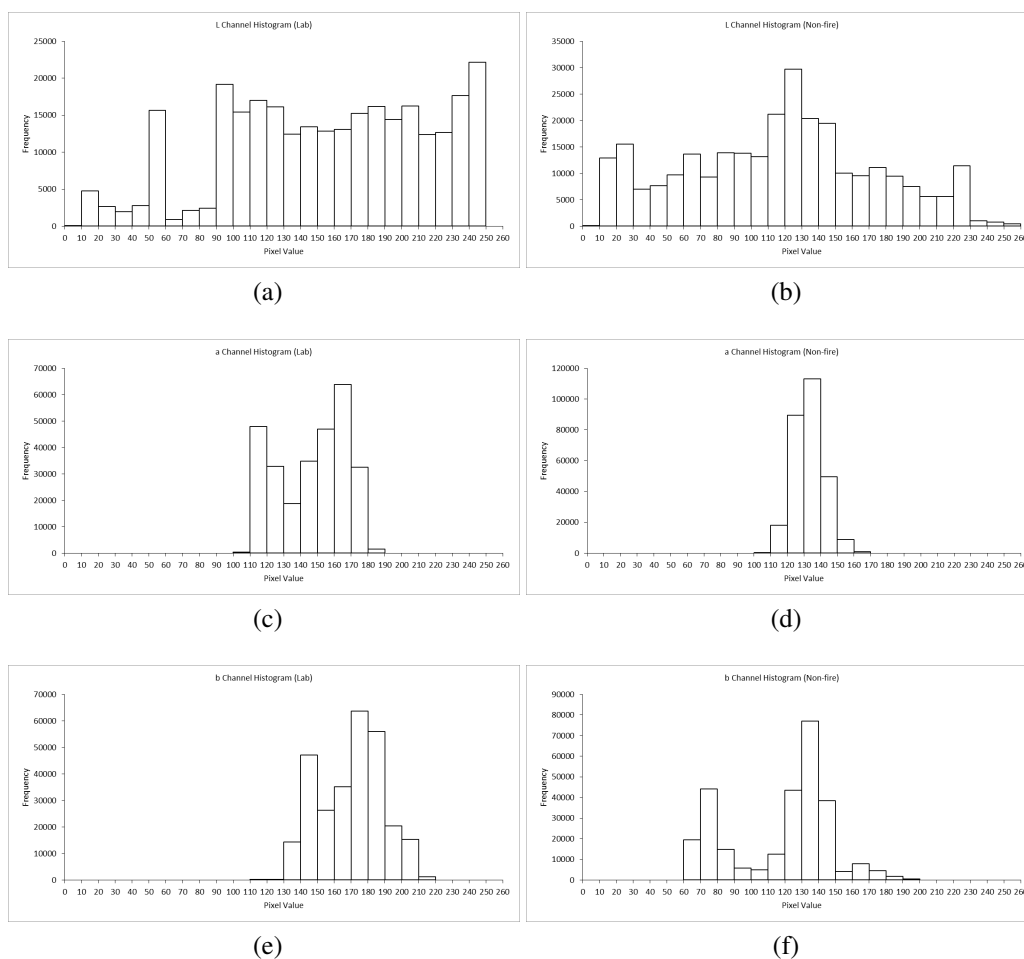


Figura 12: Histogramas de regiões de fogo para o *color space* Lab. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

representa uma diferença de apenas 34,08. O desvio padrão no entanto é de 59,98, fazendo com que dificilmente o valor de H possa ser utilizado na decisão. Nota-se também que existe um abrandamento da assimetria, isso implica dizer que dificilmente poderá ser útil para a extração de texturas. Já os componentes de crominância apresentados aqui são melhores descritores que os apresentados no *color space* Lab.

Existe uma diferença de 42,44 entre as médias para pixels de fogo e não fogo. Conforme pode ser observado na Tabela 2, para Cb o desvio padrão para as regiões de fogo é de 21,87 fazendo deste um bom descritor, pois apresenta uma distância significativa entre as cores. Já para os valores de Cr nota-se que no histograma da Fig. 14 existe uma grande área de intersecção, mas como os dados ficam bastante compactados e apresentam um desvio padrão baixo, existe a possibilidade desses fatores auxiliarem na classificação dos pixels.

Concluída esta análise baseada na estatística descritiva e conhecimento específico sobre o comportamento das regiões de chama, utilizou-se também a ferramenta de mine-

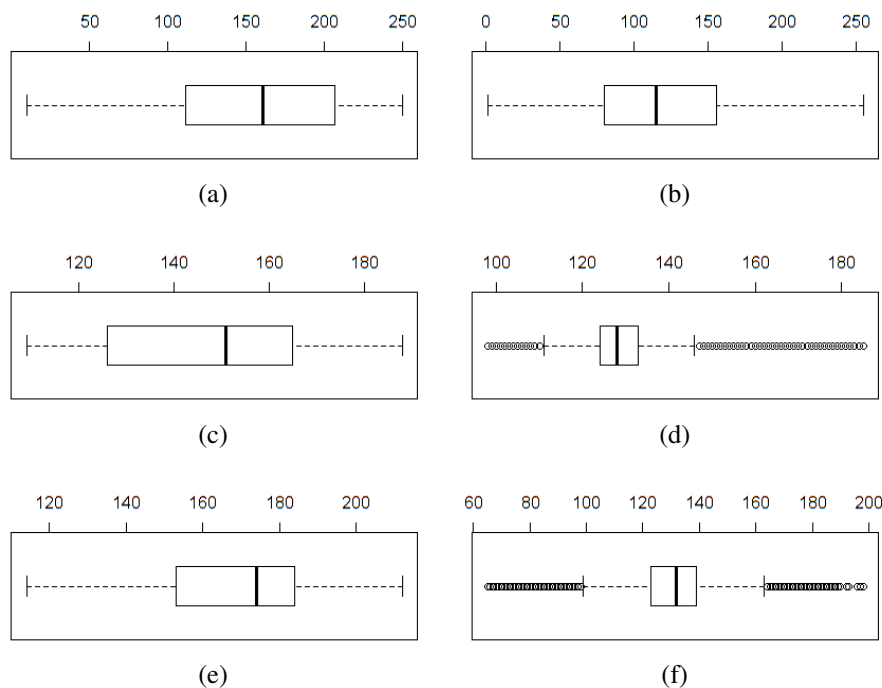


Figura 13: *Boxplot* de regiões de fogo para o *color space* Lab. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

ração de dados WEKA apresentada em (HALL et al., 2009) para fazer a escolha dos melhores atributos dentro dos espaços de cor acima apresentados. A principal vantagem da utilização da ferramenta é que ela permite a aplicação de classificadores, agrupadores e técnicas de aprendizagem de máquina sobre um *dataset* com um *setup time* mínimo.

Para a seleção dos melhores atributos, utilizou-se o algoritmo proposto por HALL (1999), que avalia o valor de um subconjunto de atributos considerando a capacidade que cada atributo tem de individualmente predizer a classificação correta, assim como a redundância entre os atributos. Atributos que apresentam uma alta correlação com a classe e uma pequena intercorrelação são os selecionados. Este algoritmo, combinado com um método guloso, conhecido como *Best First*, mostrou que os atributos que, individualmente, oferecem a melhor representação de chama, e por isso mais discriminantes, são o canal b do *color space* Lab, H do HSV, Cb do YCrCb. Vale notar, no entanto, que existe uma correlação alta entre os canais b e Cb conforme já visto nos histogramas. Os cálculos da correlação com base nos valores de pixel extraídos, mostram que Lab e YCbCr são muito próximos, apontando uma correlação de 0,997 entre L e Y , 0,915 entre a e Cr e -0,988 entre b e Cb . Desta forma, não existe um ganho de informação significativo quando Lab e YCbCr são utilizados em conjunto.

Como os resultados da classificação utilizando um método guloso não permitiram atingir uma conclusão sobre qual *color space* que permite a melhor classificação, utilizou-

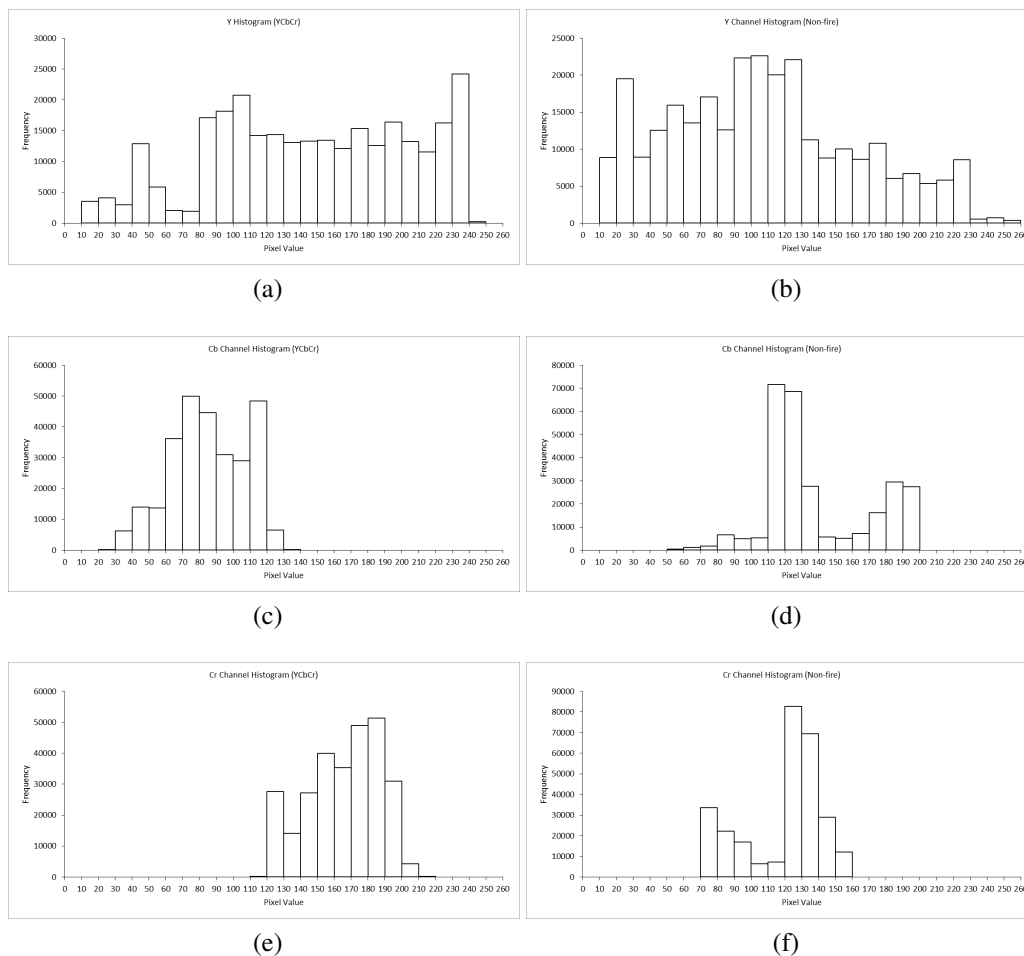


Figura 14: Histogramas de regiões de fogo para o *color space* YCbCr. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

se também a Análise dos Componentes Principais (PCA). Com PCA um número grande de variáveis independentes pode ser reduzido sistematicamente de forma a resultar em um conjunto coerente. PCA é uma técnica estatística que utiliza uma transformação ortogonal para converter um conjunto de observações de variáveis possivelmente correlacionadas a um conjunto de valores de variáveis linearmente decorrelacionadas chamadas componentes principais (DUNTEMAN, 1989).

