

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E ELETRÔNICA
CURSO DE GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Petrus Semprebom Massabki

Deteção de fogo em vídeo: um método de classificação baseado em descritores estatísticos e máquinas de vetores de suporte

Florianópolis

2019

Petrus Semprebom Massabki

Detecção de fogo em vídeo: um método de classificação baseado em descritores estatísticos e máquinas de vetores de suporte

Trabalho de Conclusão de Curso submetido ao Departamento de Engenharia Elétrica e Eletrônica da Universidade Federal de Santa Catarina para obtenção do título de Bacharel em Engenharia Elétrica.
Orientador: Prof. Dr. Joceli Mayer

Florianópolis

2019

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Massabki, Petrus Semprebom

Detecção de fogo em vídeo : um método de classificação baseado em descritores estatísticos e máquinas de vetores de suporte / Petrus Semprebom Massabki ; orientador, Joceli Mayer, 2019.

92 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia Elétrica, Florianópolis, 2019.

Inclui referências.

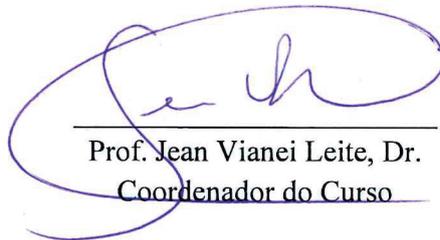
1. Engenharia Elétrica. 2. Detecção de fogo em vídeo. 3. Computação visual. 4. Análise de descritores. 5. Aprendizagem de máquina. I. Mayer, Joceli. II. Universidade Federal de Santa Catarina. Graduação em Engenharia Elétrica. III. Título.

Petrus Semprebom Massabki

Detecção de fogo em vídeo: um método de classificação baseado em descritores estatísticos e máquinas de vetores de suporte

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do Título de Bacharel em Engenharia Elétrica e aprovado em sua forma final pela Banca Examinadora

Florianópolis, 06 de dezembro de 2019.



Prof. Jean Viane Leite, Dr.
Coordenador do Curso

Banca Examinadora:



Prof. Joceli Mayer, Dr.
Orientador
Instituição UFSC



Prof. Bruno Catarino Bispo, Dr.
Avaliador
Instituição UFSC



Prof. Danilo Silva, Dr.
Avaliador
Instituição UFSC

Dedico este trabalho ao meu querido amigo Rodrigo Martins do
Amaral Varela

AGRADECIMENTOS

Agradeço aos meus pais, Antônio e Marlene, pelo suporte de uma vida toda, por sua preocupação e infindável paciência. À minha irmã, Nárria, por sua sutil, mas sempre notável capacidade de me transmitir segurança. Aos meus amigos Antônio e Stéfano, parceiros da faculdade, por permanecerem presentes até a conclusão deste trabalho. À Priscilla, por ter compartilhado carinhosamente seu tempo comigo, por ter tornado suportáveis os dias mais difíceis e por ter feito de pequenas vitórias imensas alegrias. Ao meu orientador, professor Joceli Mayer, por sua solicitude e por suas valiosas contribuições. Ao professor Pacheco e ao Maycon, por sua prestatividade em todas as visitas que precisei fazer à secretaria do departamento. Aos meus amigos do colegial e outros de mais longa data que ainda me apoiam diariamente - especialmente à Amanda Corazza e ao Lucas Correia, por sua contribuição direta neste trabalho; a ela, pela ajuda prestada na anotação manual de parte do conjunto de máscaras de segmentação, a ele, por ter cedido gentilmente algumas de suas fotografias. Aos demais colegas, amigos e professores da faculdade, por terem participado de algum modo em minha formação ou por terem simplesmente facilitado meu caminho até aqui.

RESUMO

Este trabalho de conclusão de curso apresenta um método de detecção de fogo em vídeo utilizando descritores estatísticos e máquinas de vetores de suporte. O projeto enfatiza o estudo e a comparação de diversas técnicas de segmentação por cor sugeridas na literatura. Um conjunto de medidas estatísticas provenientes de informações de cor e movimento é avaliado no reconhecimento de fogo em vídeo. Investiga-se, principalmente, a aplicabilidade do histórico e do gradiente de movimento na discriminação entre movimento ordinário e o observado nas chamas. É mostrado que tanto os descritores de cor quanto os de movimento produzem bons resultados quando usados separadamente, mas que sua associação é preferível. Confirma-se que a metodologia apresentada permite a classificação em tempo real de fogo em vídeo.

Palavras-chave: Detecção de fogo em vídeo. Computação visual. Análise de descritores. Aprendizagem de máquina. Máquina de vetores de suporte.

ABSTRACT

This undergraduate thesis presents a video fire detection method using statistical descriptors and support vector machines. The project emphasizes the study and comparison of several color segmentation techniques suggested in literature. A set of statistical measures derived from color and motion information is evaluated on the recognition of fire in video. It is investigated, mainly, the applicability of motion history and motion gradient in discriminating ordinary motion from the one observed in flames. Both color and motion descriptors are shown to produce good results when used separately, but their association is preferable. It is confirmed that the proposed methodology allows real time classification of fire in video.

Keywords: Video fire detection. Computer vision. Feature analysis. Machine learning. Support vector machines.

LISTA DE FIGURAS

Figura 1 - Exemplo de uma Imagem Binária	23
Figura 2 - Imagem colorida e divisão de canais	23
Figura 3 - Subconjuntos de pixels de uma imagem binária.....	25
Figura 4 - Filtragem espacial	27
Figura 5 - Exemplo de histograma	30
Figura 6 - Melhor margem de separação entre classes	33
Figura 7 - Gesto acompanhado das imagens do histórico e da energia do movimento.....	54
Figura 8 - Gradiente do movimento	55
Figura 9 - Exemplo de imagem do <i>dataset</i> proposto.....	63
Figura 10 - Histogramas das regiões anotadas como "fogo" e "não fogo”.....	65
Figura 11 - Separação das classes "fogo" e "não fogo" no plano R-S.....	67
Figura 12 – Curvas COR para otimização de <i>RT</i> e <i>ST</i>	69
Figura 13 – Curva COR para otimização do limiar em Phillips, Shah e Lobo (2000).....	70
Figura 14 - Representação da "nuvem" de pixels cor de fogo.....	71
Figura 15 – Exemplos dos resultados obtidos por cada método de segmentação avaliado	74
Figura 16 - Marcação das regiões candidatas a fogo.....	76
Figura 17 - Interface de rotulação.....	77
Figura 18 - Instâncias de um exemplo do conjunto de dados.....	78
Figura 19 - Diferença entre o movimento ordinário e o desempenhado pelas chamadas	81
Figura 20 - Fluxograma	85

LISTA DE QUADROS

Quadro 1 - Conceitos de TP, TN, FP e FN.....	60
--	----

LISTA DE TABELAS

Tabela 1 - Modelos de cores na literatura da segmentação de fogo em imagens.....	41
Tabela 2 - Dados estatísticos dos pixels anotados como fogo.....	66
Tabela 3 - Dados estatísticos dos pixels anotados como não-fogo.....	66
Tabela 4 - Tempo relativo de execução sobre o <i>dataset 1</i>	72
Tabela 5 - Dados de desempenho por método de segmentação (<i>dataset 1</i>).....	72
Tabela 6 - Dados de desempenho por método de segmentação (<i>dataset 2</i>).....	72
Tabela 7 - Parâmetros ótimos do modelo SVM por cenário	83
Tabela 8 - Resultado da classificação do modelo sobre o <i>dataset</i> de teste	84

LISTA DE ABREVIATURAS E SIGLAS

BSD – *Berkeley Software Distribution*

COR – Característica de operação do receptor

FP – *False positive*

FPR – *False positive rate*

FN – *False negative*

MCC – *Matthews Correlation Coefficient*

MEI – *Motion Energy Image*

MHI – *Motion History Image*

RBF – *Radial Basis Function*

ROI – *Region of interest* (região de interesse)

SVM – *Support Vector Machine*

TP – *True positive*

TPR – *True positive rate*

TN – *True negative*

SUMÁRIO

1	INTRODUÇÃO	15
1.1	MOTIVAÇÃO	16
1.2	OBJETIVOS	19
1.2.1	Objetivo Geral.....	19
1.2.2	Objetivos Específicos	19
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	CONCEITOS DE PROCESSAMENTO DIGITAL DE IMAGENS	21
2.1.1	Imagem digital	21
2.1.1.1	<i>Imagem binária.....</i>	22
2.1.1.2	<i>Imagem colorida.....</i>	23
2.1.2	Relações entre pixels.....	24
2.1.2.1	<i>Vizinhança de um pixel.....</i>	24
2.1.2.2	<i>Conectividade, adjacência, regiões e contornos</i>	24
2.1.3	Operações com imagens	25
2.1.3.1	<i>Operações entre imagens binárias</i>	26
2.1.3.2	<i>Operações orientadas a vizinhança.....</i>	26
2.1.4	Transformações de intensidade e filtragem espacial.....	27
2.1.4.1	<i>Limiarização ou thresholding.....</i>	28
2.1.5	Modelos de cores	29
2.1.6	Histograma de uma imagem	29
2.1.7	Textura.....	30
2.2	APRENDIZAGEM DE MÁQUINA	31
2.2.1	Máquina de vetores de suporte (SVM)	32
2.2.2	Descritores	34
2.2.3	Validação cruzada	34

2.3	CARACTERÍSTICAS VISUAIS DO FOGO	35
3	RECURSOS UTILIZADOS	37
3.1	LINGUAGEM DE PROGRAMAÇÃO.....	37
3.1.1	NumPy	37
3.1.2	Scikit-learn	37
3.1.3	OpenCV	38
3.2	HARDWARE	38
4	LITERATURA DA DETECÇÃO DE FOGO BASEADA EM VÍDEO E ASPECTOS RELACIONADOS	39
4.1	SEGMENTAÇÃO POR COR	40
4.1.1	Segmentação baseada em regras	41
4.1.2	Segmentação por agrupamento de dados	44
4.1.3	Segmentação via histograma	46
4.1.4	Outras abordagens	48
4.2	SEGMENTAÇÃO E ANÁLISE DO MOVIMENTO	49
4.2.1	Subtração de fundo.....	50
4.2.2	Cintilação.....	51
4.2.3	Histórico do movimento	52
<i>4.2.3.1</i>	<i>Imagem da Energia do Movimento (MEI).....</i>	<i>53</i>
<i>4.2.3.2</i>	<i>Imagem do Histórico do Movimento (MHI).....</i>	<i>53</i>
4.2.4	Gradiente do histórico do movimento.....	54
4.3	DESCRITORES	55
4.4	TREINAMENTO	58
4.5	CLASSIFICAÇÃO	59
4.6	MÉTRICAS DE AVALIAÇÃO	59
4.6.1	F₁, F-score ou F-measure	60
4.6.2	MCC (Matthews Correlation Coefficient).....	61
4.6.3	Outras métricas de avaliação.....	61

5	DESENVOLVIMENTO.....	62
5.1	ALGORITMO PROPOSTO.....	62
5.2	ESTUDO DA SEGMENTAÇÃO PELA COR.....	63
5.2.1	Datasets.....	63
5.2.1.1	<i>Dataset 1 (proposto).....</i>	<i>64</i>
5.2.1.2	<i>Dataset 2 (“BoWFire”).....</i>	<i>64</i>
5.2.2	Dados estatísticos das regiões de interesse.....	64
5.2.3	Estudo comparativo dos métodos de segmentação por cor.....	68
5.2.3.1	<i>Processo de otimização.....</i>	<i>68</i>
5.2.3.2	<i>Desempenho dos métodos de segmentação.....</i>	<i>72</i>
5.3	SEGMENTAÇÃO DO MOVIMENTO.....	74
5.4	VARREDURA E ROTULAÇÃO.....	75
5.5	DESCRITORES.....	77
5.5.1	Dados estatísticos da cor.....	78
5.5.2	Dados estatísticos do movimento.....	80
5.6	VALIDAÇÃO E TREINAMENTO.....	82
5.7	RESULTADOS DE OTIMIZAÇÃO E CLASSIFICAÇÃO.....	82
5.7.1	Treinamento e otimização da classificação.....	83
5.7.2	Resultados da classificação sobre o dataset de teste.....	83
5.8	PANORAMA DE FUNCIONAMENTO.....	84
6	CONSIDERAÇÕES FINAIS.....	86
	REFERÊNCIAS.....	88

1 INTRODUÇÃO

A linha de frente da segurança contra incêndios é convencionalmente formada por detectores pontuais de fumaça. E não à toa: essa solução ainda é, na maior parte dos casos, a única comercialmente viável e de eficácia estatisticamente comprovada. Suas deficiências, no entanto, não devem ser ignoradas, especialmente quando o cenário tecnológico favorece a inspeção de novas alternativas. A computação visual, nesse contexto, atua com protagonismo crescente, não somente pela evolução das ferramentas e da capacidade de processamento, mas também em virtude da popularização das câmeras de vídeo. Julgando pelo grande número de estruturas residenciais, comerciais e industriais já monitoradas por circuitos internos de segurança, parece apropriado que essas instalações atuem também na detecção automática de incêndios. Embora a conveniência seja um fator comercial relevante, a alternativa baseada em vídeo apresenta também benefícios técnicos importantes: em ambientes amplos, fechados ou abertos, onde a detecção convencional de fumaça é insuficiente, o monitoramento por vídeo ainda é possível.