Ao contrário do que acontece quando considera-se apenas a capacidade individual de classificação para cada variável, os resultados da análise utilizando PCA apontam que as melhores características (canais de cor) estão nos *color spaces* RGB e HSV. Por ordem de classificação, os atributos selecionados foram R , G , B , H e S . Estas informações foram consideradas posteriormente para a criação dos classificadores com base na cor apresentados no Cap. 6.

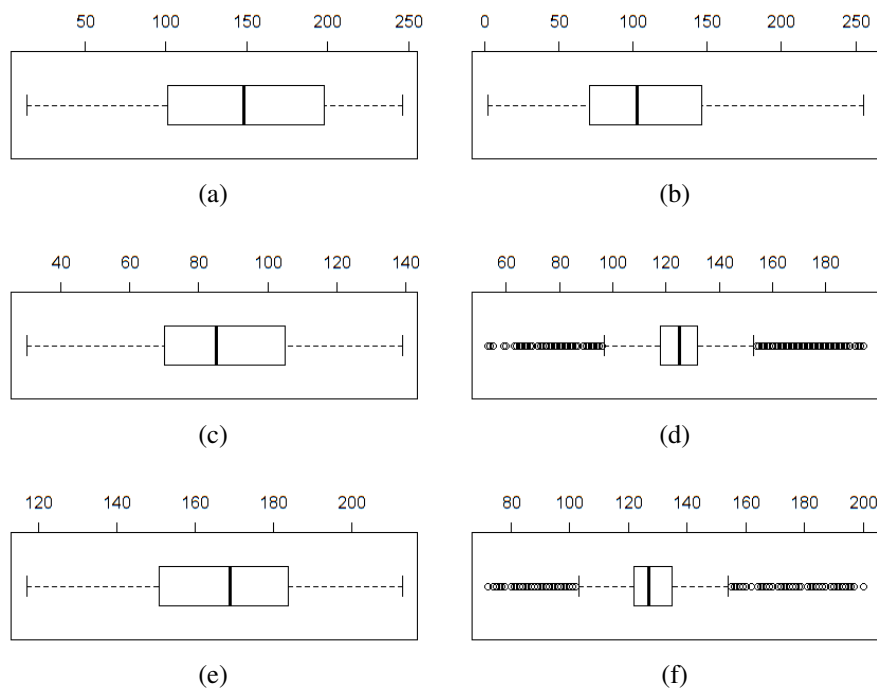


Figura 15: *Boxplot* de regiões de fogo para o *color space* RGB. A coluna da esquerda (a,c,e) representa os valores do canal para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

Tabela 4: Oscilação dos píxels no tempo

Estatística	R		G		B	
	F	NF	F	NF	F	NF
Médias	19,27	3,71	18,98	2,79	19,27	3,71
Mínimo	0,93	0,22	1,12	0,00	0,94	0,00
Máximo	76,63	24,00	45,47	12,57	76,63	24,00
Mediana	19,48	2,60	19,06	2,10	19,48	2,60
Desvio Padrão	11,57	3,36	7,74	2,36	11,57	3,36

5.2 Classificação da Oscilação

Conforme introduzido no Cap. 3, a pulsação e oscilação da chama é outra característica amplamente utilizada para sistemas baseados em câmeras estáticas. No sentido de obter mais informações sobre o comportamento da chama foram realizados experimentos considerando a oscilação do valor no píxel nos últimos 20 *frames*. Para este teste, apenas o espaço de cor RGB foi utilizado.

A Tabela 4 sumariza os dados para oscilação dos píxels em cada um dos canais RGB para regiões de fogo e não fogo. Foram utilizados para esta estatística dados de 295 píxels que representam regiões onde ocorre chama e dados de 806 píxels de regiões sem chamas. Nas regiões sem chamas foram incluídas, além de áreas praticamente estáticas, áreas onde

ocorre movimentação de pessoas, automóveis e árvores balançando, que são usualmente complicadores. O desvio padrão para cada píxel foi computado e utilizado para a geração das estatísticas.

Fica claro que nos três canais existe uma oscilação maior nas regiões de fogo que nas regiões de não fogo. Em uma escala de 0 a 255, o desvio padrão da intensidade do píxel em todos os canais é em média menor que 4 em regiões sem chamas. Já para regiões de fogo essa média se aproxima de 20, apresentando uma diferença pequena entre os canais. Pode-se afirmar que, nas regiões de fogo, a oscilação média em 20 *frames* nos canais *R* e *B* é praticamente idêntica.

Com base nos histogramas apresentados na Fig. 16 pode-se afirmar que o canal verde é o que apresenta o menor desvio padrão entre os canais. Isto se deve em partes à seleção de regiões com oscilação da copa de árvores, onde existe uma tendência de a cor variar entre o verde escuro das folhas e o branco do céu, ambos apresentando a intensidade de *G* próxima de 255. Nota-se também que mesmo nas regiões de chama existe uma grande quantidade de píxels cujo desvio padrão está entre 0 e 10. Esta é uma característica apresentada devido a saturação que ocorre no processo de transformação do sinal analógico para o digital. Por fim, é possível observar que existe uma distorção positiva nos canais *R* e *B*, indicando que existem mais píxels com variação acima da média. Já para o canal *G* a média está muito próxima da mediana, levando a distorção a um valor próximo de zero.

Vale ressaltar no entanto que esta avaliação foi feita extraíndo dados dos vídeos estáticos ou quase estáticos do *dataset*. Para vídeos onde existe um movimento constante de câmera, com muita trepidação e movimento nos 3 eixos como é o caso de câmeras acopladas à um robô móvel este tipo de avaliação com base na variância sobre o valor do píxel ou de regiões pré-definidas pode facilmente induzir a falsos positivos. Abordagens utilizando filtros temporais baseadas na variância, frequência ou mesmo em padrões de oscilação minerados como proposto em LI; WU (2012) e LI; WU (2011) são projetadas apenas para câmeras estáticas.

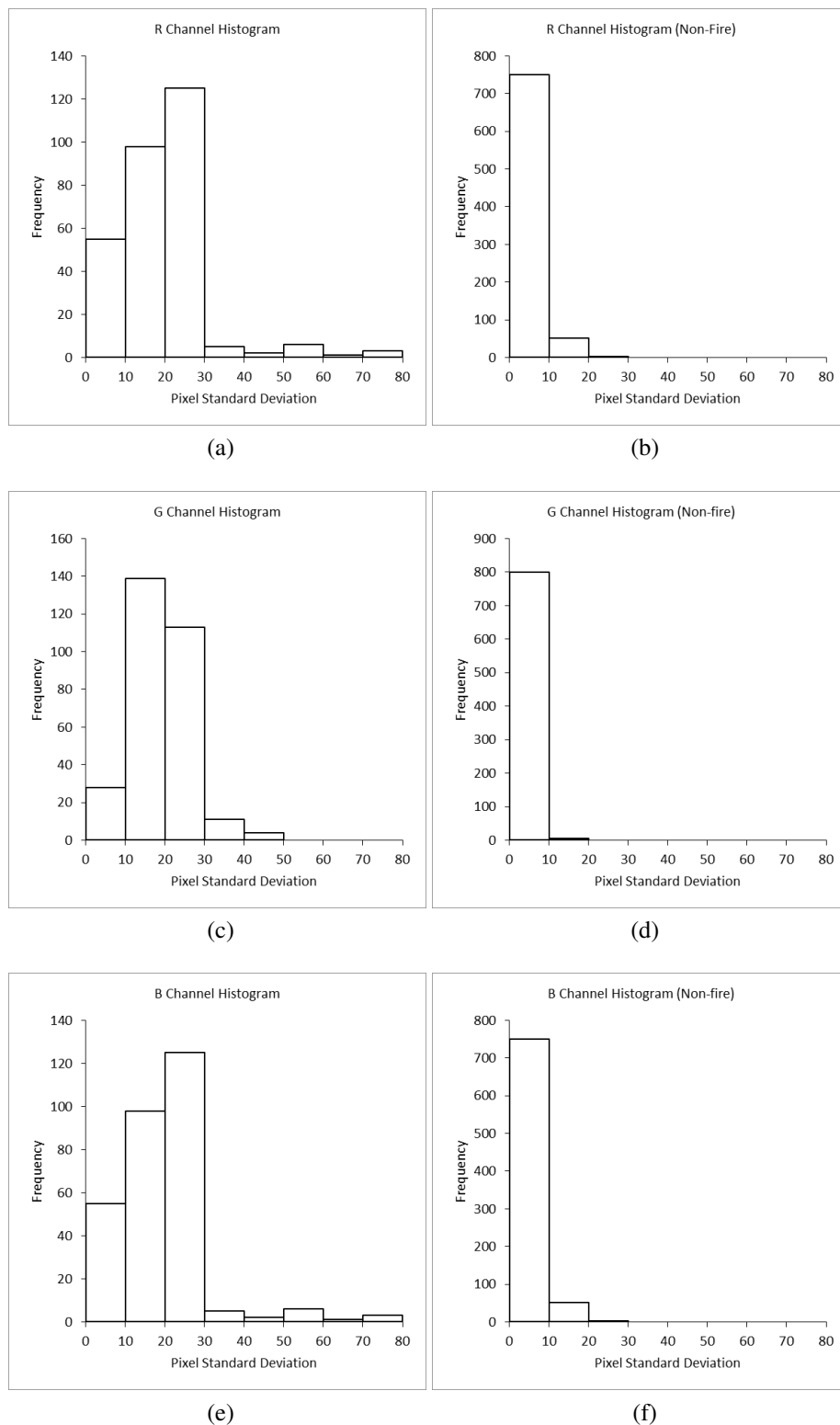


Figura 16: Histogramas de oscilação do valor de píxel no tempo *color space* RGB. A coluna da esquerda (a,c,e) apresenta os valores de desvio padrão do píxel nos canais R, G e B para regiões que apresentam fogo. A coluna da direita (b,d,f) representa os valores do canal para regiões que não apresentam fogo.

6 MÉTODO PROPOSTO

Uma visão geral do funcionamento do detector de fogo proposto é apresentada na Fig. 17. A primeira etapa do processo consiste na classificação do píxel com base na cor utilizando *Random Forest*. A partir das regiões classificadas utilizando informações de matiz, saturação e brilho, são extraídos contornos para serem validados quanto a seu formato, área, média, mediana, variância e assimetria no valor dos píxels da região. Se forem encontradas regiões de fogo esses dados são guardados em um *buffer* e o processo pode assumir dois fluxos distintos dependendo do tipo de imagens com que está trabalhando.