O interesse prático em tais sistemas teve início na década de 1990, quando surgiram na literatura soluções como a concebida por Healey et al. (1993). Entretanto, muitas das propostas sobre as quais este trabalho se embasou foram apresentadas somente uma década mais tarde: Phillips, Shah e Lobo (2000); Chen, Kao e Chang (2003); Horng, Peng e Chen (2005); Celik et al. (2007); Borges e Izquierdo (2010). Nesse período, a complexidade de alternativas variou conforme a evolução da computação visual, mas a estrutura básica foi preservada: em geral, a detecção de fogo em vídeo envolve etapas de segmentação, extração de atributos e classificação. A segmentação preliminar é quase sempre baseada em um perfil de cores concebido por análise estatística. Em associação, comumente se realiza a segmentação do movimento. Esse recorte é então submetido à etapa mais particular do projeto: a extração de descritores. Nesse ponto, são frequentes as análises de borda, forma, textura e frequência. Por fim, a classificação pode ser baseada em aprendizagem de máquina ou em critérios explícitos de detecção. Seja qual for o caso, é recorrente a busca por uma boa razão entre acurácia e complexidade computacional.

Embora aborde todas as etapas da estrutura clássica de um projeto de detecção de fogo em vídeo, este trabalho concentra esforços em um estudo comparativo de métodos de segmentação por cor. Sete métodos foram implementados e avaliados segundo sua velocidade e precisão. Para tanto, foram utilizados dois conjuntos distintos de imagens – um desenvolvido

especificamente para este trabalho e outro emprestado da literatura. No algoritmo proposto, a segmentação da cor é associada à segmentação do movimento, realizada por meio da imagem do histórico do movimento. A escolha dos descritores foi pautada pela intenção de se desenvolver uma solução de detecção em tempo real, o que motivou a pesquisa por medidas de fácil obtenção. Desse modo, optou-se uma abordagem baseada em medidas estatísticas simples, como a média, o desvio padrão e a obliquidade de grandezas conhecidamente úteis na discriminação do fogo, como alguns canais de cor e o perfil de movimento, particularmente o evidenciado pelo gradiente de movimento. Uma ferramenta foi desenvolvida para anotações manuais em uma série de vídeos selecionados, a partir dos quais foram produzidos os conjuntos de dados necessários para treinamento, validação e teste de um modelo de classificação por SVM. A decisão final do algoritmo, que se baseia na classificação por aprendizagem, ainda passa pela verificação de um critério explícito de detecção.

A princípio, este texto revisa alguns conceitos introdutórios do processamento digital de imagens, apresenta noções de aprendizagem de máquina e discorre sobre algumas características visuais do fogo (capítulo 2). Em sequência (capítulo 3), são apresentadas informações a respeito das ferramentas utilizadas no desenvolvimento – linguagem de programação, bibliotecas e hardware. A revisão bibliográfica (capítulo 4) concentra-se em apresentar os principais elementos da literatura explorados no desenvolvimento, priorizando a exposição dos métodos de segmentação baseados na cor. Dando continuidade a esse enfoque, o desenvolvimento (capítulo 5) apresenta o procedimento adotado no estudo dos métodos de segmentação. Este é, no entanto, somente o ponto de partida: seguem o processo de segmentação do movimento, a escolha de descritores, o treinamento e a exposição dos resultados de classificação. Na última seção (capítulo 6) são feitas as considerações finais.

1.1 MOTIVAÇÃO

Assim como em boa parte dos problemas de engenharia, a detecção de incêndios carece de soluções completas e infalíveis. A proposta que aqui se apresenta (baseada em vídeo), figura, portanto, dentro desse contexto: é meramente um complemento às soluções tradicionais, cada qual com suas deficiências específicas. O que sustenta uma pesquisa desta natureza, no que tange o valor acadêmico, é justamente a busca por preencher essas lacunas. E quando uma nova área do conhecimento aflora, a ciência encontra o melhor terreno para cultivar suas perguntas e colher suas respostas: a computação visual não se apresenta somente como uma

alternativa à visão biológica; também cria novas perspectivas. Diante desse cenário, é feito um convite - que não pode ser negado -, à exploração de um potencial latente. Lança-se, então, o questionamento: a máquina é capaz de reconhecer fogo visualmente? A literatura dá sinais de que sim. Uma breve consulta aos autores que avaliaram o problema é suficiente para encorajar essa linha de pesquisa, seja pela inspiração que os resultados positivos trazem, seja pela confiança no valor do que resta a ser explorado.

Ainda que se encontre inspiração na literatura e que as possibilidades de inovação por meio da computação visual sejam diversas, resta considerar a utilidade da solução que aqui se pretende propor. Por isso, sugerem-se as seguintes perguntas: as soluções tradicionais são insuficientes a ponto de justificarem uma nova investida? As propostas oferecidas pela literatura já não são exaustivas? O problema em questão merece mais atenção?

Sobre a insuficiência dos métodos tradicionais, não há muita margem para controvérsias: quando a solução almeja garantir a preservação da vida, qualquer insuficiência é grave e qualquer aperfeiçoamento é bem-vindo. Ao se avaliar, por exemplo, o típico cenário em que a detecção de incêndio é realizada exclusivamente por detectores de fumaça, uma grande lacuna fica evidente: por serem aparelhos que dependem do contato com a fumaça, sua atuação está sempre sujeita à proximidade com o foco de incêndio; quanto maior a distância, maior o tempo médio de reação. Isso não significa, é claro, que a validade dos detectores de fumaça esteja em questão - tampouco que qualquer outra alternativa possa substituí-los completamente -, mas é fundamental ter ciência de suas falhas para que se possa elaborar um aperfeiçoamento. Em um estudo publicado pelo *National Institute of Standards and Technology*, conduzido por Bukowski et al. (2007), o tempo de resposta de alguns tipos de detectores de fumaça foi aferido em uma série de condições. Na melhor delas, quando a residência possuía sensores instalados em todos os cômodos, o tempo médio mínimo de resposta foi de 30 segundos para detectores iônicos. O mesmo estudo avaliou também sensores de calor, cujos resultados foram ainda piores: para o estágio de chamas de um incêndio, o tempo de resposta foi de, no mínimo, 163 segundos (BUKOWSKI et al., 2007). Tendo em vista que se tratam de cenários otimistas, e que 30 segundos – ou 163 - já correspondem a um tempo significativo de alastramento das chamas, há claramente margem para melhorias. A solução por reconhecimento visual tem justamente as características que poderiam ser aproveitadas em tais condições: a possibilidade de operar à distância - especialmente útil em ambientes amplos - e a capacidade de atuar assim que um incêndio for discernível em vídeo, oferecendo uma possível redução considerável no tempo de resposta.

A literatura da detecção de fogo em vídeo, embora recente, é bastante ampla, o que pode gerar a suspeita de saturação nessa área. A verdade, no entanto, é que poucos desses estudos têm verdadeira significância: pode-se dizer que a grande maioria das publicações orbita em torno de uma dezena de trabalhos realmente relevantes. Daí vem a necessidade de explorar alternativas e de incentivar a produção de dados e resultados replicáveis e comparáveis.

Se o que foi exposto ainda não fosse o bastante, este trabalho se justificaria somente com base no conhecimento das consequências catastróficas dos acidentes envolvendo fogo. Infelizmente, no Brasil e no restante da América Latina não existem boas estatísticas sobre incêndios; conseqüentemente, o problema ganha um falso caráter de insignificância. A importância devida, ainda que com ressalvas, fica restrita a incidentes de grandes proporções, que ora resultam na morte de dezenas de pessoas, ora na perda de preciosos bens culturais. Do pouco conhecimento que se tem – e apesar disso –, transparece quase sempre a íntima relação desses acidentes com a imprudência e a negligência. Parte do que explica esse descuido é justamente a pouca importância que se dá à ativa compreensão do problema. Sem uma base de dados sólida e unificada, não existem meios de amarrar uma boa política pública de segurança contra incêndios, até mesmo quando um conjunto de normas especifica claramente as melhores práticas de instalação. Para se ter ideia da dimensão do que está sendo tratado, é preciso recorrer a informações esparsas: segundo reportagem divulgada pela Folha de São Paulo, “o sistema unificado de informações registra menos de 1% dos casos estimados [de incêndios no Brasil]. Calcula-se que sejam 300 mil incêndios (entre florestais e urbanos) [...] por ano, que resultam em cerca de 1.000 mortes. Mas, no site do Ministério da Integração constam apenas 56 incidentes em edificações, com uma vítima e sete feridos desde setembro de 2007.” (PAULUZE, 2019). O Instituto Sprinkler Brasil (2019) também oferece dados importantes compilados a partir do monitoramento diário dos incêndios noticiados no Brasil: em 2018 foram contabilizados 531 incêndios estruturais. Destes, a maioria envolveu estabelecimentos comerciais (190); em seguida, as estruturas mais afetadas foram depósitos, com 114 incidentes reportados. As estatísticas por estado mostram que Santa Catarina está atrás apenas de São Paulo em número de ocorrências: 77 contra 112. O próprio instituto reconhece, no entanto, que esses números devem representar menos de 3% da quantidade real de incidentes.

1.2 OBJETIVOS

Partindo de uma extensiva revisão bibliográfica e sugerindo modestos alongamentos ao estado da arte, o trabalho aqui desempenhado almeja reunir algum valor acadêmico. Também espera oferecer utilidade prática ao problema que aborda, sustentado pela convicção de sua relevância social. A seguir são especificados alguns dos objetivos que guiaram seu desenvolvimento.

1.2.1 Objetivo Geral

O esforço dedicado a esta pesquisa busca dar forma a um método completo de detecção de fogo baseado em vídeo. Das suas qualidades, espera-se principalmente que reconheça fogo de forma satisfatória, em tempo real, prevendo e lidando com a ocorrência de falsos positivos. Como não existe a pretensão de se apresentar uma solução geral, restringe-se o domínio de atuação a incêndios estruturais monitorados a curta distância.

1.2.2 Objetivos Específicos

Permeiam o desenvolvimento atividades que se mostram oportunas ou que, no mínimo, são inerentes ao processo. Como contêm em si um valor final, independente do cumprimento do objetivo geral, essas atividades assumem o caráter de objetivos específicos, os quais são listados a seguir:

- 1) Estudo e aplicação da linguagem Python e da biblioteca OpenCV no desenvolvimento de um método de detecção de fogo baseado em vídeo;
- 2) Estudo, implementação e avaliação de métodos de segmentação de fogo em imagens propostos na literatura;
- 3) Elaboração de um *dataset* destinado ao cumprimento do objetivo listado no item 2, composto por imagens e máscaras correspondentes;

- 4) Implementação de uma interface de rotulação, com a qual se espera acelerar o enfadonho processo de treinamento supervisionado.

2 FUNDAMENTAÇÃO TEÓRICA

Os fundamentos que compõem este capítulo fornecem elementos ao entendimento da discussão que se apresenta em seguida. Alguns dos conceitos abordados são certamente triviais a quem trabalha rotineiramente com computação visual e processamento digital de imagens, mas estão presentes por se acreditar que, a nível de graduação, um trabalho acadêmico deva servir de referência também aos leitores mais leigos. Em contrapartida, esta seção não deve ser encarada como uma introdução geral ao tema, dado que muitos dos conceitos tidos como introdutórios são omitidos por não serem relevantes à compreensão do texto.

2.1 CONCEITOS DE PROCESSAMENTO DIGITAL DE IMAGENS

Gonzalez e Woods (2008) entendem que o processamento digital de imagens envolve processos cujas entradas e saídas são imagens e, além disso, envolve processos de extração de atributos de imagens até - e inclusive - o reconhecimento de objetos individuais. Segundo esse critério, o processamento digital de imagens se limita à transformação, segmentação e descrição de imagens. Por sua vez, a visão computacional teria o papel de analisar e fazer sentido dos dados extraídos na etapa de processamento. Diante desse entendimento, fica claro que um projeto de visão computacional, como o compreendido por este trabalho, não pode existir sem a base ferramental composta pelo processamento digital de imagens. Portanto, como não poderia ser diferente, a fundamentação teórica se inicia com a exposição de alguns conceitos básicos da representação e manipulação de imagens digitais.

2.1.1 Imagem digital

Imagens são expressas matematicamente como funções bidimensionais na forma $f(x, y)$. Para imagens monocromáticas, o valor de f em uma coordenada (x, y) corresponde à sua intensidade I , cujos valores vão de $I_{mín}$ a $I_{máx}$:

$$I = f(x, y) \mid I_{mín} \leq I \leq I_{máx} \quad (1)$$

O intervalo $[I_{\min}, I_{\max}]$ é chamado de escala de cinza e é comumente deslocado para $[0, L - 1]$, tal que 0 é considerado preto e $L - 1$ é considerado branco. Os valores intermediários são níveis de cinza.

Para ser expressa em formato digital, uma imagem contínua é digitalizada no espaço, ao longo de seus eixos, e em amplitude. A digitalização da amplitude recebe o nome de *quantização*. Uma vez digitalizada, essa imagem pode ser representada de três formas básicas: como uma superfície, como uma matriz numérica ou como uma matriz de intensidade visual (ilustração típica da imagem em um monitor). A representação em forma de matriz numérica é especialmente útil no desenvolvimento de algoritmos em processamento digital de imagens:

$$f(x, y) = \begin{bmatrix} f(0, 0) & f(0, 1) & \dots & f(0, N - 1) \\ f(1, 0) & f(1, 1) & \dots & f(1, N - 1) \\ \vdots & \vdots & \ddots & \vdots \\ f(M, 0) & f(0, 0) & \dots & f(M - 1, N - 1) \end{bmatrix} \quad (2)$$

A equação (2) representa uma imagem de dimensões $M \times N$, tal que M e N são valores inteiros e positivos. Cada elemento dessa matriz é chamado de *pixel* e pode assumir, usualmente, 2^k valores distintos (níveis de intensidade), sendo k também um número inteiro e positivo. Costuma-se dizer que uma imagem com 2^k níveis possíveis de intensidade é uma “imagem de k bits”. Na maioria das aplicações, k é igual a oito. Isso significa que uma imagem monocromática típica possui pixels com intensidades que variam dentro do intervalo $[0, 2^8 - 1] = [0, 255]$.