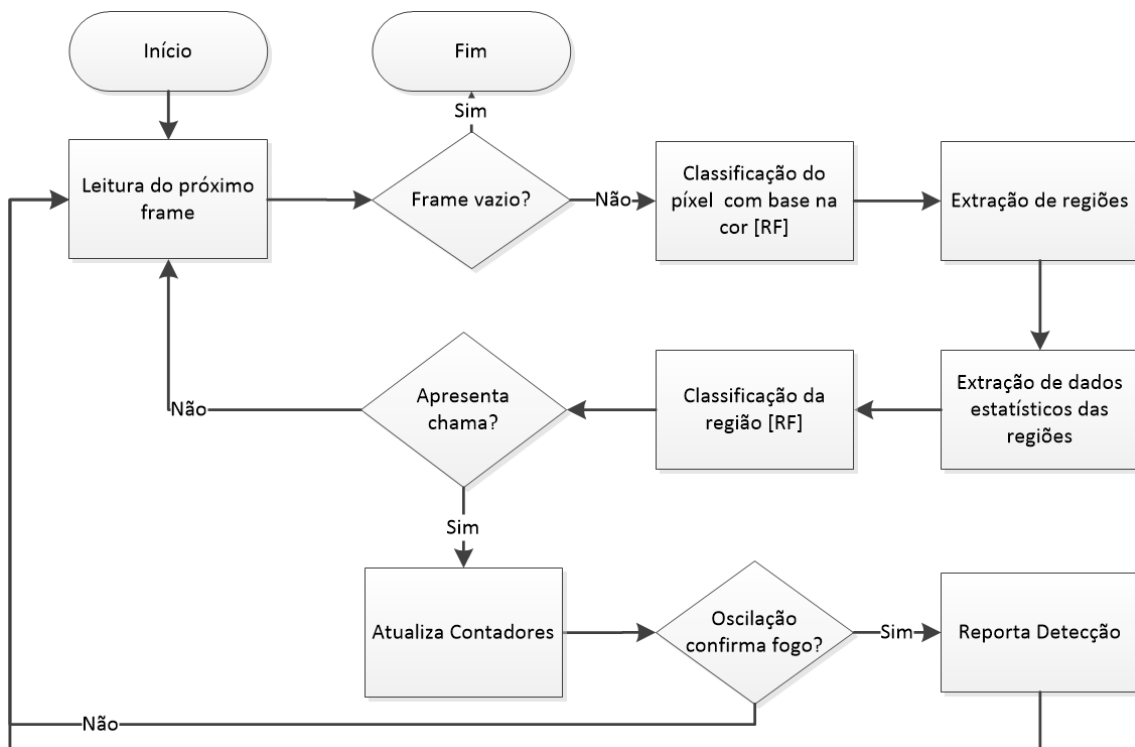


Figura 17: Fluxograma do método proposto: depois de passar por uma classificação dos píxels com base na cor, os contornos são extraídos e as regiões são verificadas quanto à variância e assimetria. Em seguida, uma análise da sequencia de frames é realizada, consistindo essencialmente de filtros temporais no domínio da frequência para vídeos de câmeras estáticas e da computação de detecções subsequentes para vídeos de câmeras móveis.

Com a utilização da Análise dos Componentes Principais foi possível identificar que as variáveis mais discriminantes na classificação são os canais de cor R , G , B , H e S dos *color spaces* RGB e HSV. Os valores de cada um dos canais destes *color spaces* foram utilizadas como entrada para a criação de um modelo de classificação utilizando *Random Forest*. *Random Forests* são uma combinação de classificadores de árvores, tal que cada uma das árvores dependa dos valores de um conjunto de instâncias amostrados de forma independente e com a mesma distribuição para cada uma das árvores treinadas. A taxa de erros converge ao passo em que o número de árvores na floresta aumenta. A acurácia da floresta depende de quão acurada é cada uma das árvores que a compõe e da correlação entre elas. A saída do classificador é obtida através de votação, onde a classe é dada pela moda BREIMAN (2001).

No *Random Forests* cada árvore da floresta é treinada utilizando os mesmos parâmetros, mas com subconjuntos de instâncias diferentes. Estes subconjuntos são criados utilizando a técnica de *bootstrap*, onde, para cada subconjunto são selecionados cerca de $\frac{2}{3}$ das instâncias originais. Estas instâncias são selecionadas com substituição, o que faz com que algumas instâncias sejam selecionadas mais de uma vez e algumas delas não sejam utilizadas. A parte remanescente da amostra original, cerca de $\frac{1}{3}$ dos dados que não são incluídos no subconjunto de treinamento, é utilizada posteriormente para a avaliação do erro. A taxa de erros obtida através desta técnica é conhecida por OOB (out-of-bag).

Utilizando subconjuntos diferentes de instâncias várias árvores são criadas seguindo a mesma abordagem proposta por BREIMAN et al. (1984). Cada nó de decisão pode utilizar um ou mais atributos, que podem ser selecionados seguindo duas abordagens. Na primeira abordagem a escolha dos atributos do nó é completamente aleatória. Na segunda abordagem leva-se em consideração a correlação linear entre os atributos. Cada nó é selecionado utilizando o ganho de informação considerando o índice Gini. Na prática os resultados mostram que a quantidade de atributos selecionados por nó não está diretamente relacionada com os resultados obtidos. Utilizando 1 ou 2 atributos por árvore é possível construir uma floresta com precisão alta. Ao final, a classificação atribuída à instância testada é dada por votação (BREIMAN, 2001).

No modelo de classificação dos pixels pela cor construído para este trabalho, tem-se como atributos os valores de cor em cada pixel. O atributo classe é binário, onde 0 e 1 indicam se a cor medida no pixel corresponde ou não a fogo. Os parâmetros do algoritmo de classificação utilizados são apresentados na Tabela 5. Para a criação do modelo foram utilizados aproximadamente 1.870.000 instâncias de treinamento, onde buscou-se fazer uma separação 50/50 entre as classes.

Para o teste do classificador de pixels com base na cor utilizou-se um segundo conjunto de dados com aproximadamente 534.000 instâncias. Do total de instâncias utilizadas para teste, 86.39% foram classificadas corretamente, sendo que a taxa de falsos positivos ficou em 12.99% e a taxa de falsos negativos ficou em 0.61%. Esses valores são

Tabela 5: Parâmetros da *Random Forest* para classificação de cor

Parâmetro	Valor
Número máximo de níveis da árvore	10
Erro OOB	0.9
Suporte mínimo	500
Número de variáveis utilizadas por nó	3
Número de árvores na floresta	100
Atributos selecionados conforme correlação	Sim

aceitáveis dado que a classificação pela cor é só a primeira etapa de um fluxo em cascata e que a taxa de falsos negativos é insignificante. A classificação com base na cor serve como um pré-processamento para permitir a extração de dados mais precisos, dado que as características do fogo muitas vezes impedem a segmentação em regiões de forma direta.

Após a classificação individual dos pixels, obtém-se uma imagem binária sobre a qual aplica-se o detector de bordas de Canny. Desta forma é possível extrair contornos (Fig. 18b) para as regiões com coloração de fogo encontradas. Para cada contorno são extraídos dados sobre a área, forma e proporção (altura/largura). Cada contorno corresponde a uma região sobre as quais são extraídos os seguintes dados estatísticos:

- Mediana por canal RGB;
- Media por canal RGB;
- Desvio Padrão por canal RGB;
- Assimetria de Pearson por canal RGB;

Aceitando a ideia de que uma região de fogo raramente apresentará um formato de alguma forma geométrica, utilizou-se para a verificação da irregularidade do contorno a mesma abordagem utilizada por BORGES; IZQUIERDO (2010), onde a relação entre o perímetro do contorno e o perímetro do fecho convexo (Convex Hull) são utilizados como medida de rugosidade. A rugosidade da borda é definida através da equação 29 onde $2P_S$ é o perímetro da borda e $2P_{CHS}$ é o perímetro do fecho convexo. Para calcular o perímetro usa-se a contagem de pixels conectados horizontal e verticalmente, mais $\sqrt{2}$ vezes o número de pixels na diagonal. Quanto mais distante de 1 for o resultado desta equação, mais convexidades o contorno apresenta.

$$CR = 2P_S/2P_{CHS} \quad (29)$$

Sabe-se também que a distribuição da intensidade da chama não é uniforme. Foram feitas tentativas de gerar valores representativos desta característica utilizando bancos de

filtros Gabor e bancos de filtros Schmid, que foram inicialmente apresentados em SCHMID (2001). Uma avaliação experimental, no entanto, demonstrou que estes descritores eram incapazes de fornecer informações adequadas, uma vez que uma das propriedades da chama é a falta de escala, podendo apresentar tamanhos muito diferentes na imagem.

Para representar esta característica visualmente perceptível, utilizou-se de uma abordagem que divide a área de chama resultante da classificação dos pixels em 9 subáreas, conforme apresentado na Fig. 18c. Depois de obter um retângulo, que define a região de interesse englobando os pixels conectados de fogo, esta região é dividida em subáreas, numeradas de 1 a 9 a partir do canto superior esquerdo. O valor de cada sub-região é calculado utilizando o valor em escala de cinza conforme a Eq. 30, em que \bar{R} é a média da região, e r_i é a média em cada uma das sub-regiões. Pixels que não apresentam cor de fogo são desconsiderados e não influenciam o cálculo da média.

$$v_i = \bar{R} - \bar{r}_i \quad (30)$$

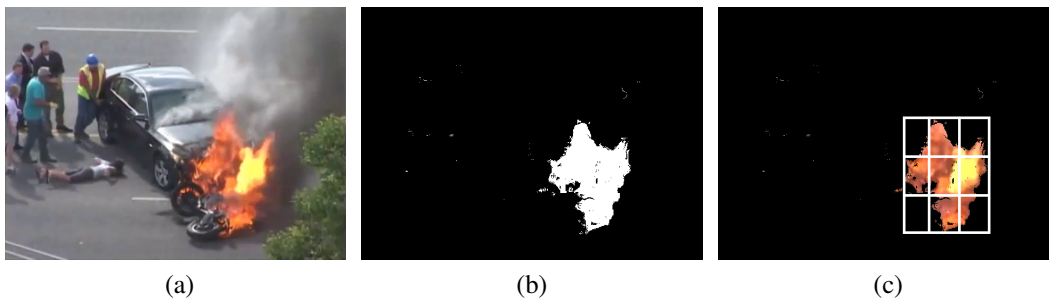


Figura 18: Distribuição das regiões de chama: Imagem original (a), saída da segmentação de cor (b), definição das regiões de interesse (c) e separação das regiões em sub-regiões para extração das médias (d).

Os dados de área, proporção e irregularidade do contorno, combinados com as informações estatísticas da região, são então processadas através de um segundo classificador *Random Forests* completando a avaliação dos dados do *frame* e retornando quais das regiões selecionadas realmente representam fogo. Para o treinamento desta árvore foram extraídas cerca de 22.090 instâncias, das quais 6.709 descrevem regiões de chama e as 13.381 restantes descrevem regiões que apresentam cor de chama mas são falsos positivos resultantes da etapa anterior (classificação pixel-a-pixel com base na cor). Os parâmetros utilizados para a geração do modelo de classificação das regiões são apresentados na Tabela 6. Para evitar sobreajuste (*overfitting*) utilizou-se a precisão e o número de níveis da árvore como critério de poda.