2.1.1.1 Imagem binária

Uma imagem binária é uma imagem digital cujos pixels podem assumir apenas dois níveis de intensidade. Geralmente, a intensidade máxima é representada pela cor branca, de valor *um*, e a intensidade mínima é representada pela cor preta, de valor *zero*. Imagens binárias também podem ser chamadas de *máscaras*, quando são utilizadas para “recortar” objetos de interesse.

Figura 1 - Exemplo de uma Imagem Binária



Fonte (imagem original, à esquerda): *dataset "BoWFire"* (CHINO et al., 2015). À direita, máscara binária resultante da segmentação dos pixels cor-de-fogo pelo método proposto por Phillips, Shah e Lobo (2002)

2.1.1.2 Imagem colorida

Em uma imagem colorida, cada ponto previsto pela função $f(x, y)$ é uma associação de três grandezas que variam de um modelo para outro.

Figura 2 - Imagem colorida e divisão de canais



Fonte (imagem original): Lucas Correia – disponível em <http://lucascorreia.net/fotojornalismo>. No sentido horário, a partir do canto superior esquerdo: imagem original; canal vermelho; canal verde; canal azul.

A representação mais conhecida é a fornecida pelo modelo RGB, que adota as cores primárias vermelho, verde e azul para formar um domínio aproximado do espectro visível. Cada uma dessas grandezas pode ser expressa na forma de uma imagem monocromática, usualmente chamada de *canal*. Desse modo, é comum dizer que uma imagem colorida possui três canais de cor.

2.1.2 Relações entre pixels

Esta seção apresenta os conceitos de *vizinhança*, *conectividade*, *adjacência*, *regiões* e *contornos*, necessários ao entendimento de algumas das operações que surgem no decorrer do texto. Em certos casos, têm influência direta no funcionamento dos algoritmos utilizados.

2.1.2.1 Vizinhança de um pixel

Um pixel p de coordenadas (x, y) possui quatro vizinhos horizontais e verticais, cujas coordenadas são: $(x - 1, y)$, $(x + 1, y)$, $(x, y - 1)$ e $(x, y + 1)$. Esse conjunto de pixels é chamado de *vizinhança-4* de p e é expresso por $N_4(p)$. Os pixels imediatamente diagonais a p são expressos por $N_D(p)$ e têm as seguintes coordenadas: $(x + 1, y + 1)$, $(x + 1, y - 1)$, $(x - 1, y + 1)$ e $(x - 1, y - 1)$. A junção das vizinhanças $N_4(p)$ e $N_D(p)$ é chamada de *vizinhança-8* de p e é expressa por $N_8(p)$.

2.1.2.2 Conectividade, adjacência, regiões e contornos

Antes de se avaliar a adjacência de pixels em uma imagem, é preciso escolher o conjunto de valores de intensidade a se considerar. Em uma imagem binária, por exemplo, avalia-se a adjacência de pixels com valores iguais a um.

Além disso, os conceitos de *conectividade* e *adjacência* entre pixels e subconjuntos de pixels apenas fazem sentido quando se define um modelo de vizinhança. Dado um subconjunto S , dois pixels são ditos conexos em S se houver um caminho entre eles composto inteiramente de pixels em S . Este caminho pode, no entanto, ser um *caminho-4* (segundo a vizinhança-4) ou um *caminho-8* (segundo a vizinhança-8). Para qualquer pixel p em S , o conjunto de pixels conectados a ele é chamado de um *componente conexo* de S . Se um subconjunto for composto inteiramente de pixels conexos, ele é considerado um *conjunto conexo*. Por último, se um

subconjunto de pixels for um conjunto conexo, ele é chamado de *região*. Duas regiões são ditas adjacentes se sua união formar também um conjunto conexo. Na Figura 3 são representados quatro subconjuntos de pixels rotulados numericamente.

Figura 3 - Subconjuntos de pixels de uma imagem binária

1						3	3
1	1	1			3	3	3
	1	1			3	3	
		2	2				
		2	2		4	4	
	2	2		4	4	4	4

Fonte: autoria própria.

No exemplo ilustrado, cada subconjunto é um componente conexo em si, independente da vizinhança escolhida. Considerando-se a vizinhança-4, os subconjuntos 1 e 2 formam um conjunto conexo. Considerando-se a vizinhança-8, o subconjunto 4 também é agregado. Dessa forma, os subconjuntos 1, 2 e 4 formam uma região. O subconjunto 3, no entanto, é *disjunto* em relação aos demais e forma uma região independente.

O conjunto de regiões disjuntas é chamado de *frente (foreground)* da imagem. Os pixels complementares constituem o *fundo (background)*. O *contorno*, também chamado de *borda* ou *fronteira* de uma região, é o conjunto de pontos adjacentes dessa região com seu complemento.

2.1.3 Operações com imagens

Embora imagens digitais sejam representadas na forma matricial, suas operações não são realizadas segundo a teoria de álgebra linear. A multiplicação, exemplificada na equação a seguir, ilustra bem essa diferença:

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \times \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11} \times b_{11} & a_{12} \times b_{12} \\ a_{21} \times b_{21} & a_{22} \times b_{22} \end{bmatrix} \quad (3)$$

Observa-se que cada elemento (pixel) da primeira imagem é multiplicado pelo correspondente na segunda. A mesma ideia pode ser expandida para as demais operações aritméticas – soma, subtração e divisão. Vale ressaltar que esse tipo de operação exige que as imagens tenham dimensões iguais.

2.1.3.1 Operações entre imagens binárias

Quando se manipula imagens binárias, é comum adotar os conceitos de operações lógicas no cálculo da união, intersecção ou complemento das regiões de interesse.

Neste trabalho interessa-se particularmente pela operação lógica “e”, utilizada no cálculo da intersecção entre as máscaras de segmentação da cor e do movimento, que resulta no recorte final das regiões de interesse (ver seção 5.4).

2.1.3.2 Operações orientadas a vizinhança

Operações espaciais são operações realizadas diretamente aos pixels de uma imagem, podendo ser de três categorias: ponto a ponto; por vizinhança; e transformações geométricas.

Uma operação ponto a ponto consiste na alteração do valor de um pixel com base em um critério que não depende dos demais (o negativo de uma imagem é um exemplo do resultado de uma operação desse tipo).

Operações por vizinhança alteram o valor de um pixel com base nos valores de seus vizinhos. Um exemplo seria o cálculo da média local de uma imagem. Supondo uma região de 3×3 pixels, a intensidade do pixel central é dada pela média aritmética de cada um dos 9 pixels que a compõem:

$$\begin{bmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{bmatrix}$$

$$I_{22}' = \frac{I_{11} + I_{12} + I_{13} + I_{21} + I_{22} + I_{23} + I_{31} + I_{32} + I_{33}}{9} \quad (4)$$

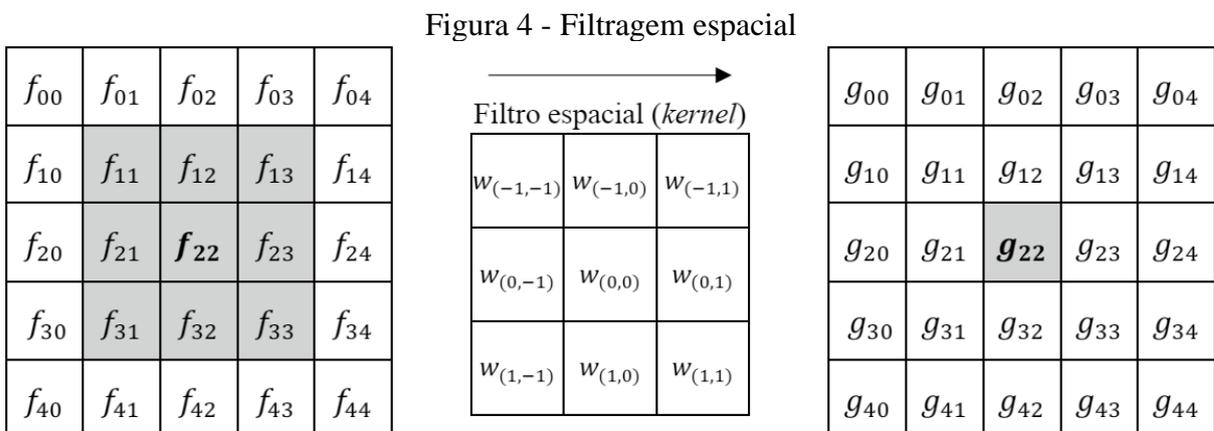
Transformações geométricas, por sua vez, são operações que alteram a relação espacial entre pixels (como escalonamento, rotação e translação).

2.1.4 Transformações de intensidade e filtragem espacial

Uma operação no domínio espacial pode ser expressa pela equação (5): $g(x, y)$ é o resultado de uma operação em $f(x, y)$ através de um operador T definido na vizinhança do ponto (x, y) . Em geral, a vizinhança é retangular e centrada em (x, y) , assim como no exemplo da média local dado anteriormente.

$$g(x, y) = T[f(x, y)] \quad (5)$$

Cada ponto da imagem é alterado de acordo com a vizinhança e o operador escolhido, num processo de varredura que se inicia no canto esquerdo superior da imagem e continua horizontalmente, linha a linha. Esse procedimento é chamado de *filtragem espacial*. O conjunto vizinhança-operador é chamado de *filtro espacial*, *máscara espacial*, *kernel*, *template* ou *janela*. Se a operação realizada sobre os pixels da imagem for linear, o filtro é chamado de *filtro espacial linear*. Caso contrário, o filtro é *não linear*.



Fonte: autoria própria.

Tomando o exemplo da Figura 4, a imagem resultante da filtragem espacial é gerada da seguinte forma: cada pixel f_{xy} é substituído pela soma da multiplicação do kernel genérico w com sua vizinhança-8. O elemento f_{22} , cuja vizinhança-8 está destacada, seria transformado dessa maneira:

$$g_{22} = w_{(-1,-1)}f_{11} + w_{(-1,0)}f_{12} + w_{(-1,1)}f_{13} + w_{(0,-1)}f_{21} + w_{(0,0)}f_{22} + w_{(0,1)}f_{23} + w_{(1,-1)}f_{31} + w_{(1,0)}f_{32} + w_{(1,1)}f_{33} \quad (6)$$

É importante ressaltar que os pixels das bordas da imagem são tratados de forma especial, já que não possuem uma vizinhança-8 completa. Esse problema pode ser resolvido de várias maneiras, conforme a aplicação. Uma solução simples consiste em definir uma nova borda ao redor da imagem original com valores predeterminados.

Quando a vizinhança tem dimensões 1×1 , o resultado $g(x, y)$ depende apenas do ponto (x, y) em $f(x, y)$. Nesse caso, T é uma *função de transformação de intensidade* da forma

$$j = T(i) \quad (7)$$

sendo j e i , respectivamente, os valores de intensidade em um ponto qualquer de g e f .

2.1.4.1 Limiarização ou *thresholding*

Uma função de transformação de intensidade muito importante é a de limiarização (ou *thresholding*), que consiste na criação de uma máscara binária de acordo com as seguintes regras:

$$g(x, y) = \begin{cases} 1 & \text{se } f(x, y) \geq k \\ e & \\ 0 & \text{se } f(x, y) < k \end{cases} \quad (8)$$

Cada valor de intensidade da imagem $f(x, y)$ é comparado à constante k , sendo transformado em 1 se a intensidade for maior ou igual a k e em 0 se a intensidade for menor que k . Um exemplo de aplicação dessa função pode ser encontrado na seção 5.3, onde consta o cálculo da máscara de movimento via limiarização.

2.1.5 Modelos de cores

Definido de maneira simples, “um modelo de cor é essencialmente uma representação tridimensional na qual cada cor é especificada por um ponto no sistema de coordenadas tridimensionais.” (PEDRINI; SCHWARTZ, 2008, p. 469). Destacam-se no texto os modelos RGB, HSV, YUV e YCbCr, razão pela qual são aqui introduzidos.

O modelo RGB, talvez o mais difundido entre os exemplos citados, tem seu nome derivado das iniciais das cores primárias que o compõem, *red*, *green* e *blue*. Baseia-se em um sistema tridimensional de coordenadas cartesianas e tem seu domínio delimitado por um cubo. O vértice que se encontra na origem representa a cor preta; o mais distante da origem, na outra ponta da diagonal principal, representa a cor branca. Os demais vértices que coincidem com os eixos do sistema de coordenadas correspondem às cores primárias, vermelho, verde e azul. No restante do cubo se encontram todas as outras combinações dessas componentes primárias. Por convenção, o valor de cada componente é normalizado entre 0 e 1; em formato digital de 8 bits, no entanto, assumem valores entre 0 e 255.