Para imagens obtidas a partir de câmeras móveis, como não é possível fazer uma avaliação da oscilação do pixel ao longo do tempo, utilizando por exemplo frequência e variância, estas informações são suficientes para reportar a detecção. É parametrizável a

Tabela 6: Parâmetros da *Random Forest* para classificação de regiões utilizando dados estatísticos

Parâmetro	Valor
Número máximo de níveis da árvore	10
Erro OOB	0,9
Suporte mínimo	50
Número de variáveis utilizadas por nó	2
Número de árvores na floresta	100
Atributos selecionados conforme correlação	Sim

contagem mínima de *frames* sequenciais para que a detecção seja dada como definitiva. Avalia-se também se a região apresentou alguma alteração no valor do píxel, de forma a suprimir falsos positivos causados por lâmpadas, névoa ou fumaça presentes na imagem. Já para câmeras estáticas, o processo pode ser levado adiante, melhorando a precisão através de um filtro temporal.

Na avaliação temporal é utilizada a variância do píxel nos últimos 3 *frames*. O objetivo aqui é identificar as regiões de borda do fogo dado que é comum os valores de píxel no centro da chama serem saturados não apresentando assim variância alguma. Utiliza-se para isto o canal *R* do *color space* RGB, medindo a distância em relação às médias apresentadas na Tabela 4, 19.27 para regiões de chama e 3.71 para regiões de não chama. A média entre estes valores é 11.49. Uma imagem binária é gerada através da Eq. 31, onde σ_P é o desvio padrão no píxel.

$$P(x, y) = \begin{cases} 1, & \text{se } \sigma_P \geq 11.49 \\ 0, & \text{caso contrário} \end{cases} \quad (31)$$

A saída produzida pode ser utilizada dentro de uma arquitetura maior, para construir sistemas de detecção e extinção automática de fogo. Da mesma forma, os resultados gerados podem ser armazenados para revisão futura e inspeção humana sem exigir deslocamento até o local. O método proposto foi implementado no *framework CvWork* utilizando a linguagem de programação C++ com as bibliotecas OpenCv e Boost. A implementação de protocolos de comunicação industrial, leitura de sensores analógicos, integração com sistemas de supervisão e controle, reconhecimento de faces e outros objetos foge ao escopo deste trabalho, mas é, conforme introduzido anteriormente parte do *framework*.

7 RESULTADOS

A primeira etapa do trabalho realizado consistiu em revisar a bibliografia e implementar parcial ou completamente os algoritmos propostos de forma a verificar os resultados gerados e como os mesmos são impactados por uma situação de câmera não estática. É previsível que alguns métodos previamente propostos falhem devido às condições para os quais foram desenvolvidos. A validação dos classificadores feita para este trabalho não pode, no entanto, ser generalizada de forma a diminuir os trabalhos citados, servindo apenas como referência na busca da solução para um problema novo. Da mesma forma, alguns dos trabalhos citados anteriormente no Cap. 3, pecam ao omitir detalhes, impedindo assim a implementação completa do detector para fins de comparação.

A Fig. 19 apresenta um resultado comparativo da primeira parte do processo de classificação. Ao separar os píxeis que apresentam cor de fogo dos que não apresentam cor de fogo, nota-se que o método proposto (f) é visivelmente melhor que as técnicas propostas por CHEN; WU; CHIOU (2004) (b), CELIK et al. (2007) (c), CELIK; DEMIREL (2009) (d), CELIK (2010) (e), suprimindo a maior parte dos falsos positivos. Já a Fig. 20 apresenta a utilização dos mesmos classificadores em uma imagem que apresenta fogo. Novamente, nota-se que a segmentação realizada apresenta resultados superiores, removendo, por exemplo, as regiões com laranja constante, como a que ocorre nos coletes utilizados pela equipe de resgate. Da mesma forma, as regiões de fumaça também foram suprimidas.

Os resultados da classificação pela cor são de extrema importância dentro da proposta apresentada, pois é a partir destes que é feita a extração e processamento das outras características avaliadas. Uma segmentação ruim acaba prejudicando a extração das estatísticas por região e conseqüentemente todo o restante do processo. O método proposto como um todo, foi testado utilizando o *dataset* introduzido no Cap. 4. Este *dataset* é composto por vídeos de câmeras não estáticas, sendo que a maioria deles apresenta deslocamento nos três eixos. As anotações do vídeo foram feitas quadro a quadro, na forma de retângulos que contornam toda a área que apresenta chama.

Para validar o método proposto os resultados foram comparados com o método de detecção de chamadas em vídeo não estacionário, apresentado em CHENEVERT; BRECKON;

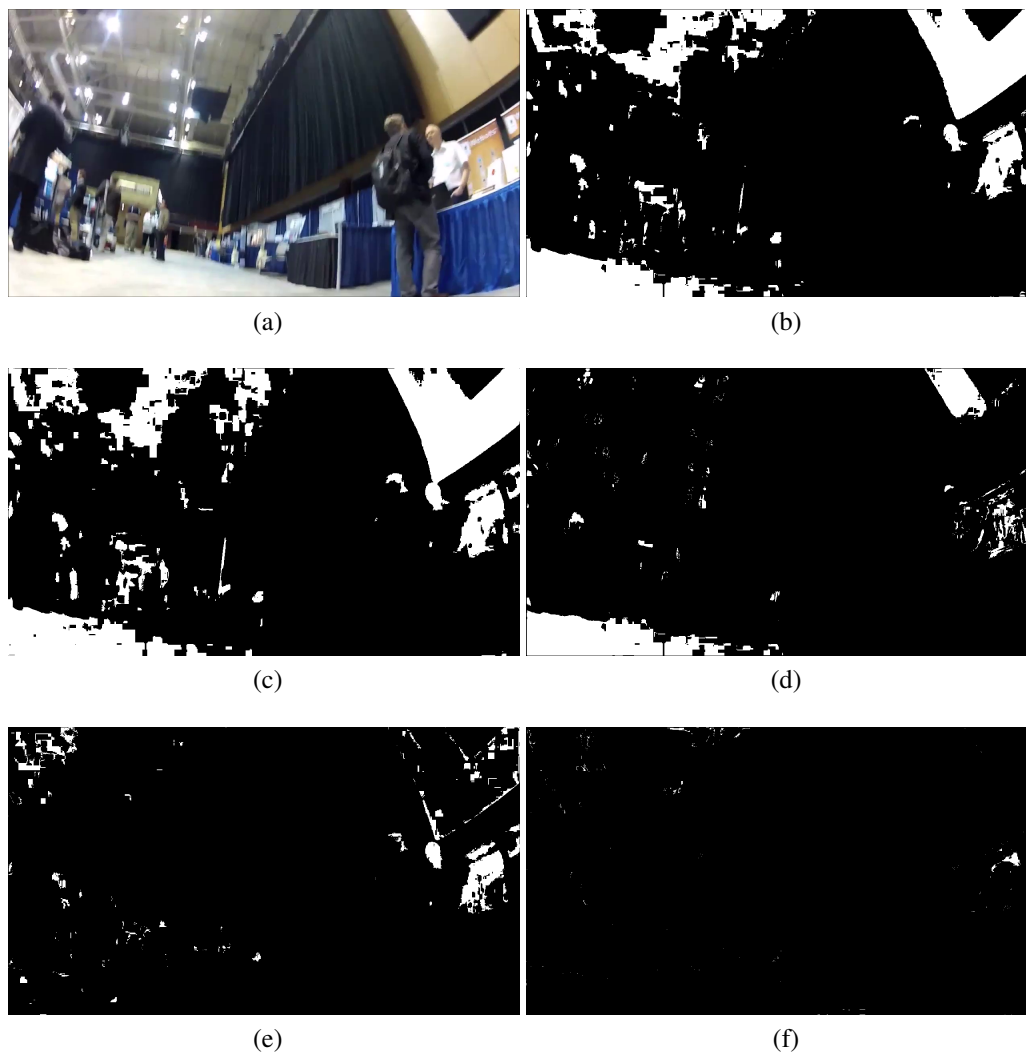


Figura 19: Comparativo de segmentação de cor em imagens sem fogo obtida por CHEN; WU; CHIOU (2004) (b), CELIK et al. (2007) (c), CELIK; DEMIREL (2009) (d), CELIK (2010) (e) e pelo método proposto (f).

GASZCZAK (2011). Este, conforme apresentado na Sec. 3.3 é, também, um método que divide a detecção em duas etapas. Na primeira delas, utiliza uma equação com limiares fixos nos espaços de cor RGB e HSV. Na segunda etapa, utilizando descritores de textura como histogramas e dados de energia, entropia, contraste, homogeneidade e correlação GLMC em 4 direções, um classificador de árvore é treinado e utilizado. Desta forma, foram utilizados os mesmos dados para o treinamento e avaliação de ambos os classificadores.

7.1 Casos de teste

Dois casos de teste foram criados para avaliar a capacidade de classificação e generalização do sistema proposto. No primeiro caso, para cada um dos vídeos do *dataset*,