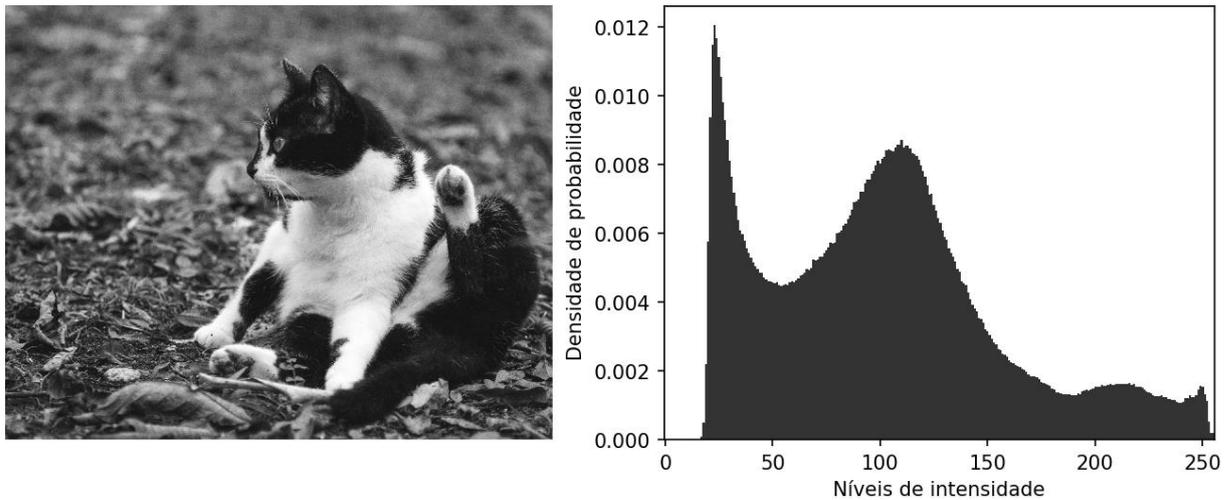
No modelo HSV, a cromaticidade e a luminância são representadas independentemente, de forma análoga à percepção humana das cores. Suas componentes, matiz (H, *hue*), saturação (S, *saturation*) e luminância (V, *value*), formam uma pirâmide de base hexagonal, que delimita o espaço de cores do seguinte modo: cada vértice da base corresponde a uma cor primária ou primária complementar (vermelho, verde, azul, ciano, magenta e amarelo); a matiz é, portanto, representada angularmente, entre 0 e 360°; a luminância acompanha o eixo vertical, crescendo no sentido vértice-base, e a saturação acompanha o eixo horizontal, crescendo no sentido centro-periferia; ambas são normalizadas entre 0 e 1.

Os modelos YUV e YCbCr possuem representações semelhantes: em ambos, Y corresponde à luminância, enquanto as demais componentes, U, V, Cb e Cr, codificam a cromaticidade. As transformações entre cada um desses modelos e o RGB podem ser dadas, linearmente, por meio de equações matriciais.

2.1.6 Histograma de uma imagem

O histograma de uma imagem digital é uma função discreta que relaciona os valores de intensidade com sua frequência de ocorrência. Um histograma pode ser compreendido como uma função de probabilidade, se for normalizado pela quantidade total de pixels da imagem.

Figura 5 - Exemplo de histograma



Fonte: autoria própria.

O modelo mais simples de histograma é o de uma imagem monocromática, representado por uma curva de duas dimensões (Figura 5). Imagens com mais de um canal possuem histogramas que relacionam a frequência de ocorrência a uma combinação de valores de cada canal. Em geral, a contagem é realizada para cada valor de intensidade individual, isto é, uma imagem monocromática de 8 bits, por exemplo, possui um histograma de 256 valores. Se for colorida, a mesma imagem no espaço RGB possui um histograma de $256^3 = 16.777.216$ valores – um para cada cor. Esse número deixa claro que, em aplicações que exigem alta velocidade, o cálculo do histograma completo pode se tornar um impedimento. Neste caso, é desejável agrupar pixels em faixas de valores de intensidade, conhecidas como *bins*, ainda que isso cause perda de informação.

2.1.7 Textura

A fim de poder descrever efetivamente agrupamentos de pixels, o processamento digital de imagens recorre a propriedades facilmente reconhecidas e interpretadas pelo sistema visual humano. Embora seja um caminho lógico, formalizar conceitos ou descritores capazes de traduzir essa percepção intuitiva não é uma tarefa simples. Em geral, a maior dificuldade consiste em elaborar soluções que incorporem domínios de aplicações diversos, o que acaba resultando em um conjunto de abordagens com vantagens e desvantagens particulares (PEDRINI; SCHWARTZ, 2008). A *textura*, por exemplo, é uma propriedade explorada por métodos de caracterização que podem ser divididos entre quatro grupos: estatísticos, geométricos, baseados em processamento de sinais e baseados em modelos paramétricos.

Diante desse cenário, o projetista deve avaliar qual abordagem satisfaz melhor as necessidades de sua aplicação. Neste trabalho, optou-se somente pela abordagem estatística, que compreende técnicas que objetivam representar as regiões de uma imagem segundo a distribuição e a relação entre os níveis de cinza dos pixels que as preenchem. Algumas dessas técnicas empreendem matrizes de coocorrência, capazes de fornecer descritores de textura relativamente robustos. Certas aplicações, entretanto, toleram o uso de medidas estatísticas mais simples, como a média, a variância e a assimetria. A seção 5.5 apresenta descritores que exploram justamente essas medidas.

2.2 APRENDIZAGEM DE MÁQUINA

Aprendizagem de máquina é o ramo da ciência responsável pelo estudo de algoritmos e modelos que buscam fornecer a um sistema computacional a capacidade de executar tarefas a partir da análise e compreensão de dados, recorrendo à inferência, ao reconhecimento de padrões e à estatística. Seu propósito é espelhar nas máquinas a habilidade humana de tomar decisões e prever resultados com base na experiência adquirida.

Aqui, *experiência* se refere à informação prévia viabilizada ao aprendiz, que tipicamente assume a forma de dados eletrônicos coletados e disponibilizados para análise. Esses dados podem estar na forma de conjuntos digitalizados e rotulados pelo homem, ou outros tipos de informações obtidas via interação com o ambiente. Em todo caso, seu tamanho e qualidade é crucial para o sucesso das previsões realizadas pelo aprendiz (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012, p. 1, tradução nossa).

De modo geral, os algoritmos abordados pela aprendizagem de máquina são divididos entre *supervisionados* e *não supervisionados*, que diferem tanto no grau de conhecimento prévio dos dados como na forma de tratá-los. A primeira categoria lida com dados cujas características e rótulos já são conhecidos e que, portanto, podem ser devidamente agrupados; a segunda lida com dados sobre os quais ainda se buscam informações que permitam seu devido tratamento. A fim de compreender melhor essa distinção, o resultado de um sistema de aprendizagem de máquina pode ser expresso na forma $y(x)$, em que x representa o vetor de atributos de uma amostra e y o rótulo associado a ela (BISHOP, 2006). Em outras palavras: o algoritmo deve reconhecer a natureza (*rótulo*) de uma instância a partir de suas características (*descritores*). No caso supervisionado, tanto as entradas x quanto as saídas $y(x)$ são conhecidas,

ou seja, existe um mapeamento entre as características e a natureza de cada amostra; no caso não supervisionado, somente as entradas são conhecidas, isto é, o sistema ainda precisa interpretar os dados e aprender seus atributos a fim de determinar suas saídas. Na aprendizagem supervisionada, quando as saídas de um modelo são restritas por um conjunto limitado de valores, o algoritmo correspondente busca resolver um problema de *classificação*; por outro lado, quando podem assumir quaisquer valores numéricos dentro de um intervalo, o algoritmo correspondente busca resolver um problema de *regressão*. (Este trabalho trata de um caso típico de classificação binária com treinamento supervisionado, visto que pretende realizar a distinção entre duas classes bem definidas – fogo e não fogo). A etapa de aprendizagem, comumente chamada de *treinamento*, é a que determina a forma precisa da função $y(x)$. A partir da definição de um modelo, este pode então determinar a identidade de novas amostras, que compõem o chamado *conjunto de teste*. Em condições adequadas, espera-se que o modelo resultante possua a habilidade de categorizar corretamente novos exemplos, diferentes daqueles usados para treinamento – qualidade conhecida por *generalização* (BISHOP, 2006). As seções seguintes elaboram um pouco mais os conceitos e ferramentas que surgem no decorrer deste trabalho, como os algoritmos de classificação e treinamento.

2.2.1 Máquina de vetores de suporte (SVM)

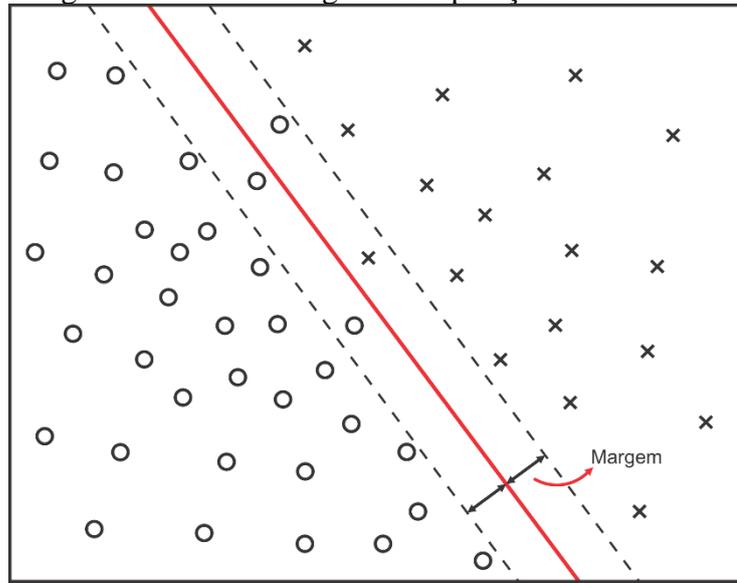
Máquinas de vetores de suporte são algoritmos de aprendizagem supervisionada utilizados para classificação de dados. A partir de um conjunto de treinamento, sua tarefa é determinar o modelo de hiperplano que melhor separa as classes envolvidas.

Em seu guia prático, Hsu, Chang e Lin (2003) declaram acreditar não ser necessário que um usuário compreenda a teoria fundamental por trás das SVMs a fim de aplicá-las com sucesso. Em concordância com essa opinião, são tratados aqui somente alguns conceitos básicos que ditam superficialmente o procedimento utilizado na classificação de fogo em vídeo.

Dado um conjunto de pares instância-rótulo, a máquina de vetores de suporte se propõe a resolver um problema de otimização que consiste em mapear os vetores de instâncias a um espaço de dimensão mais elevada e encontrar, nesse espaço, um hiperplano de separação com margem máxima. Ao problema estão associados um parâmetro $C > 0$, que atribui um fator de penalização ao erro, e uma função *kernel* K , que pode assumir, tipicamente, quatro formas básicas: linear, polinomial, base radial (RBF) e sigmoide (HSU; CHANG; LIN, 2003). Alguns desses *kernels*, por sua vez, possuem parâmetros específicos. A função de base radial, por

exemplo, apresenta o parâmetro γ (gamma), que pode ser entendido como o inverso do raio de influência das amostras selecionadas pelo modelo. Desse modo, um γ pequeno implica na interpretação de duas amostras distantes como semelhantes. Inversamente, quando γ é grande, dois pontos são considerados semelhantes somente quando estão próximos um do outro.

Figura 6 - Melhor margem de separação entre classes



Fonte: autoria própria.

Em síntese, o projeto de um modelo de classificação por SVM pode ser dividido nas seguintes etapas: processamento dos dados, otimização de parâmetros, treinamento e teste.

Antes que os dados sejam submetidos ao treinamento, é importante que passem por um processo de normalização. Evita-se, desse modo, que atributos de diferentes faixas de valores numéricos possuam pesos discrepantes. Além disso, o escalonamento dos dados ajuda a prevenir dificuldades numéricas (HSU; CHANG; LIN, 2003).

A otimização dos parâmetros está condicionada à escolha dos *kernels* avaliados pelo problema. No caso da função de base radial, por exemplo, o processo consiste em encontrar os melhores parâmetros C e γ . Para isso, uma das abordagens possíveis é a varredura sistemática de parâmetros. Sugere-se, além disso, o uso de validação cruzada, detalhada na seção 2.2.3.

Em posse dos parâmetros otimizados, o modelo de classificação pode, então, ser definido pelo conjunto completo de dados para treinamento. Por fim, o desempenho deve ser aferido a partir de um conjunto de teste, independente do primeiro, de modo a verificar com segurança a capacidade de *generalização*.

2.2.2 Descritores

No contexto da aprendizagem de máquina, um *descriptor* é uma propriedade, geralmente numérica, associada a um objeto de interesse; é a medida de uma característica específica de um exemplo do conjunto de dados. Duas ou mais características formam um *vetor de descritores*. Este corresponde, na prática, à instância quantitativa submetida a um algoritmo de reconhecimento.

Associar descritores relevantes a um exemplo “é uma etapa crítica no projeto de soluções de aprendizagem de máquina. Descritores úteis podem efetivamente guiar o algoritmo de aprendizagem, enquanto os pobres ou pouco informativos podem ser enganosos.” (MOHRI; ROSTAMIZADEH; TALWALKAR, 2012). Além de conter informações discriminatórias significativas, espera-se também que os descritores possam ser computados de forma suficientemente rápida, de acordo com as necessidades da aplicação. Neste trabalho, por exemplo, cujo objetivo é realizar a detecção de fogo em tempo real, a escolha dos descritores deve priorizar a eficiência.

No processamento de imagens, em geral, descritores são associados a propriedades de regiões de pixels, normalmente resultantes de um processo de segmentação. Por apresentarem características específicas, relacionadas à natureza do problema, a escolha de um descriptor deve ser feita de modo coerente. Descritores de forma, por exemplo, são úteis somente quando o problema apresenta propriedades geométricas facilmente observáveis. Outros modos de representação podem explorar a cor, a textura (ver seção 2.1.7) ou a borda de uma região.

2.2.3 Validação cruzada

Quando devidamente executado, um projeto de aprendizagem de máquina demanda três conjuntos de dados: um para treinamento, outro para validação e um terceiro para teste. O conjunto intermediário, de validação, tem a finalidade de evitar que a otimização do modelo seja influenciada pelo conjunto de teste, pois, na sua ausência, o desempenho obtido na etapa de teste acaba por servir de apoio à regulagem dos parâmetros de treinamento. Isso não é desejável porque a capacidade de generalização de um modelo deve ser verificada sobre um conjunto totalmente independente. Essa condição é facilmente satisfeita quando a disponibilidade de dados é abundante, pois basta que uma parcela aleatória seja destinada a cada um dos três conjuntos. Por outro lado, se a disponibilidade de dados for limitada e, por

consequência, o conjunto de validação for pequeno, a estimativa de desempenho pode ser relativamente ruidosa. Nesse caso, uma solução possível é o uso de *validação cruzada* (BISHOP, 2006). Na prática, a validação cruzada é utilizada tanto no treinamento quanto na escolha dos melhores parâmetros de um modelo.

O método de *validação cruzada de k-dobras (k-fold)* realiza, de forma aleatória, o particionamento dos dados disponíveis em k subgrupos de mesmo tamanho. Destes, $k - 1$ são usados para treinamento e o remanescente fica retido para validação. Repete-se o procedimento k vezes, cada qual com um grupo diferente de validação, de forma que um dado modelo seja treinado e validado com a participação de todos os exemplos observados. O desempenho do modelo, segundo o índice escolhido pelo projetista, é o resultado da média dos valores alcançados em todas as iterações. Dessa forma, cada combinação de parâmetros e algoritmos pode ser avaliada sistematicamente. Uma inconveniência deste método é a proporcionalidade entre o número de iterações e o de subdivisões realizadas no conjunto de dados: quando o treinamento é, em si, computacionalmente custoso, sucessivas repetições podem ser inviáveis (BISHOP, 2006).

2.3 CARACTERÍSTICAS VISUAIS DO FOGO

Fogo pode ser definido como um processo rápido de combustão que se manifesta em forma de luz, calor e chamas. Suas características físico-químicas dependem das substâncias presentes na reação e das condições em que a mesma ocorre, podendo variar em intensidade, temperatura e coloração. A parte visível e gasosa do fogo, que recebe o nome de *chama*, tende a apresentar movimento caótico e turbulento quando a queima é descontrolada – estado designado de *incêndio*. No tipo mais comum de chama, a de hidrocarboneto, a coloração depende principalmente da disponibilidade de oxigênio fornecido à reação; em condições normais, como no caso de incêndios estruturais, o espectro visível do fogo compreende tonalidades vermelho amareladas (principalmente amareladas). Por essa razão, a detecção de fogo em vídeo costuma se concentrar nessa faixa de cores.

Além da cor, outra marca visual do fogo é a cintilação. Embora não se possa determinar com exatidão um formato típico para o fogo - tampouco um modelo único de cores que o descreva -, há em sua dinâmica uma certa padronização. Desse modo, a análise em frequência é usualmente empregada no reconhecimento visual do fogo, com algumas ressalvas:

“A fim de capturar cintilação em 10 Hz, o vídeo deve capturar pelo menos 20 quadros por segundo (*fps*). Entretanto, em alguns sistemas de vigilância, a taxa de captura está abaixo de 20 Hz. Se o vídeo está disponível em uma taxa menor de captura, *aliasing* ocorre, mas a cintilação causada pelas chamas ainda pode ser observada em vídeo. Por exemplo, uma senoide de 8 Hz aparece como uma senoide de 2 Hz em um vídeo com 10 *fps*.” (GÜNAY et al., 2009, p. 557, tradução nossa).

A afirmação anterior menciona a frequência de 10 Hz em razão do conhecimento, já difundido, de que a frequência de oscilação das chamas repousa sobre esse valor: “[...] a taxa de vibração para as chamas de todos os gases investigados é da ordem de 10 por segundo. Essa taxa não é muito afetada pela atmosfera em que a chama queima ou pela ponta a partir da qual o gás é emitido, ou pela taxa de fluxo de gás.” (CHAMBERLIN; ROSE, 1948, p. 27, tradução nossa).

3 RECURSOS UTILIZADOS

A seguir são brevemente apresentados os recursos empregados na execução do trabalho proposto. Devido à ausência de um produto concreto, como um protótipo, esses recursos se limitam aos meios de implementação do projeto: linguagem de programação, bibliotecas e hardware.

3.1 LINGUAGEM DE PROGRAMAÇÃO

As metodologias investigadas na literatura, ferramentas auxiliares e blocos que compõem o algoritmo final de detecção foram todos programados em linguagem Python (versão 3.7), escolhida principalmente por ser de fácil aprendizagem. Sua difusão na academia também a torna atraente pela disponibilidade de conteúdo e pelo constante suporte oferecido a suas bibliotecas. Dentre elas, três foram de fundamental importância ao desenvolvimento deste trabalho: NumPy, Scikit-learn e OpenCV.

O ambiente integrado de desenvolvimento utilizado foi o PyCharm Community Edition (versão 2019.1.3), publicado sob a Licença Apache, licença permissiva de software livre.

3.1.1 NumPy

Em sua documentação, o NumPy é descrito como o pacote fundamental para computação científica em Python. Oferece suporte a arrays e matrizes multidimensionais e conta com uma coleção de funções para lidar com tais estruturas. Por essa razão, bem como por sua eficiência, é essencial a aplicações que envolvem a manipulação de imagens e vídeos, de modo que a biblioteca OpenCV está condicionada à sua instalação. É publicado sob a licença de código aberto BSD, que permite sua reutilização com poucas restrições. Neste trabalho, a versão 1.16.4 foi utilizada.

3.1.2 Scikit-learn

O scikit-learn é um módulo para Python que integra uma grande variedade de algoritmos de aprendizagem de máquina. Inclui soluções para classificação, regressão,

agrupamento e pré-processamento de dados. Assim como o NumPy, está condicionado à licença BSD. Neste trabalho, a versão 0.21.2 foi utilizada.

3.1.3 OpenCV

A OpenCV (Open Source Computer Vision Library) é uma biblioteca multiplataforma, totalmente livre ao uso acadêmico e comercial, destinada ao desenvolvimento de aplicações que englobam computação visual e aprendizagem de máquina. Sua grande coleção de algoritmos otimizados e o suporte prestado por uma comunidade crescente a tornam uma alternativa bastante atraente. Não à toa, é extensivamente utilizada por empresas, grupos de pesquisa e órgãos governamentais. Embora tenha sido desenvolvida nas linguagens C/C++, também oferece suporte à linguagem Python. Sua licença também é a BSD. A versão 4.1.0 foi utilizada no desenvolvimento deste trabalho.

3.2 HARDWARE

O desenvolvimento foi inteiramente realizado em um notebook *Dell Inspiron 14z 5423* com as seguintes especificações:

- Processador Intel® Core™ i3-3227U CPU 1,90 GHz;
- Memória RAM de 4,00 GB;
- Sistema operacional Windows 8 de 64 bits.

O *dataset* proposto, apresentado na seção 5.2.1.1, foi produzido com o auxílio de uma mesa digitalizadora *Wacom Intuos*, modelo CTL-480.

4 LITERATURA DA DETECÇÃO DE FOGO BASEADA EM VÍDEO E ASPECTOS RELACIONADOS

Na exposição deste capítulo, ficam estabelecidos os seguintes critérios de notação:

- 1) Canais de cores tratados de forma genérica são representados por letras maiúsculas (ex.: R, G, B, H, S, I) e não têm relação direta com os pixels de uma imagem.
- 2) Canais de cores que indicam os níveis de intensidade de uma imagem colorida são representados por letras minúsculas (ex.: r, g, b, h, s, i). Neste caso, fica implícito que cada componente de uma imagem $f(x, y)$ é uma função que retorna a intensidade da imagem, naquele canal, no ponto (x, y) , ou seja, $f(x, y) = [c_1(x, y), c_2(x, y), c_3(x, y)]$. Por clareza, a coordenada (x, y) é omitida quando conveniente (ex.: $r(x, y) \equiv r$).

A detecção de fogo baseada em vídeo ainda é uma área recente da visão computacional, tendo se desenvolvido com mais força somente a partir do início dos anos 2000. Esse desenvolvimento “tardio” pode ser atribuído principalmente às limitações impostas pela tecnologia e pelos custos de implementação, uma vez que, quando tratada como artifício de segurança, a metodologia proposta deve funcionar em tempo real, exigindo sempre o favorecimento da menor complexidade computacional em detrimento da acurácia. Em razão disso, a afirmação de Çetin et al. (2013) continua válida: ainda não é possível, com a tecnologia de hoje, contar com um sistema de detecção de fogo baseado em vídeo completamente confiável sem o auxílio de um operador humano. Em razão dessa insuficiência, a literatura tem elaborado sistemas com funcionamento de caráter complementar, isto é, que auxiliam métodos tradicionais de detecção de incêndio e combinam outras faixas do espectro luminoso – em geral, a expressão “detecção de fogo baseada em vídeo” sinaliza a detecção realizada no espectro de luz visível.

Embora a literatura sobre o tema seja vasta e apresente abordagens bastante diversas, é possível observar um padrão nas metodologias propostas (quanto à sua divisão). As detecções por cor e movimento estão presentes na grande maioria dos estudos e são realizadas, geralmente, como etapa preliminar. Uma vez obtido esse primeiro recorte, são extraídas informações das regiões de interesse - de acordo com os aspectos estudados por cada autor -,

resumidas aos chamados *descritores*. A etapa de decisão, na qual por fim se apresenta o resultado da detecção, pode ser baseada em regras ou contar com técnicas de aprendizagem de máquinas. As seções a seguir detalham cada uma dessas etapas e servem de base para o entendimento do método proposto por este trabalho.

4.1 SEGMENTAÇÃO POR COR

Das características visuais do fogo, a mais descritiva é certamente sua cor. No entanto, não se pode supor, a partir dessa afirmação, que as chamas de um incêndio exibem sempre as mesmas tonalidades - sabe-se que a coloração das chamas é condicionada a fatores como temperatura, circulação de ar, composição do material em combustão, presença de fumaça, iluminação ambiente, etc. (ver seção 2.3). Outras considerações importantes podem ser feitas não sobre o incêndio, propriamente, mas sobre o processo de captura das imagens, que depende dos dispositivos e sofre com suas limitações - em determinadas condições, por exemplo, as regiões de fogo apresentam-se fortemente saturadas, causando perda de informação. Esse discernimento, quanto à natureza do incêndio e qualidade das amostras, é crucial, pois interfere diretamente no desempenho da segmentação das regiões de interesse. Numa abordagem ampla, é aconselhável que todos esses cenários sejam avaliados; em contrapartida, aplicações mais restritas podem se beneficiar do estreitamento da gama de possibilidades. A literatura relacionada, em sua grande maioria, simplifica o problema ao limitar-se à análise das chamas vermelho-amareladas, amparando-se no fato de que essa coloração está presente em quase todos os incidentes envolvendo fogo. No modelo de cores RGB, a referida banda é descrita segundo a relação $R > G > B$ (ÇETIN et al., 2013; CHEN; KAO; CHANG, 2003).

Por se tratar de uma característica facilmente explorada, no contexto da detecção de fogo em imagens, a cor é utilizada quase sempre no primeiro estágio de segmentação dos métodos propostos na literatura. Apesar dessa constância, há divergência com relação às técnicas e modelos de cores utilizados, visto que cada conjunto de elementos apresenta peculiaridades que os tornam adequados a determinadas aplicações e inadequados a outras. Quando a operação em tempo real de um sistema é desejada, por exemplo, prioriza-se o modelo RGB, tendo em vista a menor complexidade computacional (CHEN; WU; CHIOU, 2004). O favorecimento do modelo RGB pode ser explicado, em outras palavras, por sua popularidade e conveniência, já que quase todas as câmeras de vídeo de espectro visível capturam imagens nesse modelo e há um conteúdo espectral óbvio associado a ele (ÇETIN et al. 2013). Outros

domínios foram também explorados repetidamente, como o HSI, o YCbCr, o YUV, e o CIELAB.

Embora a escolha do modelo de cores resulte em diferenças práticas, seja por sua capacidade de diferenciar um dado conjunto de pixels ou por sua representação matemática, este se resume apenas ao domínio sobre o qual o método de segmentação se baseia, sendo mais determinante, nesse caso, a escolha das técnicas de discriminação entre as classes (fogo e não-fogo). Dentre essas técnicas, é comum a segmentação baseada em regras, fazendo-se uso de simples limiarização, bem como a utilização de técnicas de agrupamento de dados e *look-up tables*. Métodos mais elaborados empregam aprendizagem de máquina. Em todo caso, é necessária uma análise estatística prévia dos pixels de interesse, de forma que o resultado da segmentação está sempre sujeito ao conjunto de dados amostrados.

A seguir são detalhadas algumas dessas técnicas de segmentação por cor, cuja seleção foi realizada prezando pela relevância (com base no número de citações) e aplicabilidade no presente trabalho. Todas elas foram implementadas e verificadas, conforme é descrito mais adiante (ver capítulo 5.2).