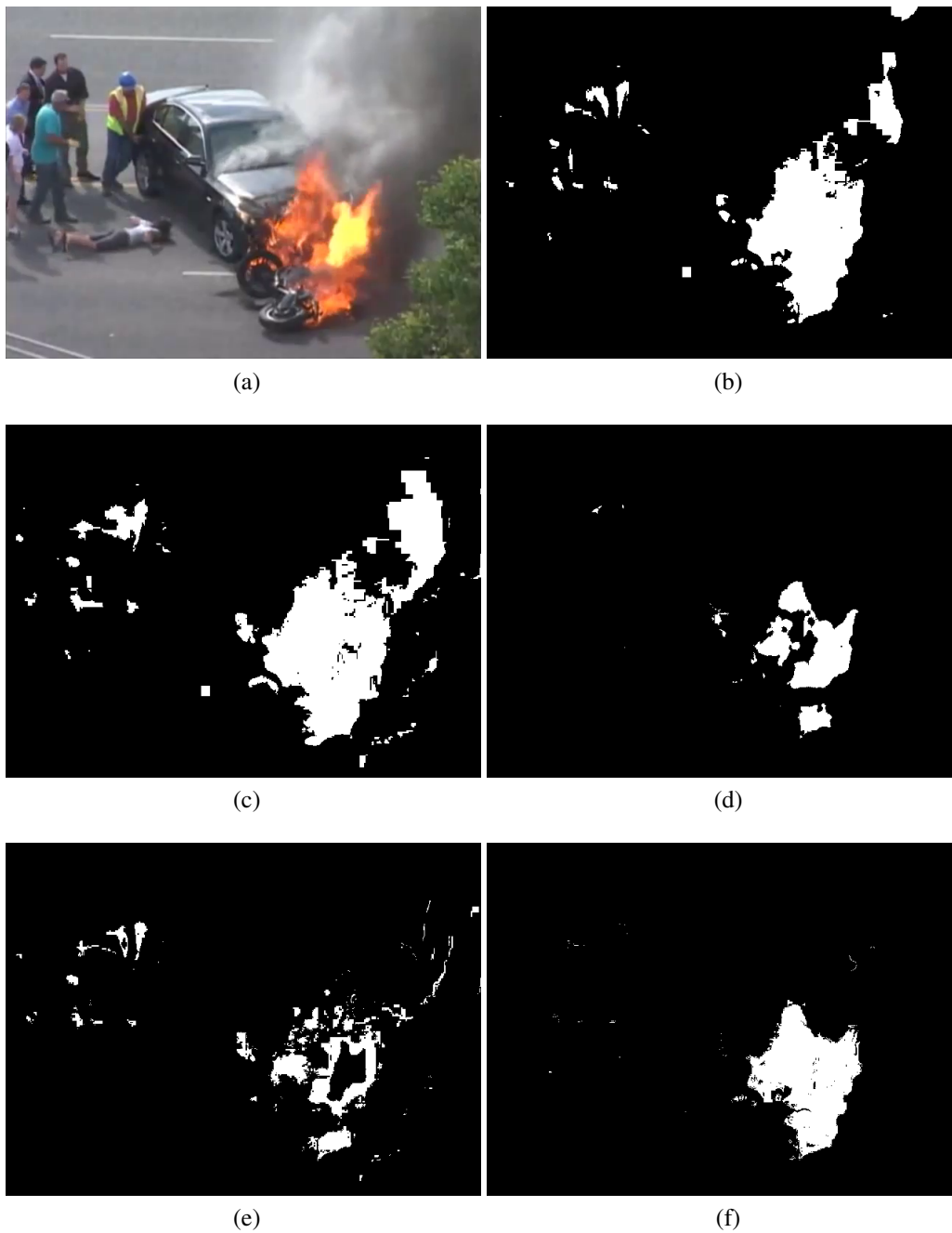


Figura 20: Comparativo de segmentação de cor em imagens com fogo obtida por CHEN; WU; CHIOU (2004) (b), CELIK et al. (2007) (c), CELIK; DEMIREL (2009) (d), CELIK (2010) (e) e pelo método proposto (f).

os primeiros 300 *frames* (aproximadamente $\frac{1}{3}$ do total) foram utilizados para gerar os dados de treinamento. A parte remanescente de cada um destes vídeos foi utilizada para os testes. Já para a criação do segundo caso de teste, 2 vídeos, foram completamente separados e não utilizados no processo de treinamento. Os mesmos modelos de classificação gerados para o caso de teste I foram utilizados para classificar os vídeos do caso de teste II.

Os dados de treinamento para a classificação das regiões utilizando textura são gerados a partir da saída do classificador de píxeis utilizando a cor. Para o treinamento, os dados foram gerados automaticamente levando em conta a intersecção das regiões detectadas com os retângulos de *ground-truth*. Como instâncias positivas considerou-se que as regiões que representam fogo, utilizadas para o treinamento deveriam ter uma intersecção mínima de 0.8. Já as regiões que apresentam uma intersecção menor ou igual a 0.1 foram consideradas como instâncias negativas para o treinamento. Regiões que não satisfazem nenhuma dessas condições não foram utilizadas. O desbalanceamento entre as instâncias de fogo e não-fogo, foi tratado utilizando a técnica de *bootstrap*, onde as instâncias são selecionadas aleatoriamente com substituição.

O comparativo entre os resultados atingidos utilizando os métodos de ZHOU et al. (2010), CELIK (2010), CHENEVERT; BRECKON; GASZCZAK (2011) e o método proposto, para o caso de teste I, pode ser visto na Tab. 7, que apresenta a avaliação quadro a quadro, e na Tab. 8, que apresenta os resultados considerando a localização das detecções. Estes dados levam em consideração que pode existir uma relação de 1-N entre as detecções e as anotações do *ground truth* (ver Fig. 21). No total, 18691 *frames*, originados de 13 vídeos distintos, foram testados. Destes, 8095 *frames* contém anotações sobre regiões de fogo no *ground truth* e 10596 não contém anotações. Entende-se que, conforme as métricas de avaliação propostas, o resultado é um falso positivo quando existe uma detecção, mas não existe nenhuma anotação de *ground truth* para o *frame* correspondente. Entende-se por falso negativo a situação em que existe uma anotação feita por um humano que aponta a região de fogo, mas o detector não aponta tal região.

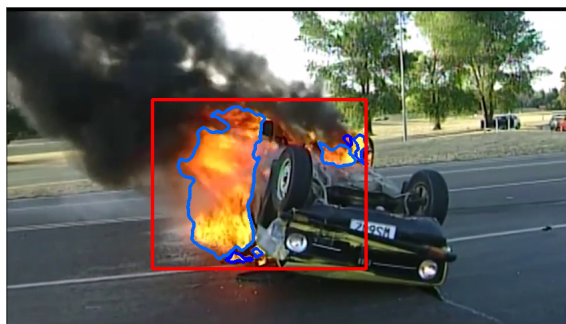


Figura 21: Método de avaliação do classificador proposto. O retângulo em vermelho mostra o gabarito para este quadro. As regiões em azul mostram as regiões retornadas pelo detector proposto. Nota-se que existem várias detecções para uma única anotação de *ground-truth*.

Dentre os métodos testados, ZHOU et al. (2010) e CELIK (2010) são voltados para sistemas detecção de chamas utilizando câmeras estacionárias, demonstrando-se inadequadas para o uso em câmeras móveis. Com relação aos sistemas VFD voltados para câmeras móveis citados na Sec. 3.3, o método apresentado em BORGES; IZQUIERDO

(2010), não pôde ser implementado para fins de comparação pois existem parâmetros do classificador Bayesiano que não foram explicitados no artigo e os dados de treinamento não são publicamente disponíveis. O método de CHENEVERT; BRECKON; GASZCZAK (2011) foi utilizado na avaliação, sendo superado pelo método proposto em todas as métricas avaliadas.

Na avaliação quadro a quadro, utilizando exatamente as mesmas amostras para ambos os métodos, o detector proposto resultou em 122 falsos positivos e 1.365 falsos negativos. Já o detector proposto por Chenevert resultou em 1.685 falsos positivos e 2.243 falsos negativos. Assim, o *Recall*, ou sensibilidade, que é a métrica que indica a capacidade do modelo para classificar instâncias de uma determinada classe mostra que o percentual de *frames* corretamente classificados como fogo é 14,93 maior. Já a especificidade, que é dada pela taxa de negativos reais, é 17,47% maior. Da mesma forma, a precisão indica que a probabilidade de as detecções reportadas pelo método proposto serem corretas é 26,54% maior que aquela atingida por CHENEVERT; BRECKON; GASZCZAK (2011), passando de 0,776 para 0,982. Com relação ao valor preditivo negativo, que é, neste caso, o equivalente da precisão aplicado somente aos *frames* que não contém fogo, pode-se afirmar que a melhora é de 10,77% aumentando de 0,798 para 0,884.

Em relação ao *fall-out*, que é calculado dividindo-se o número de falsos positivos pelo total de instâncias negativas no *dataset*, nota-se que existe uma redução de 15,90% para 1,20%. A taxa de falsa descoberta exibe valores parecidos em decorrência da distribuição dos dados de teste (43% fogo/56% não fogo). Já a taxa de falsos negativos apresentou uma queda de 0,277 para 0,168, o que representa uma melhora de 21,29%.

A acurácia, que é obtida ao dividir a soma dos positivos verdadeiros e negativos verdadeiros pelo total de *frames* testados, mostrou que o método proposto classificou corretamente 92% das instâncias testadas enquanto no método proposto em CHENEVERT; BRECKON; GASZCZAK (2011) 79% das instâncias foram corretamente classificadas. Isso representa uma redução de 13% no número de *frames* classificados erroneamente. A acurácia, embora seja amplamente utilizada na avaliação de modelos de classificação, não pode não fornecer uma representação precisa do classificador e não leva em consideração a matriz de confusão. Em outras palavras, isto significa que é impossível, utilizando apenas a acurácia, avaliar a qualidade de um detector de fogo. Por este motivo, foram calculadas também as métricas F_1 Score e coeficiente de correlação de Matthews.

A *F-measure* foi proposta em CHINCHOR (1992), como uma métrica para combinar precisão e *recall* em uma única medida através da média harmônica. Quando utiliza-se a precisão e o *recall* com o mesmo peso, esta medida recebe o nome de F_1 Score. Considerando esta métrica, o método proposto melhora os resultados em 20,16% passando de 0,749 para 0,900. Já o coeficiente de correlação de Matthews, apresentado inicialmente em MATTHEWS (1975), é também uma métrica balanceada de avaliação das predições dadas por um classificador binário, preparado para lidar com o desbalanceamento das ins-

Tabela 7: Resultados do método proposto na avaliação *frame a frame*. A coluna A apresenta os resultados de CELIK (2010), a coluna B apresenta os resultados de ZHOU et al. (2010) e a coluna C apresenta os resultados de CHENEVERT; BRECKON; GASZCZAK (2011). Todos os métodos foram avaliados utilizando os mesmos dados de teste.

Métrica	A	B	C	Proposto
Recall (TPR)	0,434	0,843	0,723	0,831
Especificidade (SPC)	0,241	0,187	0,841	0,988
Precisão (PPV)	0,434	0,442	0,776	0,982
Valor preditivo negativo (NPV)	0,571	0,610	0,798	0,884
Fall-out (FPR)	0,758	0,812	0,159	0,012
Taxa de falsa descoberta (FDR)	0,565	0,557	0,224	0,018
Miss Rate (FNR)	0,236	0,156	0,277	0,168
Acurácia (ACC)	0,467	0,471	0,790	0,920
F_1 Score	0,553	0,580	0,749	0,900
Coefficiente de correlação de Matthews (MCC)	0,005	0,041	0,569	0,843

tâncias. A escala do MCC varia de -1 a 1, onde -1 indica que a saída do classificador binário tem correlação inversa com o resultado esperado, 0 indica que o classificador é aleatório e 1 indica que as classes retornadas pelo classificador são exatamente iguais ao *ground truth*. Utilizando esta métrica, o classificador proposto também mostra uma melhora significativa em relação ao estado da arte. Os classificadores de ZHOU et al. (2010) e CELIK (2010) mostraram resultados apenas um pouco melhores que uma predição aleatória.

A relação entre precisão e *recall* é um dilema recorrente. Enquanto sistemas com *recall* alto e precisão baixa são aceitáveis para aplicações que dependem da intervenção humana para acionamento, quando se trata de sistemas totalmente automatizados necessitam de uma precisão alta para evitar o acionamento indevido do sistema. Por este motivo, dado que o sistema proposto não foi projetado para uma aplicação específica, o uso de métricas balanceadas como F_1 Score e o coeficiente de correlação de Matthews mostra-se como mais adequado para mensurar a qualidade dos sistemas de detecção de fogo avaliados.