Tabela 1 - Modelos de cores na literatura da segmentação de fogo em imagens

	RGB	YCbCr	HSI	YUV
Phillips, Shah e Lobo (2002)	✓			
Chen, Kao e Chang (2003)	✓		✓	
Hornng, Peng e Chen (2005)			✓	
Çelik e Demirel (2009)		✓		
Rossi e Akhloufi (2009)	✓			✓
Wirth e Zaremba (2010)		✓	✓	
Rudz et al. (2013)	✓	✓		

4.1.1 Segmentação baseada em regras

Ainda que haja certa obviedade no espectro visível do fogo, é essencial, no contexto da segmentação de imagens, que este seja descrito com a maior exatidão possível. Caso contrário, a extração das regiões de interesse pode ser excessivamente abrangente ou restritiva. Essa é a razão pela qual o processo de segmentação com base na cor se inicia, preferencialmente, por um estudo estatístico dos seus níveis de intensidade - em posse dos dados de frequência de ocorrência, é possível reduzir o fogo à sua região cromática mais

provável. Com esse objetivo, Horng, Peng e Chen (2005) analisam 70 imagens de chamas no modelo HSI (normalizado a $0 \leq H \leq 360^\circ$, $0 \leq S \leq 100$ e $0 \leq I \leq 255$) e determinam que a matiz do fogo na faixa do vermelho ao amarelo está delimitada, geralmente, por 0 e 60° . Os valores de saturação e brilho, no entanto, são mais difíceis de precisar, segundo os autores, sendo melhor representados por dois cenários de iluminação distinta: para ambientes mais claros, a saturação e o brilho ficam, respectivamente, entre $[40, 100]$ e $[127, 255]$. Já para ambientes mais escuros, as faixas típicas são $[20, 100]$ e $[100, 255]$. Propondo que os pixels de uma imagem $f(x, y) = (h, s, i)$ sejam classificados como fogo somente se seus valores estiverem presentes num conjunto C_f delimitado por $[h_1, h_2]$, $[s_1, s_2]$ e $[i_1, i_2]$, cujos limites são deduzidos a partir do estudo estatístico citado anteriormente, a imagem $g(x, y)$ resultante da segmentação é obtida da seguinte forma:

$$C_f([h_1, h_2], [s_1, s_2], [i_1, i_2]) = \{(h, s, i) \mid h_1 \leq h \leq h_2, s_1 \leq s \leq s_2, i_1 \leq i \leq i_2\} \quad (9)$$

$$g(x, y) = \begin{cases} 0, & \text{se } f(x, y) \notin C_f \\ f(x, y), & \text{caso contrário} \end{cases} \quad (10)$$

Chen, Kao e Chang (2003) também adotam o modelo HSI no desenvolvimento do método de segmentação e apresentam resultados quase idênticos aos de Horng, Peng e Chen (2005). No entanto, a fim de evitar o custo computacional da conversão entre modelos de cores, formulam regras que exploram o modelo RGB associado apenas ao canal S do modelo HSI. A observação feita anteriormente sobre a matiz dos pixels de fogo ($0 \leq H \leq 60^\circ$) é traduzida pelas condições $R \geq G$ e $G > B$. “Além do mais, deve haver um valor de R mais forte nas imagens de fogo [...]. Portanto, a componente R deve estar acima de um limiar, R_T .” (CHEN; KAO; CHANG, 2003, p. 107, tradução nossa). Por último, é destacada a influência da iluminação ambiente na saturação das imagens contendo fogo, razão pela qual se estabelece uma relação entre os canais S e R em concordância com a observação de que a saturação das imagens de fogo diminui conforme a intensidade do vermelho aumenta. A classificação dos pixels de fogo fica, portanto, condicionada às seguintes relações:

$$r > R_T \quad (11)$$

$$r \geq g > b \quad (12)$$

$$s \geq (255 - r) \frac{S_T}{R_T} \quad (13)$$

A constante S_T é definida como a saturação no ponto da reta $s(r)$ no qual $r = R_T$. Aqui, vale ressaltar que a equação da reta apresentada pelos autores não corresponde exatamente à observada, e que a relação (13) é melhor expressa por:

$$s \geq (255 - r) \frac{S_T}{255 - R_T} \quad (14)$$

Embora o método descrito por Chen, Kao e Chang (2003) produza bons resultados, o modelo RGB apresenta supostamente uma desvantagem ao não distinguir componentes de brilho e cromaticidade. Com o intuito de reduzir a influência da iluminação no reconhecimento da cor, Çelik e Demirel (2009) propõem um método baseado no modelo de cores YCbCr, no qual o brilho e a cromaticidade são representados independentemente. Desse modo, sejam \bar{y} , \bar{c}_b e \bar{c}_r os valores médios de cada canal de uma imagem, definidos por

$$\bar{m} = \frac{1}{N} \sum_{i=1}^N m(w_i, z_i) \mid m \in \{y, c_b, c_r\} \quad (15)$$

onde N é o número total de pixels da imagem e (w_i, z_i) a coordenada espacial de cada pixel; e sejam τ uma constante definida experimentalmente e f_u , f_d e f_l funções no plano $Cr - Cb$ que delimitam a região estatisticamente formada por pixels de “fogo”, os autores consideram haver fogo em um pixel $p = (y, c_b, c_r)$ se

$$y > c_b \quad (16)$$

$$c_r > c_b \quad (17)$$

$$y > \bar{y} \quad (18)$$

$$|c_r - c_b| \geq \tau \quad (19)$$

$$c_b \geq f_u(c_r) \cap c_b \leq f_d(c_r) \cap c_b \leq f_l(c_r) \quad (20)$$

A condição imposta em (19) traduz o fato de que as regiões de interesse apresentam, tipicamente, uma grande diferença entre os canais C_r e C_b . Por meio de uma curva de Característica de Operação do Receptor (COR), é deduzido que $\tau = 40$ produz a melhor relação entre as taxas de verdadeiros positivos e falsos positivos. Já as inequações apresentadas em (20) surgem da observação de que os pixels extraídos das regiões contendo fogo ocorrem com maior frequência dentro de uma região no plano $C_r - C_b$ confinada pelas curvas f_u , f_d e f_l .

4.1.2 Segmentação por agrupamento de dados

Métodos de agrupamento também são explorados pela literatura na segmentação de fogo em imagens, como nos trabalhos de Rossi e Akhloufi (2009) e Rudz et al. (2013). Estes comentam:

Em uma aplicação como essa, a segmentação de imagens pode ser considerada um problema de agrupamento focado na combinação de pixels apresentando as mesmas características de cor. Para imagens de fogo, o objetivo é combinar todos os pixels pertencentes a regiões de fogo dentro de um grupo. (RUDZ et al., 2013, p. 3, tradução nossa).

Em ambos os trabalhos, aplica-se o método de agrupamento *k-means* sobre uma das componentes da imagem – aquela considerada mais promissora na distinção entre as classes – e, posteriormente, submete-se o resultado a um processo de filtragem, cujo objetivo é eliminar as regiões erroneamente classificadas. As metodologias diferem apenas na escolha dos atributos da etapa de agrupamento e nos critérios de filtragem. Rossi e Akhloufi (2009) apontam que o canal V do modelo de cores YUV é um candidato interessante na discriminação de fogo em imagens. O estudo conduzido por Rudz et al. (2013), no entanto, mostra que a componente C_b do modelo YCbCr pode gerar melhores resultados. Neste último, o agrupamento é realizado sobre quatro classes ($k = 4$); no primeiro, sobre apenas duas ($k = 2$). Rossi e Akhloufi (2009) não detalham como foi realizado o processo de decisão. Por outro lado, Rudz et al. (2013) explicam que a avaliação dos métodos e a escolha do número de grupos foi feita com base nos critérios de Martin et al. (2001) e Hafiane et al. (2007).

A etapa de agrupamento proporciona uma boa extração preliminar das regiões de interesse, porém, uma vez que está restrita a um único canal da imagem, não é capaz de eliminar uma grande parcela de falsos positivos. Isso leva à necessidade de verificação das regiões extraídas, considerando sua proximidade com valores de referência. Rossi e Akhloufi (2009) realizam essa verificação em RGB utilizando um modelo Gaussiano 3D que representa os pixels dentro da região segmentada, conforme segue: um dado pixel $p = (r, g, b)$ somente é considerado fogo se satisfizer a condição

$$\|p - \bar{m}\| \leq k \times \sigma \quad (21)$$

onde k é uma constante, $\bar{m} = (\mu_r, \mu_g, \mu_b)$ e $\sigma = \text{máx}(\sigma_r, \sigma_g, \sigma_b)$, tal que μ_i e σ_i são a média e o desvio padrão do canal $i \in \{r, g, b\}$, respectivamente. O lado esquerdo da inequação, $\|p - \bar{m}\|$, é definido como a distância euclidiana entre p e \bar{m} :

$$\|p - \bar{m}\| = \sqrt{(r - \mu_r)^2 + (g - \mu_g)^2 + (b - \mu_b)^2} \quad (22)$$

Rudz et al. (2013) efetuam a verificação por regiões conexas – e não pixel a pixel –, estabelecendo regras diferentes conforme seu tamanho; uma região grande (com mais de 256 pixels), é classificada como fogo se as seguintes condições forem satisfeitas:

$$\|hist_R^{ref} - hist_R\| < \tau_R \quad (23)$$

$$\|hist_G^{ref} - hist_G\| < \tau_G \quad (24)$$

$$\|hist_B^{ref} - hist_B\| < \tau_B \quad (25)$$

Em contrapartida, uma região pequena (com menos de 256 pixels), deve respeitar os seguintes critérios:

$$\|\mu_R^{ref} - \mu_R\| < k_R \times \sigma_R^{ref} \quad (26)$$

$$\|\mu_G^{ref} - \mu_G\| < k_G \times \sigma_G^{ref} \quad (27)$$

$$\|\mu_B^{ref} - \mu_B\| < k_B \times \sigma_B^{ref} \quad (28)$$

Nessas relações, $hist^{ref}$, μ^{ref} e σ^{ref} correspondem ao histograma, à média e ao desvio padrão de referência, respectivamente, e foram calibrados sobre um terço do banco de dados; τ_R , τ_G , τ_B , k_R , k_G e k_B são constantes definidas experimentalmente, otimizadas sobre a mesma parcela do banco de dados.

4.1.3 Segmentação via histograma

Tendo em vista que o histograma de uma imagem armazena a informação de seus níveis de cor - e que um histograma de referência pode ser construído a partir de um conjunto de imagens de interesse -, é evidente que a segmentação baseada em cor possa ser realizada por meio de uma comparação entre a imagem analisada e o histograma de referência. Essa ideia é explorada por Swain e Ballard (1990) em *Indexing via color histograms*, artigo no qual apresentam uma abordagem para o reconhecimento e a localização de objetos em imagens, com base na correspondência entre histogramas, e introduzem o conceito de “retroprojeção de histograma” (*histogram backprojection*). Na retroprojeção de histograma, a imagem modelo é representada por seu histograma multidimensional M . O histograma da imagem-alvo, I (com a mesma dimensão de M), bem como um terceiro histograma, R , que é a razão entre M e I , são computados. O histograma R é então retroprojetado na imagem-alvo, conferindo a cada pixel um valor proporcional à sua correspondência com o histograma modelo. Em outras palavras, a retroprojeção consiste na substituição dos valores da imagem pelos valores por eles indexados em R . Por fim, é realizada uma convolução entre o resultado e uma máscara, cujo pico deve apresentar a localização mais provável do objeto investigado (SWAIN; BALLARD, 1990, p. 393). Mais precisamente, seja $h(c)$ uma função que mapeia uma cor “ c ” a um *bin* de um histograma, seja D^r um disco de raio r definido por

$$D_{x,y}^r = \begin{cases} 1, & \text{se } \sqrt{(x^2 + y^2)} < r \\ 0, & \text{caso contrário} \end{cases} \quad (29)$$

e seja “*loc*” uma função que retorna um pixel (x, y) com o valor de seu argumento, a retroprojeção de histograma pode ser assim descrita (SWAIN; BALLARD, 1990, p. 393):

- 1) Para cada *bin* j do histograma, faça $R_j = M_j/I_j$
- 2) Para cada x, y , faça $b_{x,y} = \min(R_{h(c_{x,y})}, 1)$
- 3) $b = D^r * b$
- 4) $(x_t, y_t) = \text{loc}(\max_{x,y}(b_{x,y}))$

É importante destacar que os passos 3 e 4, responsáveis por determinar a posição mais provável do objeto investigado, não são relevantes a este trabalho. Como a segmentação proposta nesta seção visa à extração do maior número possível de pixels candidatos a “fogo”, são considerados apenas os passos 1 e 2, associados a uma etapa de limiarização, conforme é aprofundado mais adiante.

A aplicabilidade da referida técnica na segmentação de fogo em imagens é avaliada por Wirth e Zaremba (2010). Dada a suposição de que a informação contida num histograma 3D completo de uma imagem no modelo RGB (contendo 256^3 bins) é dispersa demais para ser utilizada efetivamente, os autores concluem que os modelos HSV e YCbCr são opções mais adequadas, já que possibilitam que a crominância seja expressa em histogramas de apenas duas componentes, totalizando 256^2 bins no pior caso. Propõe-se, ainda, que cada componente seja quantizada a 32 níveis de intensidade. Logo, embora a retroprojeção de histograma possa ser empregada em um histograma 3D, o que permitiria a utilização de todos os canais dos modelos analisados, os autores optam por trabalhar com histogramas de apenas duas dimensões, reduzidos a 32^2 níveis de intensidade, com o intuito de diminuir a complexidade computacional. Segundo a avaliação da segmentação conduzida em ambos os modelos de cor, determina-se que o modelo HSV não é tão adequado quanto o YCbCr na detecção de chamas (WIRTH; ZAREMBA, 2010). Pressupõe-se que essa afirmação seja válida somente em referência à técnica utilizada.

Uma aplicação mais direta de histogramas na segmentação de fogo em imagens pode ser vista em Phillips, Shah e Lobo (2000). Aqui, fica claro que um histograma 3D de 256^3 bins

no modelo RGB pode, sim, ser eficiente na segmentação de imagens, desde que o histograma de referência seja construído adequadamente. Ao abordar, indiretamente, a questão da dispersão dos dados apontada por Wirth e Zaremba (2010), o trabalho referenciado por Phillips, Shah e Lobo (2000) apresenta uma solução:

Se fosse possível obter dados de treinamento perfeitos, um simples histograma e a limiarização dos pixels na região de treinamento criariam um bom predicado de cor (PC). Infelizmente, os dados de treinamento raramente são perfeitos, e nossos primeiros experimentos utilizando somente exemplos positivos produziram resultados muito ruins, principalmente devido a regiões de falsos positivos no PC. As tentativas de limpá-los com um limiar maior resultaram na criação de regiões de falsos negativos. Adicionar exemplos negativos de treinamento e suavização gaussiana [...] melhorou drasticamente os resultados (KJELDEN; KENDER, 1996, p. 314, tradução nossa).