Complementando a avaliação da qualidade do classificador, propõe-se também a comparação de métricas que avaliam a localização da chama no *frame*. Conforme detalhado na Sec. 4.2 a avaliação da localização considera a intersecção da região detectada com as anotações de *ground truth*. Aceita-se a ideia de que uma anotação de *ground truth* possa estar atrelada à várias detecções. As métricas de localização são especialmente interessantes quando se considera as possíveis aplicações do detector em um sistema de visão ativa.

Entende-se por erro de localização os casos em que a área detectada está localizada em sua maior parte fora da região anotada como fogo no *ground truth*. Esse tipo de erro é geralmente causado por regiões de baixo contraste entre a chama e o *background*, fa-

Tabela 8: Resultados do método proposto considerando localização. A coluna A apresenta os resultados de CELIK (2010), a coluna B apresenta os resultados de ZHOU et al. (2010) e a coluna C apresenta os resultados de CHENEBERT; BRECKON; GASZCZAK (2011). Todos os métodos foram avaliados utilizando os mesmos dados de teste.

Métrica	A	B	C	Proposto
Total de regiões detectadas	92356	102862	16397	12864
Total de regiões no <i>ground truth</i>	8425	8425	8425	8425
Precisão (L_{PPV})	0,105	0,124	0,648	0,971
Recall (L_{TPR})	0,523	0,334	0,650	0,805
Taxa de falsa descoberta (FDR)	0,894	0,875	0,351	0,028
Taxa de falsos negativos (FNR)	0,477	0,665	0,349	0,194
F_1 Score	0,184	0,211	0,709	0,928
Similaridade Média	0,105	0,130	0,602	0,937

zendo com que a área da região detectada não corresponda com a área do *ground truth*. Conforme pode-se observar na Tab. 8 os resultados mostram também uma melhora significativa do sistema de detecção proposto em relação ao estado da arte, melhorando as características de similaridade entre as detecções e o *ground truth*. Os valores de precisão e *recall* mostram que 97,1% das detecções foram corretas, no entanto, o sistema deixou de encontrar 19,5% das regiões anotadas. Além disso, o número de regiões detectadas, mostra que o contorno das regiões é extraído de forma mais precisa, reduzindo a fragmentação.

Já no caso de teste II, foram comparados apenas os detectores propostos em CHENEBERT; BRECKON; GASZCZAK (2011) e o detector proposto. Como ambos os detectores dependem de classificadores treinados, buscou-se, neste caso de teste, identificar o potencial destes classificadores quando aplicados a vídeos diferentes daqueles de onde foram extraídos os dados de treinamento. Desta forma, busca-se verificar se não houve um sobreajuste na etapa de treinamento. Os resultados para este caso de teste são dados na Tab. 9 e Tab. 10 que apresentam respectivamente o desempenho na classificação *frame-a-frame* e as métricas por localização. Para este caso de teste foram utilizados 3 vídeos, totalizando 1719 *frames* que contém chamadas (2054 regiões de chamada anotadas no *ground truth*) e 1032 *frames* que não contém chamadas.

Nota-se neste segundo caso de teste, que os resultados do detector proposto são bastante próximos aos resultados do detector proposto em CHENEBERT; BRECKON; GASZCZAK (2011). Nas métricas *recall* e *miss rate* por exemplo, o detector de CHENEBERT; BRECKON; GASZCZAK (2011) mostra resultados melhores. No entanto, quando são consideradas as métricas balanceadas, este é superado pelo detector proposto que apresenta uma acurácia, F_1 Score e coeficiente de correlação de Matthews maior. Isto se deve, em grande parte, à diferença 0,053 na precisão, que varia de 0,857 para 0,910, indicando uma redução no número de falsos positivos.

Ainda para o segundo caso de teste, quando são consideradas as métricas por locali-

Tabela 9: Resultados do método proposto na avaliação *frame a frame* para o caso de teste II. A coluna A apresenta os resultados CHENEBERT; BRECKON; GASZCZAK (2011). Os métodos foram avaliados utilizando os mesmos dados.

Métrica	A	Proposto
Recall (TPR)	0,990	0,966
Especificidade (SPC)	0,724	0,842
Precisão (PPV)	0,857	0,910
Valor preditivo negativo (NPV)	0,979	0,937
Fall-out (FPR)	0,275	0,157
Taxa de falsa descoberta (FDR)	0,142	0,089
Miss Rate (FNR)	0,009	0,033
Acurácia (ACC)	0,890	0,919
F_1 Score	0,919	0,937
Coefficiente de correlação de Matthews (MCC)	0,773	0,827

Tabela 10: Resultados do método proposto considerando localização para o caso de teste II. A coluna A apresenta os resultados de CHENEBERT; BRECKON; GASZCZAK (2011). Os métodos foram avaliados utilizando os mesmos dados.

Métrica	A	Proposto
Total de regiões detectadas	3461	3620
Total de regiões no <i>ground truth</i>	2054	2054
Precisão (L_{PPV})	0,832	0,800
Recall (L_{TPR})	0,979	0,969
Taxa de falsa descoberta (FDR)	0,167	0,199
Taxa de falsos negativos (FNR)	0,020	0,030
F_1 Score	0,902	0,880
Similaridade Média	0,801	0,794

zação, pode-se observar na Tab. 10 que os resultados do detector proposto são um pouco inferiores ao detector proposto em CHENEVERT; BRECKON; GASZCZAK (2011). Quando o modelo gerado utilizando um subconjunto dos vídeos do *dataset* e aplicado em outro subconjunto, tentando desta forma emular os resultados da aplicação deste detector em uma situação real (em que não existe um treinamento específico) a precisão do detector proposto é de 0,8, com uma taxa de *recall* de 0,969. A similaridade média entre as regiões detectadas é de 0,794.

Durante os testes algumas deficiências do detector puderam ser observadas. A primeira particularidade diz respeito ao reflexo que ocorre na água e no vidro. Este fenômeno, por ser facilmente ignorado em uma análise humana não é incluído no *ground truth* gerando falsos positivos na análise quantitativa. Esta é uma característica do detector, decorrente da abordagem proposta, para a qual ainda não foi vislumbrada uma solução.

Outra situação que implica na geração de falsos positivos é a presença de coletes reflexivos, como os utilizados por bombeiros e socorristas, na cena. Este tipo de vestimenta, quando em condições de iluminação adequadas, é capturado pela câmera com uma cor e textura muito próxima da que é apresentada pelo fogo verdadeiro, satisfazendo inclusive as regras que buscam pela obliquidade e formato da região.

Um terceiro gerador de falsos positivos é a presença de ruído estruturado no vídeo, normalmente uma consequência do uso de câmeras e *encoders* baratos. Quando existe ruído estruturado a definição das bordas é prejudicada. Dado que o sistema é baseado primariamente na classificação por cor e nos dados estatísticos da região, a quantidade de falsos positivos e falsos negativos aumenta quando existe movimento da câmera.

Em decorrência da análise do formato, notou-se que o detector falha sistematicamente quando existe oclusão parcial da região de fogo. No *dataset* utilizado, estes casos correspondem a fogo visto através de janelas ou dentro de tanques que impedem a visualização da base do fogo. Este problema é similar ao que já havia sido reportado na bibliografia revisada. Por fim, notou-se também a ocorrência de falsos negativos quando a chama é muito translúcida.

8 TRABALHOS FUTUROS

O trabalho apresentado representa uma contribuição significativa para o estado da arte em relação à detecção de chamas utilizando câmeras móveis. Visto que não foi possível encontrar na bibliografia consultada nenhuma referência que trabalhasse com objetivos próximos que são a detecção em tempo de execução, permitindo a tomada de decisão pelo robô e o acionamento de uma cadeia de eventos associados, não foi possível estabelecer um comparativo entre os resultados alcançados. A título de referência, cabe citar que BORGES; IZQUIERDO (2010) busca apenas classificar o arquivo de vídeo obtido de uma biblioteca de mídias como contendo fogo ou não.

Já no que se refere à detecção de fogo a partir de câmeras estacionárias pode-se afirmar que não existiu um grande avanço, dado que a série de métodos baseados na oscilação que foram apresentados no Cap. 3 já havia praticamente esgotado o assunto. Quando utiliza-se câmeras fixas e com pouca trepidação, o processo de detecção é significativamente mais simples.

Com relação às melhorias específicas do detector, sabe-se que o custo computacional da classificação com base na cor utilizando *Random Forests* para a classificação de cada pixel individualmente pode ser um impeditivo para a aplicação da solução em imagens de alta resolução. Por este motivo é interessante procurar uma abordagem que possa produzir resultados semelhantes, mas que não exija tantas operações sobre cada pixel. Uma das alternativas viáveis neste caso seria explorar massivamente o processamento paralelo utilizando a plataforma CUDA. Além disto, considerando os valores de *recall* discutidos no Cap. 7 pode-se focar na busca por detectores que minimizem a taxa de falsos negativos, melhorando a qualidade geral do detector.

Apesar dos resultados alcançados, existem diversas linhas de trabalho que não foram completamente exploradas até aqui. Entre elas pode-se destacar a integração completa com o *framework CvWorks*, adequando-o perfeitamente à arquitetura proposta pelo mesmo e adequando a documentação de forma a permitir o reaproveitamento do código no futuro. Neste processo incluem-se também a colaboração para o desenvolvimento dos módulos em compatibilidade com ROS (*Robot Operating System*) e com as arquiteturas e protocolos industriais utilizados. Da mesma forma, o suporte à múltiplas câmeras é

desejável.

Devido a falta de critérios objetivos para definir a área da região de fogo, uma das propostas futuras é que o *dataset* seja anotado por uma segunda pessoa, utilizando então as estatísticas de concordância para melhorar a consistência das anotações de *ground-truth*. Já no campo de pesquisa da aprendizagem de máquina, o problema de detecção de fogo pode ser também abordado utilizando técnicas de *deep learning*, como redes neurais profundas ou redes neurais profundas convolucionais.

9 CONCLUSÃO

Desde a década de 60, quando iniciaram-se os primeiros estudos no campo da visão computacional, houve uma grande evolução. A popularização de câmeras digitais e outros eletrônicos capazes de gravar e reproduzir vídeos fez com que tarefas que envolvessem a extração de dados a partir de imagens e vídeos se tornassem relevantes e permitissem a automação de diversas tarefas. Atualmente existe uma tendência de se usar o processamento de imagens nas mais diversas aplicações, desde o reconhecimento de caracteres até a detecção de fogo, passando por sistemas de autenticação e reconhecimento biométrico, sistemas de vigilância e monitoramento ou até mesmo processamento de imagens médicas.

Ao contrário do que acontece por exemplo com a detecção de face, para detecção e localização de incêndio, fogo ou fumaça não existem *datasets* públicos completos, com anotações que permitam uma validação automática dos resultados. Da mesma forma, também não existe uma convenção sobre as métricas. Por este motivo, precisou-se criar um *dataset* próprio composto por vídeos de câmeras não estacionárias para verificar as saídas produzidas. Considera-se que o trabalho foi satisfatório e permitiu atingir um maior conhecimento sobre a área de pesquisa como um todo, referenciando a maior parte dos algoritmos atualmente utilizados e elucidando suas principais vantagens e pontos fracos.

Com relação ao objetivo principal do trabalho, pode-se afirmar que o desenvolvimento de um detector de chamas que produz bons resultados em vídeos obtidos a partir de câmeras móveis foi atingido. Por outro lado, não é possível afirmar que o sistema permite a detecção em tempo real. Os detectores propostos na literatura foram estudados e seus principais pontos positivos e negativos foram elucidados. A comparação quantitativa do detector proposto com os trabalhos anteriores, utilizando um *dataset* desafiador, mostrou que as técnicas baseadas na textura da imagem fornecem melhores resultados. A avaliação qualitativa permitiu identificar quais os pontos em o detector proposto pode ser incrementado em trabalhos futuros.

Ao minimizar a dependência da avaliação temporal, a solução proposta mostra-se funcional tanto para vídeos obtidos a partir de câmeras móveis quanto para vídeos obtidos a partir de câmeras estáticas. A técnica proposta supera as principais desvantagens apre-

sentadas pelos outros métodos estudados. Os próximos passos, apresentados com maiores detalhes no Cap. 8, estão primariamente relacionados com a otimização em nível de código e a integração completa com o *framework* CvWorks.

REFERÊNCIAS

AGRAWAL, R.; SRIKANT, R. Mining sequential patterns. In: DATA ENGINEERING, 1995. PROCEEDINGS OF THE ELEVENTH INTERNATIONAL CONFERENCE ON, 1995. **Anais...** [S.l.: s.n.], 1995. p.3–14.

BERG, T. L.; BERG, A. C.; EDWARDS, J.; MAIRE, M.; WHITE, R.; TEH, Y.-W.; LEARNED-MILLER, E.; FORSYTH, D. A. Names and faces in the news. In: COMPUTER VISION AND PATTERN RECOGNITION, 2004. CVPR 2004. PROCEEDINGS OF THE 2004 IEEE COMPUTER SOCIETY CONFERENCE ON, 2004. **Anais...** [S.l.: s.n.], 2004. v.2, p.II–848.

BOOST C++ Libraries. [S.l.: s.n.], 2014. <http://www.boost.org>.

BORGES, P. V. K.; IZQUIERDO, E. A probabilistic approach for vision-based fire detection in videos. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.l.], v.20, n.5, p.721–731, 2010.

BOUWMANS, T.; EL BAF, F.; VACHON, B. Background modeling using mixture of gaussians for foreground detection-a survey. **Recent Patents on Computer Science**, [S.l.], v.1, n.3, p.219–237, 2008.

BRADSKI, G. The OpenCV Library. **Dr. Dobb's Journal of Software Tools**, [S.l.], 2000.

BREIMAN, L. Random Forests. **Machine Learning**, [S.l.], v.45, n.1, p.5–32, 2001.

BREIMAN, L.; FRIEDMAN, J.; STONE, C. J.; OLSHEN, R. A. **Classification and regression trees**. [S.l.]: CRC press, 1984.

BRICE, C. R.; FENNEMA, C. L. Scene analysis using regions. **Artificial intelligence**, [S.l.], v.1, n.3, p.205–226, 1970.

CANNY, J. A computational approach to edge detection. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], n.6, p.679–698, 1986.

CELIK, T. Fast and Efficient Method for Fire Detection Using Image Processing. **ETRI journal**, [S.l.], v.32, n.6, 2010.

CELIK, T.; DEMIREL, H. Fire detection in video sequences using a generic color model. **Fire Safety Journal**, [S.l.], v.44, n.2, p.147–158, 2009.

CELIK, T.; DEMIREL, H.; OZKARAMANLI, H.; UYGUROGLU, M. Fire detection using statistical color model in video sequences. **Journal of Visual Communication and Image Representation**, [S.l.], v.18, n.2, p.176–185, 2007.

ÇELIK, T.; OZKARAMANLI, H.; DEMIREL, H. Fire and smoke detection without sensors: image processing-based approach. In: EUROPEAN SIGNAL PROCESSING CONFERENCE, EUSIPCO, 15., 2007. **Anais...** [S.l.: s.n.], 2007. p.147–158.

ÇETIN, A. E.; DIMITROPOULOS, K.; GOUVERNEUR, B.; GRAMMALIDIS, N.; GÜNAY, O.; HABIBOGLU, Y. H.; TÖREYIN, B. U.; VERSTOCKT, S. Video fire detection—Review. **Digital Signal Processing**, [S.l.], v.23, n.6, p.1827–1843, 2013.

CHEN, J.; HE, Y.; WANG, J. Multi-feature fusion based fast video flame detection. **Building and Environment**, [S.l.], v.45, n.5, p.1113–1122, 2010.

CHEN, T.-H.; WU, P.-H.; CHIOU, Y.-C. An early fire-detection method based on image processing. In: IMAGE PROCESSING, 2004. ICIP'04. 2004 INTERNATIONAL CONFERENCE ON, 2004. **Anais...** [S.l.: s.n.], 2004. v.3, p.1707–1710.

CHENEVERT, A.; BRECKON, T. P.; GASZCZAK, A. A non-temporal texture driven approach to real-time fire detection. In: IMAGE PROCESSING (ICIP), 2011 18TH IEEE INTERNATIONAL CONFERENCE ON, 2011. **Anais...** [S.l.: s.n.], 2011. p.1741–1744.

CHINCHOR, N. MUC-4 Evaluation Metrics. In: CONFERENCE ON MESSAGE UNDERSTANDING, 4., 1992, Stroudsburg, PA, USA. **Proceedings...** Association for Computational Linguistics, 1992. p.22–29. (MUC4 '92).

CHUNYU, Y.; JUN, F.; JINJUN, W.; YONGMING, Z. Video fire smoke detection using motion and color features. **Fire technology**, [S.l.], v.46, n.3, p.651–663, 2010.

COLLINS, R. T.; LIPTON, A.; KANADE, T.; FUJIYOSHI, H.; DUGGINS, D.; TSIN, Y.; TOLLIVER, D.; ENOMOTO, N.; HASEGAWA, O.; BURT, P. et al. **A system for video surveillance and monitoring**. [S.l.]: Carnegie Mellon University, the Robotics Institute Pittsburg, 2000. v.2.

CUMANI, A. Edge detection in multispectral images. **CVGIP: Graphical models and image processing**, [S.l.], v.53, n.1, p.40–51, 1991.

CYGANEK, B.; SIEBERT, J. P. **An introduction to 3D computer vision techniques and algorithms**. [S.l.]: John Wiley & Sons, 2011.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: a large-scale hierarchical image database. In: COMPUTER VISION AND PATTERN RECOGNITION, 2009. CVPR 2009. IEEE CONFERENCE ON, 2009. **Anais...** [S.l.: s.n.], 2009. p.248–255.

DUNTEMAN, G. H. **Principal Components Analysis**. [S.l.]: SAGE Publications, 1989. n.Nº 69. (A Sage Publications).

FEI-FEI, L.; FERGUS, R.; PERONA, P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. **Computer Vision and Image Understanding**, [S.l.], v.106, n.1, p.59–70, 2007.

FORSYTH, D. A.; PONCE, J. **Computer Vision: a modern approach**. [S.l.]: Prentice Hall Professional Technical Reference, 2002.

GIL, A. C. Métodos e técnicas de pesquisa social. In: **Métodos e técnicas de pesquisa social**. [S.l.]: Atlas, 2010.

GOMEZ-RODRIGUEZ, F.; ARRUE, B.; OLLERO, A. Smoke monitoring and measurement using image processing: application to forest fires. In: AEROSENSE 2003, 2003. **Anais...** [S.l.: s.n.], 2003. p.404–411.

GÓMEZ-RODRÍGUEZ, F.; PASCUAL-PEÑA, S.; ARRUE, B.; OLLERO, A.; VIEGAS, D. et al. Smoke detection using image processing. In: FOREST FIRE RESEARCH AND WILDLAND FIRE SAFETY: PROCEEDINGS OF IV INTERNATIONAL CONFERENCE ON FOREST FIRE RESEARCH 2002 WILDLAND FIRE SAFETY SUMMIT, LUSO, COIMBRA, PORTUGAL, 18-23 NOVEMBER 2002., 2002. **Anais...** [S.l.: s.n.], 2002.

GÜNAY, O.; TAŞDEMİR, K.; TÖREYİN, B. U.; ÇETİN, A. E. Fire detection in video using LMS based active learning. **Fire technology**, [S.l.], v.46, n.3, p.551–577, 2010.

GÜNAY, O.; TAŞDEMİR, K.; UĞUR TÖREYİN, B.; ENİS ÇETİN, A. Video based wildfire detection at night. **Fire Safety Journal**, [S.l.], v.44, n.6, p.860–868, 2009.

HALL, M. A. **Correlation-based feature selection for machine learning**. 1999. Tese (Doutorado em Ciência da Computação) — The University of Waikato.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA data mining software: an update. **ACM SIGKDD Explorations Newsletter**, [S.l.], v.11, n.1, p.10–18, 2009.

HAMINS, A.; YANG, J.; KASHIWAGI, T. An experimental investigation of the pulsation frequency of flames. In: SYMPOSIUM (INTERNATIONAL) ON COMBUSTION, 1992. **Anais...** [S.l.: s.n.], 1992. v.24, n.1, p.1695–1702.

HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. H. Textural features for image classification. **Systems, Man and Cybernetics, IEEE Transactions on**, [S.l.], n.6, p.610–621, 1973.

HORN, B. K.; SCHUNCK, B. G. Determining optical flow. In: TECHNICAL SYMPOSIUM EAST, 1981., 1981. **Anais...** [S.l.: s.n.], 1981. p.319–331.

HORNG, W.-B.; PENG, J.-W.; CHEN, C.-Y. A new image-based real-time flame detection method using color analysis. In: NETWORKING, SENSING AND CONTROL, 2005. PROCEEDINGS. 2005 IEEE, 2005. **Anais...** [S.l.: s.n.], 2005. p.100–105.

HOROWITZ, S. L.; PAVLIDIS, T. Picture segmentation by a tree traversal algorithm. **Journal of the ACM (JACM)**, [S.l.], v.23, n.2, p.368–388, 1976.

JAIN, V.; LEARNED-MILLER, E. **FDDDB**: a benchmark for face detection in unconstrained settings. [S.l.]: University of Massachusetts, Amherst, 2010. (UM-CS-2010-009).

JIAO, Y.; WEIR, J.; YAN, W. Flame Detection in Surveillance. **Journal of Multimedia**, [S.l.], v.6, n.1, 2011.

KAMEDA, Y.; MINOH, M. A human motion estimation method using 3-successive video frames. In: INTERNATIONAL CONFERENCE ON VIRTUAL SYSTEMS AND MULTIMEDIA, 1996. **Anais...** [S.l.: s.n.], 1996. p.135–140.

KIRSCH, R. A. Computer determination of the constituent structure of biological images. **Computers and biomedical research**, [S.l.], v.4, n.3, p.315–328, 1971.

KO, B. C.; HAM, S. J.; NAM, J. Y. Modeling and formalization of fuzzy finite automata for detection of irregular fire flames. **Circuits and Systems for Video Technology, IEEE Transactions on**, [S.l.], v.21, n.12, p.1903–1912, 2011.