Em outras palavras, a dispersão ocorre apenas em razão da limitação dos dados de treinamento. A fim de atenuar esse problema, Kjeldsen e Kender (1996) propõem que o histograma de referência seja gerado mediante a soma de distribuições gaussianas e a partir de um conjunto de imagens que contenha também exemplos negativos. Mais especificamente, conforme detalham Phillips, Shah e Lobo (2000), um histograma modelo com suavização gaussiana pode ser construído por intermédio de um par de imagens de treinamento – uma imagem colorida e uma máscara booleana especificando as localizações dos objetos de interesse. Para cada valor “verdadeiro” da máscara, adiciona-se ao histograma uma distribuição gaussiana centrada no valor da imagem colorida naquele ponto; da mesma maneira, subtrai-se do histograma uma distribuição menor para cada ponto em que o valor da máscara for “falso”. No referido trabalho, atribui-se $\sigma = 2$ aos exemplos positivos e $\sigma = 1$ aos exemplos negativos, sendo σ o desvio padrão das distribuições gaussianas. Por fim, o histograma é limiarizado no nível desejado, resultando em uma função que retorna um valor booleano indicando se um determinado pixel $p = (r, g, b)$ está ou não presente na tabela de cores de referência. Segundo os autores, dez imagens são suficientes para a construção do histograma da forma indicada.

4.1.4 Outras abordagens

Liu e Ahuja (2004) realizam a detecção da potencial região de fogo, a princípio, somente com base na cor e sua distribuição.

[...] a região de fogo em uma única imagem pode ser modelada como segue: (i) ela está em alto contraste com seus arredores; (ii) ela exibe uma estrutura de anéis de cores aninhados, variando de branco em seu núcleo até amarelo, laranja e vermelho na periferia.” (LIU; AHUJA, 2004, p. 135, tradução nossa)

O método detecta, primeiramente, regiões de alto brilho (em escala de cinza), que podem corresponder ao núcleo de um foco de incêndio. Essas regiões são expandidas, seguindo o gradiente espectral da imagem, conforme se agregam os pixels da vizinhança que possuem cores dadas pelo modelo de cores de fogo. Este modelo segue uma função densidade de probabilidade no espaço HSV. Para garantir que a segmentação realmente apresenta fogo e não somente uma região de alto brilho, os autores estabelecem que essa região deva possuir pelo menos metade dos pixels com cor de fogo.

4.2 SEGMENTAÇÃO E ANÁLISE DO MOVIMENTO

Assim como na suposição de um predicado de cor, a detecção de fogo baseada em vídeo se escora no pressuposto de que as chamas de um incêndio estão sempre em movimento. Essa premissa só se torna inválida em situações de fogo controlado - como numa vela em um ambiente sem fluxo de ar - e que, portanto, não devem mesmo ser tratadas como acidentes envolvendo fogo. A partir dessa ideia, em conjunto com a segmentação por cor discutida no capítulo anterior, é possível elaborar um sistema primitivo de detecção de fogo, como foi de fato realizado em alguns dos primeiros trabalhos apresentados na literatura. A deficiência óbvia dessa abordagem está no fato de que muitos objetos e fenômenos podem ser confundidos com fogo quando apenas esses critérios são considerados. Logo, a rigor, um método não pode ser considerado robusto sem o auxílio de informações adicionais. Embora a simples segmentação de movimento não forneça elementos suficientes, é válida e comum a utilização desse artifício na eliminação de regiões “cor-de-fogo” estáticas e na extração de regiões que posteriormente são submetidas a uma análise mais precisa do movimento, com o intuito de explorar atributos como a cintilação, a frequência de oscilação e o fluxo óptico.

As seções a seguir ilustram um apanhado de técnicas de segmentação e análise do movimento usualmente empregadas na detecção de fogo em vídeo, algumas das quais são utilizadas mais adiante na metodologia proposta por este trabalho.

4.2.1 Subtração de fundo

O termo “subtração de fundo” compreende toda técnica de extração do primeiro plano de um vídeo realizada por meio da diferenciação entre o quadro atual e um fundo de referência, supondo que a câmera seja estacionária, de forma que se considera parte do primeiro plano todo pixel cuja intensidade varia acima de um limiar especificado, conforme a relação

$$|I(x, y, t) - B(x, y, t)| > \tau(x, y, t) \quad (30)$$

sendo I o quadro atual, B o fundo estimado e τ o limiar de variação de intensidade. Em geral, pressupõe-se uma câmera estacionária.

Os métodos de subtração de fundo são classificados, basicamente, pela forma como modelam o fundo de referência, mas suas divergências vão além dessa característica. Uma delas é quanto ao número de canais considerados na operação. No contexto deste trabalho, conforme apontam Bertolini et al. (2014), essa distinção é importante porque a segmentação do movimento das cenas pode ser bastante comprometida se a subtração de fundo for realizada em escala de cinza, ignorando as variações específicas de cada canal de cor.

Embora haja uma preocupação recorrente com a eficiência dos métodos de segmentação de movimento na detecção de fogo em vídeo, não parece haver uma tendência clara pautando as escolhas apresentadas na literatura. Celik et al. (2007) baseiam-se no trabalho de Wren et al. (1997) e utilizam um modelo de fundo adaptativo gerado por três distribuições Gaussianas que representam os dados estatísticos dos pixels em cada canal de cor no sistema RGB. Töreyn et al. (2006), Xu e Xu (2007), Günay et al. (2009) e Ko, Cheong e Nam (2010) baseiam-se em Collins, Lipton e Kanade (1999), os quais formulam um fundo adaptativo com atualização seletiva, cuja modelagem é descrita da seguinte forma:

$$B_{n+1}(x, y) = \begin{cases} aB_n(x, y) + (1 - a)I_n(x, y), & M = 0 \\ B_n(x, y), & M = 1, \end{cases} \quad (31)$$

sendo M uma variável que indica se o pixel atual está se movendo ($M = 1$) ou se está estático ($M = 0$), $a \in [0, 1]$ uma constante que define o peso de cada termo na média ponderada e especifica o quão rápido novas observações sobrescrevem as antigas, $B_n(x, y)$ um pixel do fundo e $I_n(x, y)$ um pixel do quadro atual (indicado por n). Com a exceção de Xu e Xu (2007),

que não explicita essa informação, os demais adotam um limiar adaptativo, conforme sugerido por Collins, Lipton e Kanade (1999):

$$|I_n(x, y) - B_n(x, y)| > T_n(x, y) \quad (32)$$

$$T_{n+1}(x, y) = \begin{cases} aT_n(x, y) + (1 - a)(5 \times |I_n(x, y) - B_n(x, y)|), & M = 0 \\ T_n(x, y), & M = 1, \end{cases} \quad (33)$$

Dimitropoulos, Tsalakanidou e Grammalidis (2012) dão mais detalhes sobre o processo que os levou à escolha do método de subtração de fundo admitido em seu trabalho: considerando a velocidade e a acurácia, treze métodos foram avaliados com o auxílio de um *ground-truth* e tiveram suas curvas COR (Característica de Operação do Receptor) traçadas – dentre os métodos estão, inclusive, os propostos por Collins, Lipton e Kanade (1999) e Wren et al. (1997). Em função dos resultados obtidos, um método de modelagem adaptativa de fundo pela mediana, proposto por Mcfarlane e Schofield (1995), foi escolhido.

4.2.2 Cintilação

Informações valiosas podem ser extraídas da variação temporal de intensidade observada nos pixels de um vídeo. As chamas de um incêndio, mais especificamente, possuem uma característica importante: a cintilação.

A maior variação temporal possível ocorre no caso de cintilação, ou seja, quando um pixel está mudando rapidamente de um valor de intensidade para outro. Isso geralmente ocorre somente na presença de fogo. Movimento de corpos rígidos, em contraste, produz menor variação temporal. (PHILLIPS III; SHAH; LOBO, 2000, p. 226, tradução nossa).

Phillips, Shah e Lobo (2000), tirando proveito do fato que evidenciam, realizam a segmentação do movimento com base na média de variação absoluta de intensidade dos pixels. A premissa é simples: se a cintilação, presente nas regiões contendo fogo, indica um grau de variação de intensidade maior do que o observado em movimentos de outras naturezas, então a segmentação das chamas pode ser realizada pela simples limiarização da média de variação

absoluta de intensidade para cada pixel. Mais precisamente, a operação é traduzida pelas seguintes equações:

$$Dif(x, y) = \frac{\sum_{i=2}^n |I_i(x, y) - I_{i-1}(x, y)|}{n - 1} \quad (34)$$

$$Dif_{\sim fogo} = \frac{\sum_{(x,y) \notin [cor\ de\ fogo]} Dif(x, y)}{\sum_{(x,y) \notin [cor\ de\ fogo]} 1} \quad (35)$$

$$\Delta I(x, y) = Dif(x, y) - Dif_{\sim fogo} \quad (36)$$

sendo $Dif(x, y)$ a média da diferença absoluta de intensidade para cada pixel $I(x, y)$ tomada em uma sequência de quadros no intervalo que varia entre i e n , $Dif_{\sim fogo}$ a média de $Dif(x, y)$ sobre todos os pixels que não possuem cor de fogo e $\Delta I(x, y)$ a variação média absoluta de intensidade “corrigida” para cada pixel da imagem. É importante ressaltar que o cálculo dessas grandezas exige a segmentação prévia dos pixels cor-de-fogo. O termo $Dif_{\sim fogo}$ pode ser entendido como a variação média absoluta de intensidade global, ou seja, indica o grau de variação das regiões da cena sem a presença de fogo. Ao contabilizar esse termo, a função $\Delta I(x, y)$ é dita “corrigida” por eliminar a influência da variação média global de intensidade, salientando os pixels que possuem variação de intensidade acima da média. Por fim, considera-se candidato a “fogo”, pelo critério da cintilação, um *pixel* que oscila a uma taxa suficientemente superior à média, adotando-se um limiar k :

$$fogo_{cintilação}(x, y) = \begin{cases} 1, & \Delta I(x, y) > k \\ 0, & \text{caso contrário} \end{cases} \quad (37)$$

4.2.3 Histórico do movimento

Os conceitos apresentados nesta seção restringem-se aos trabalhos de caracterização do movimento humano publicados por Bobick e Davis (2001) e Bradski e Davis (2002), nos quais são apresentadas propostas de reconhecimento de gestos baseadas no histórico e no gradiente do movimento. Embora não estejam diretamente relacionados à detecção de fogo em vídeo, acredita-se que o resultado derivado desses estudos possa ser útil na tarefa de diferenciar o movimento observado nas chamas dos apresentados por outros objetos.

4.2.3.1 Imagem da Energia do Movimento (MEI)

A variação existente entre quadros subsequentes de um vídeo é obtida por sua diferenciação (subtração), e o resultado limiarizado dessa diferença é uma imagem binária que indica a região onde houve movimento acima do referido limiar. Se forem tomadas e acumuladas as diferenças de uma sequência de quadros, o que se obtém é uma imagem que carrega a memória da localização do movimento em todo o intervalo. Essa imagem binária de movimento cumulativo, com “camadas” sobrepostas, é chamada de Imagem da Energia do Movimento (*Motion-Energy Image*) (BOBICK; DAVIS, 2001), representada por $E_r(x, y, t)$ e definida por

$$E_r(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t - 1), \quad (38)$$

sendo $D(x, y, t)$ uma sequência de imagens binárias indicando as regiões de movimento e τ a duração do intervalo. Visualmente, a Imagem da Energia do Movimento apresenta a silhueta do movimento (Figura 7).

4.2.3.2 Imagem do Histórico do Movimento (MHI)

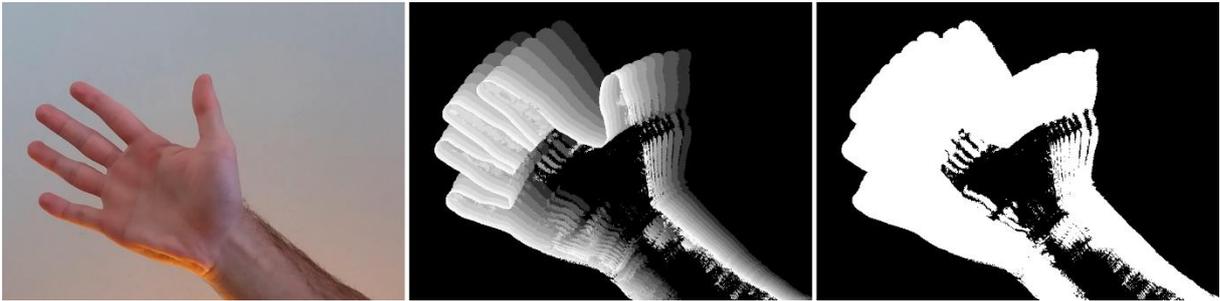
A fim de registrar como – e não somente onde – o movimento ocorre, Bobick e Davis (2001) formulam a Imagem do Histórico do Movimento (*Motion-History Image*). “Em uma MHI H_τ , a intensidade de um pixel é uma função do histórico temporal de movimento naquele ponto.” (BOBICK; DAVIS, 2001, p. 260, tradução nossa).

$$H_\tau(x, y, t) = \begin{cases} \tau, & \text{se } D(x, y, t) = 1 \\ \max(0, H_\tau(x, y, t - 1) - 1), & \text{caso contrário} \end{cases} \quad (39)$$

Além da silhueta do objeto de interesse, H_τ fornece a informação de quando o movimento ocorre: regiões de maior brilho indicam movimento recente, enquanto regiões obscurecidas indicam movimento mais antigo.

Vale ressaltar que a Imagem da Energia do Movimento também pode ser obtida por meio da limiarização da Imagem do Histórico do Movimento.

Figura 7 - Gesto acompanhado das imagens do histórico e da energia do movimento



Fonte: autoria própria. O deslocamento, da direita para a esquerda, é registrado pela MHI (centro) – as posições mais atuais são mais claras na imagem. A MEI (direita), que reproduz a segmentação do movimento, é equivalente à limiarização da MHI.