KOLESOV, I.; KARASEV, P.; TANNENBAUM, A.; HABER, E. Fire and smoke detection in video with optimal mass transport based optical flow and neural networks. In: IMAGE PROCESSING (ICIP), 2010 17TH IEEE INTERNATIONAL CONFERENCE ON, 2010. **Anais...** [S.l.: s.n.], 2010. p.761–764.

KOSCHAN, A.; ABIDI, M. Detection and classification of edges in color images. **Signal Processing Magazine, IEEE**, [S.l.], v.22, n.1, p.64–73, 2005.

KRSTINIĆ, D.; STIPANIČEV, D.; JAKOVČEVIĆ, T. Histogram-based smoke segmentation in forest fire detection system. **Information Technology and Control**, [S.l.], v.38, n.3, p.237–244, 2009.

LABATI, D.; GENOVESE, A.; PIURI, V.; SCOTTI, F. Wildfire smoke detection using computational intelligence techniques enhanced with synthetic smoke plume generation. **IEEE Transactions on Systems, Man, and Cybernetics: Systems**, [S.l.], v.43, n.4, p.1003–1012, 2013.

LEE, B.; HAN, D. Real-time fire detection using camera sequence image in tunnel environment. In: **Advanced Intelligent Computing Theories and Applications. With Aspects of Theoretical and Methodological Issues**. [S.l.]: Springer, 2007. p.1209–1220.

LI, Y.-C.; WU, W.-C. Visual Fire Detection Based on Data Mining Technique. **International Conference on Robot, Vision and Signal Processing**, Los Alamitos, CA, USA, v.0, p.328–331, 2011.

LI, Y.-C.; WU, W.-C. Sequential Pattern Technology for Visual Fire Detection. **Environments**, [S.l.], v.9, p.13, 2012.

LUCAS, B. D.; KANADE, T. et al. An iterative image registration technique with an application to stereo vision. In: IJCAI, 1981. **Anais...** [S.l.: s.n.], 1981. v.81, p.674–679.

MARBACH, G.; LOEPFE, M.; BRUPBACHER, T. An image processing technique for fire detection in video images. **Fire safety journal**, [S.l.], v.41, n.4, p.285–289, 2006.

MARR, D.; HILDRETH, E. Theory of edge detection. **Proceedings of the Royal Society of London. Series B. Biological Sciences**, [S.l.], v.207, n.1167, p.187–217, 1980.

MATTHEWS, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. **Biochimica et Biophysica Acta (BBA)-Protein Structure**, [S.l.], v.405, n.2, p.442–451, 1975.

MCFARLANE, N. J.; SCHOFIELD, C. P. Segmentation and tracking of piglets in images. **Machine Vision and Applications**, [S.l.], v.8, n.3, p.187–193, 1995.

MIGLIORE, D. A.; MATTEUCCI, M.; NACCARI, M. A reevaluation of frame difference in fast and robust motion detection. In: ACM INTERNATIONAL WORKSHOP ON VIDEO SURVEILLANCE AND SENSOR NETWORKS, 4., 2006. **Proceedings...** [S.l.: s.n.], 2006. p.215–218.

MUBARAK, S. Fundamentals of Computer Vision. **University of Central Florida. Orlando, Florida, USA**, [S.l.], 1997.

MUELLER, M.; KARASEV, P.; KOLESOV, I.; TANNENBAUM, A. Optical Flow Estimation for Flame Detection in Videos. **Image Processing, IEEE Transactions on**, [S.l.], v.22, n.7, p.2786–2797, 2013.

NGUYEN-TI, T.; NGUYEN-PHUC, T.; DO-HONG, T. Fire detection based on video processing method. In: **ADVANCED TECHNOLOGIES FOR COMMUNICATIONS (ATC), 2013 INTERNATIONAL CONFERENCE ON**, 2013. **Anais...** [S.l.: s.n.], 2013. p.106–110.

OSKOEI, M. A.; HU, H. **A survey on edge detection methods**. [S.l.]: Technical Report CES, 2010.

PAL, N. R.; PAL, S. K. A review on image segmentation techniques. **Pattern recognition**, [S.l.], v.26, n.9, p.1277–1294, 1993.

PARK, J.; KO, B.; NAM, J.-Y.; KWAK, S. Wildfire smoke detection using spatiotemporal bag-of-features of smoke. In: **APPLICATIONS OF COMPUTER VISION (WACV), 2013 IEEE WORKSHOP ON**, 2013. **Anais...** [S.l.: s.n.], 2013. p.200–205.

PARKER, J. **Algorithms for Image Processing and Computer Vision**. [S.l.]: Wiley, 2010. (IT Pro).

PEI, J.; PINTO, H.; CHEN, Q.; HAN, J.; MORTAZAVI-ASL, B.; DAYAL, U.; HSU, M.-C. Prefixspan: mining sequential patterns efficiently by prefix-projected pattern growth. In: **IEEE 29TH INTERNATIONAL CONFERENCE ON DATA ENGINEERING (ICDE)**, 2013., 2001. **Anais...** [S.l.: s.n.], 2001. p.0215–0215.

PHILLIPS III, W.; SHAH, M.; VITORIA LOBO, N. da. Flame recognition in video. **Pattern recognition letters**, [S.l.], v.23, n.1, p.319–327, 2002.

PREWITT, J. M. **Object enhancement and extraction**. [S.l.]: Academic Press, New York, 1970. v.75.

ROSENFELD, A.; THURSTON, M. Edge and curve detection for visual scene analysis. **Computers, IEEE Transactions on**, [S.l.], v.100, n.5, p.562–569, 1971.

RUSSELL, B. C.; TORRALBA, A.; MURPHY, K. P.; FREEMAN, W. T. LabelMe: a database and web-based tool for image annotation. **International journal of computer vision**, [S.l.], v.77, n.1-3, p.157–173, 2008.

SCHMID, C. Constructing models for content-based image retrieval. In: **COMPUTER VISION AND PATTERN RECOGNITION, 2001. CVPR 2001. PROCEEDINGS OF THE 2001 IEEE COMPUTER SOCIETY CONFERENCE ON**, 2001. **Anais...** [S.l.: s.n.], 2001. v.2, p.II–39.

SCHROFF, F.; CRIMINISI, A.; ZISSERMAN, A. Harvesting image databases from the web. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.33, n.4, p.754–766, 2011.

SHIDIK, G. F.; ADNAN, F. N.; SUPRIYANTO, C.; PRAMUNENDAR, R. A.; AN-DONO, P. N. Multi Color Feature, Background Subtraction and Time Frame Selection for Fire Detection. **Jurnal Informatika**, [S.l.], 2014.

SOBEL, I.; FELDMAN, G. A 3x3 isotropic gradient operator for image processing. **a talk at the Stanford Artificial Project in**, [S.l.], p.271–272, 1968.

STAUFFER, C.; GRIMSON, W. E. L. Adaptive background mixture models for real-time tracking. In: **COMPUTER VISION AND PATTERN RECOGNITION, 1999. IEEE COMPUTER SOCIETY CONFERENCE ON**, 1999. **Anais...** [S.l.: s.n.], 1999. v.2.

TÖREYİN, B. U.; CETİN, A. E. Wildfire detection using LMS based active learning. In: **ACOUSTICS, SPEECH AND SIGNAL PROCESSING, 2009. ICASSP 2009. IEEE INTERNATIONAL CONFERENCE ON**, 2009. **Anais...** [S.l.: s.n.], 2009. p.1461–1464.

TÖREYİN, B. U.; DEDEOĞLU, Y.; CETİN, A. E. Contour based smoke detection in video using wavelets. In: **EUROPEAN SIGNAL PROCESSING CONFERENCE**, 2006. **Anais...** [S.l.: s.n.], 2006. p.123–128.

TÖREYİN, B. U.; DEDEOĞLU, Y.; CETİN, A. E. et al. Wavelet based real-time smoke detection in video. In: **EUROPEAN SIGNAL PROCESSING CONFERENCE**, 2005. **Anais...** [S.l.: s.n.], 2005. p.4–8.

TÖREYİN, B. U.; DEDEOĞLU, Y.; GÜDÜKBAY, U.; CETİN, A. E. Computer vision based method for real-time fire and flame detection. **Pattern Recognition Letters**, [S.l.], v.27, n.1, p.49–58, 2006.

TRIGO, T. **Equipamento Fotografico. Teoria E Pratica**. [S.l.]: SENAC, 1998.

WREN, C. R.; AZARBAYEJANI, A.; DARRELL, T.; PENTLAND, A. P. Pfnder: real-time tracking of the human body. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, [S.l.], v.19, n.7, p.780–785, 1997.

XIAO, J.; HAYS, J.; EHINGER, K. A.; OLIVA, A.; TORRALBA, A. Sun database: large-scale scene recognition from abbey to zoo. In: **COMPUTER VISION AND PATTERN RECOGNITION (CVPR), 2010 IEEE CONFERENCE ON**, 2010. **Anais...** [S.l.: s.n.], 2010. p.3485–3492.

YAMAGISHI, H.; YAMAGUCHI, J. A contour fluctuation data processing method for fire flame detection using a color camera. In: **INDUSTRIAL ELECTRONICS SOCIETY, 2000. IECON 2000. 26TH ANNUAL CONFERENCE OF THE IEEE**, 2000. **Anais...** [S.l.: s.n.], 2000. v.2, p.824–829.

YIN, F.; MAKRIS, D.; VELASTIN, S. A. Performance evaluation of object tracking algorithms. In: IEEE INTERNATIONAL WORKSHOP ON PERFORMANCE EVALUATION OF TRACKING AND SURVEILLANCE, RIO DE JANEIRO, BRAZIL, 2007. **Anais...** [S.l.: s.n.], 2007.

ZHANG, Y. J. A survey on evaluation methods for image segmentation. **Pattern recognition**, [S.l.], v.29, n.8, p.1335–1346, 1996.

ZHANG, Z.; ZHAO, J.; ZHANG, D.; QU, C.; KE, Y.; CAI, B. Contour based forest fire detection using FFT and wavelet. In: COMPUTER SCIENCE AND SOFTWARE ENGINEERING, 2008 INTERNATIONAL CONFERENCE ON, 2008. **Anais...** [S.l.: s.n.], 2008. v.1, p.760–763.

ZHAO, J.; ZHANG, Z.; HAN, S.; QU, C.; YUAN, Z.; ZHANG, D. SVM based forest fire detection using static and dynamic features. **Computer Science and Information Systems**, [S.l.], v.8, n.3, p.821–841, 2011.

ZHOU, X.-L.; YU, F.-X.; WEN, Y.-C.; LU, Z.-M.; SONG, G.-H. Early fire detection based on flame contours in video. **Information Technology Journal**, [S.l.], v.9, n.5, p.899–908, 2010.

ZIOU, D.; TABBONE, S. et al. Edge detection techniques-an overview. **PATTERN RECOGNITION AND IMAGE ANALYSIS C/C OF RASPOZNAVANIYE OBRAZOV I ANALIZ IZOBRAZHENII**, [S.l.], v.8, p.537–559, 1998.