4.2.4 Gradiente do histórico do movimento

Em uma nova proposta de segmentação e reconhecimento de gestos, Bradski e Davis (2002) abordam o cálculo do gradiente de movimento em uma sequência de imagens por meio da MHI. Para tanto, formulam uma nova imagem de ponto flutuante, designada *timed Motion History Image* (tMHI) e calculada como segue:

$$tMHI_{\delta}(x, y, t) = \begin{cases} \tau, & \text{se houver movimento em } (x, y) \\ 0, & \text{se } tMHI_{\delta}(x, y, t') < (\delta - \tau) \end{cases} \quad (40)$$

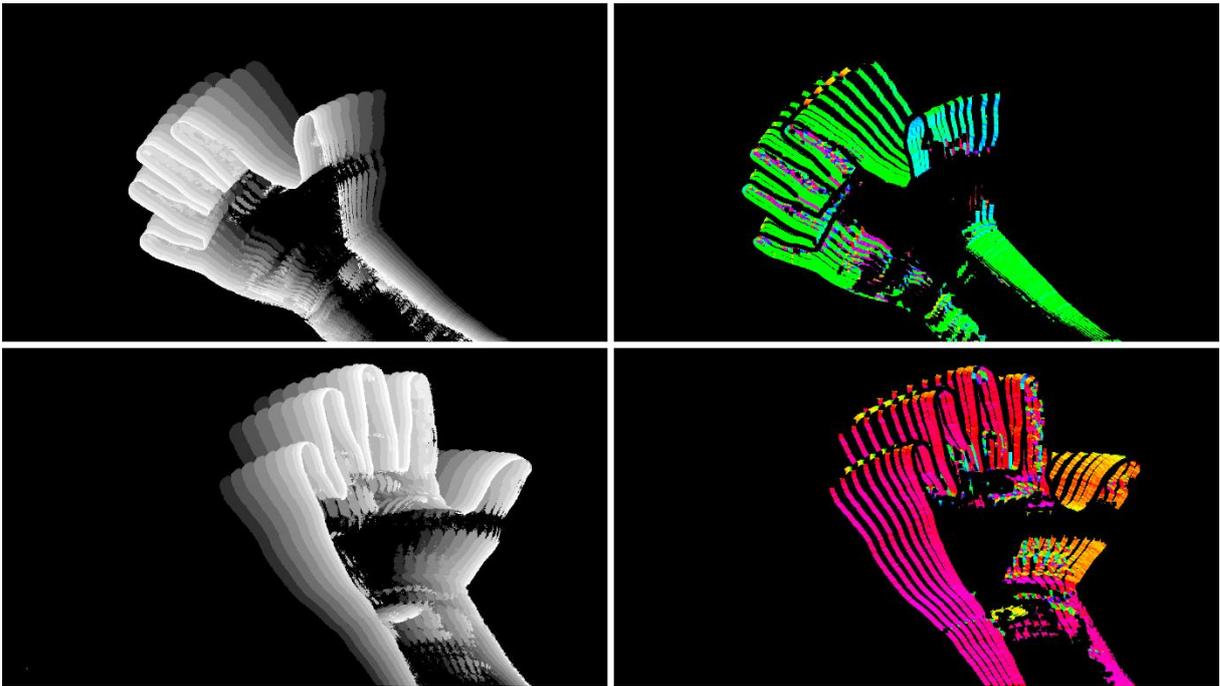
Em $tMHI_{\delta}(x, y, t)$, δ é a constante de duração máxima, associada à natureza do movimento – no caso considerado, é da ordem de alguns segundos. Essa variação da equação (39) garante que a representação de um determinado gesto seja independente da taxa de captura quando o tempo de gravação for o mesmo (BRADSKI; DAVIS, 2002).

No exemplo a seguir, observa-se claramente que a MHI registra a trilha desenvolvida pelo gesto, de maneira que o gradiente de intensidade acompanha a direção do movimento. Essa característica da tMHI fornece uma alternativa rápida para o cálculo do fluxo óptico normal (ortogonal ao contorno do objeto de interesse). A orientação $\varphi(x, y)$ do gradiente em cada ponto da tMHI é dada por

$$\varphi(x, y) = \arctg \frac{F_y(x, y)}{F_x(x, y)} \quad (41)$$

sendo F_x e F_y as derivadas espaciais calculadas pela convolução da imagem com filtros Sobel nas direções x e y .

Figura 8 - Gradiente do movimento



Fonte: autoria própria. À esquerda, a MHI; à direita, o gradiente do movimento. Originalmente, a orientação em cada ponto é representada por valores entre 0 e 360, compreendendo um círculo completo. Nesta imagem, no entanto, reduziu-se o domínio pela metade, a fim de possibilitar a representação em 8 bits. O mapeamento de cores foi realizado de acordo com a matiz do modelo HSI.

4.3 DESCRITORES

Até aqui, a cor foi tratada como um meio de identificar quais pixels poderiam fazer parte de uma região contendo fogo. Esse primeiro recorte, no entanto, carece de informações mais detalhadas sobre a aparência dessas regiões, como a *textura* e a distribuição de cores. O movimento, de forma similar, foi abordado de maneira restrita: ora limitando-se à sua ocorrência, sem a preocupação de descrevê-lo, ora analisando-o apenas no domínio do tempo. Os parágrafos seguintes resumem uma série de descritores complementares explorados pela literatura da detecção de fogo em vídeo, abordando tanto a cor quanto o movimento de maneiras mais amplas e introduzindo novas formas de representar o objeto de estudo.

Descritor é um termo que pode assumir conotações diversas. No contexto da computação visual, de forma geral, indica um atributo do objeto de interesse com o potencial de descrevê-lo visualmente. Mais especificamente, pode designar um atributo que serve de elemento a um método de classificação por aprendizagem de máquina. Neste trabalho, o termo é usado indistintamente.

Conforme mencionado na seção 2.3, sabe-se que a frequência de cintilação das chamas é de aproximadamente 10 Hz, e que esse número não é significativamente afetado pelo material em combustão. É natural, portanto, supor que tal frequência de cintilação seja um bom indicativo da presença de fogo, mas essa suposição não se sustenta em todos os cenários. No estágio inicial de um incêndio, por exemplo, observa-se um comportamento caótico e intermitente, sem uma frequência definida. Nesse caso, um incêndio turbulento pode ser caracterizado como uma atividade caótica de banda larga. De todo modo, a análise de Fourier pode ser utilizada na detecção de fogo em vídeo, desde que não se busque, necessariamente, a frequência de 10 Hz (ÇETIN et al., 2013). Ainda no contexto da cintilação, uma alternativa possível é a análise de *wavelet* temporal das chamas, como na implementação de Töreyn et al. (2006), na qual se verificou a variação temporal dos pixels no canal R do modelo RGB.

No domínio espacial, de forma análoga, observa-se alta variação de cores nas regiões contendo fogo, o que torna possível, em certa medida, distingui-las dos objetos ordinários cor-de-fogo. Töreyn et al. (2006), dando sequência à análise temporal citada no fim do parágrafo anterior, exploram também o perfil espacial via *wavelet*. Essa abordagem, no entanto, não é a única possível, conforme apontam Borges e Izquierdo (2010):

Ao contrário de outras regiões de alarme falso, como um sinal de trânsito amarelo, por exemplo, as regiões de fogo têm uma quantidade significativa de variabilidade nos valores dos pixels. Bancos de filtros são frequentemente usados na análise de textura ao se tentar descrever um determinado padrão. No caso do fogo, no entanto, é muito difícil descrever sua textura com qualquer modelo dado. A aleatoriedade observada no fogo pode variar significativamente na resposta em frequência (periodicidade muitas vezes não está presente) e no ângulo dos gradientes, por exemplo. A variância é uma métrica bem conhecida para indicar a medida da rugosidade nos valores dos pixels. (BORGES; IZQUIERDO, 2010, p. 725, tradução nossa).

Os autores argumentam, portanto, que a variância dos pixels de uma dada região da imagem, uma simples métrica estatística, serve como um descritor que informa o grau de *rugosidade* dessa região. Desse modo, presume-se que haja fogo somente quando a variância dos pixels da região avaliada estiver acima de um limiar predeterminado. Na seção referente ao estudo estatístico das regiões de fogo (seção 5.2.2), verifica-se que a variância - lá representada pelo desvio padrão - de G é significativamente maior que as de B e R, de modo que se pode adotar o canal verde como o principal representante da variabilidade dos pixels de fogo.

Borges e Izquierdo (2010) sugerem ainda a *obliquidade*, medida do grau de assimetria de uma distribuição em torno de sua média, como um potencial descritor das regiões de fogo

em imagens, sustentados pela observação do histograma (ver Figura 10) do canal vermelho dessas regiões: em virtude da saturação da cor vermelha, os níveis de intensidade dos pixels nesse canal concentram-se à direita, resultando em valores de obliquidade fortemente negativos. Esse fato também pode ser observado na tabela de dados estatísticos das regiões de fogo apresentada mais adiante.

Ambos os descritores propostos por Borges e Izquierdo (2010) citados até aqui fornecem somente informações estatísticas dos pixels que preenchem as regiões de interesse. São, nesse contexto, descritores de textura, mas não definem de maneira exata um modelo da aparência do fogo, tampouco descrevem as relações espaciais entre os pixels das regiões analisadas. Para tanto, descritores de textura mais robustos – e mais complexos - podem ser utilizados, como *padrões binários locais* (LBP) e *matrizes de coocorrência em nível de cinza* (GLCM). A fim de obter um maior grau de detalhamento do objeto de estudo sem tirar o foco da simplicidade, os autores propõem também uma métrica de forma: a *irregularidade do contorno*. Tendo em vista que o fogo turbulento não exhibe bordas definidas, argumenta-se que não é necessário empenhar-se em descrever sua forma com exatidão, dada a complexidade dessa tarefa, bastando uma medida de sua irregularidade, como a razão entre o perímetro e a envoltória convexa da região segmentada, expressa por:

$$B_R = \frac{P_S}{P_{CH_S}} \quad (42)$$

Uma alternativa semelhante é proposta por Xu e Xu (2007), que obtêm a medida de desordem da forma por meio da razão entre a circunferência e a área da região segmentada. O valor adotado como descritor é a média dessa divisão considerando n quadros consecutivos.

Liu e Ahuja (2004), em contrapartida, favorecem uma abordagem menos elementar, a qual prevê a utilização de coeficientes de Fourier obtidos pela transformação dos pontos do contorno da região segmentada como descritores. Os N pontos (x_i, y_i) de um contorno são expressos na forma complexa $\{z_i | z_i = x_i + jy_i\}_{i=0}^N$, cujos coeficientes da Transformada Discreta de Fourier são:

$$a_k = \frac{1}{N} \sum_{i=0}^N z_i e^{-j\frac{2\pi}{N}ik}, \text{ onde } k = \left[-\frac{N-1}{2}\right], \dots, \left[\frac{N}{2}\right] \quad (43)$$

Utilizando M harmônicos ($M \leq \lfloor (N - 1)/2 \rfloor$), os coeficientes $\{a_m\}_{m=-M}^M$ são os descritores de Fourier que caracterizam a forma das regiões segmentadas.

Em acidentes envolvendo fogo, quando a queima é desenfreada, há outra característica presente: a *propagação*. Regras de crescimento, como em Chen, Kao e Chang (2003) e Xu e Xu (2007), exploram essa característica e estabelecem um limiar mínimo de propagação da área segmentada antes de classificá-la como fogo. O crescimento é medido de acordo com a progressão do número de pixels das regiões detectadas, numa checagem realizada a intervalos regulares de tempo. Dessa forma, é possível, além de discernir entre fogo e outras regiões ordinárias, obter uma medida do seu alastramento, informando a gravidade do incêndio e a urgência de uma ação de combate.

4.4 TREINAMENTO

Em abordagens que envolvem aprendizagem de máquina, a classificação de cada região (*blob*) ou janela, de acordo com os descritores, é precedida de uma etapa de treinamento. Quando *supervisionado*, esse processo se inicia com a rotulação das amostras de um *dataset*, cada qual pertencente a uma classe específica. No presente caso, são apenas duas: *fogo* e *não fogo*. Isso significa que são necessários exemplos positivos e negativos das instâncias que fornecerão à máquina os dados para treinamento, que pode ser entendido como o processo que converge para a melhor divisão das instâncias entre as classes envolvidas. Submete-se para treinamento apenas o *vetor de descritores*, matriz unidimensional composta pelos elementos que descrevem cada instância, normalizado convenientemente. A fim de se otimizar a classificação, diversos parâmetros podem ser ajustados, desde os que são intrínsecos de cada método de classificação até o número e o balanceamento das instâncias de cada classe. Liu e Ahuja (2004), por exemplo, em virtude da natureza do problema, utilizam um número maior de exemplos positivos do que de negativos na etapa de treinamento, com o intuito de favorecer a detecção de falsos positivos e diminuir a de falsos negativos, visto que é mais nocivo que o sistema deixe de detectar um incêndio verdadeiro do que provoque um alarme falso.

4.5 CLASSIFICAÇÃO

Diversos métodos de classificação foram explorados pela literatura da detecção de fogo em vídeo. Dentre esses, talvez o mais popular seja o que emprega *máquina de vetores de suporte* (SVM) com kernel *função de base radial* (RBF). Exemplos dessa abordagem são encontrados nos trabalhos de Liu e Ahuja (2004), Zhao et al. (2011), Dimitropoulos, Tsalakanidou e Grammalidis (2012) e Jiang et al. (2014). Acompanhando o curso do desenvolvimento da computação visual, a utilização de redes neurais convolucionais é cada vez mais comum, tendo como exemplos os trabalhos propostos por Frizzi et al. (2016) e Muhammad et al. (2018). Outras aplicações incluem: redes bayesianas, como em Ko, Cheong e Nam (2010) e Borges e Izquierdo (2010); classificação baseada em regras, como em Phillips, Shah e Lobo (2000) e Chen, Wu e Chiou (2004).

Muitos dos sistemas propostos dependem de algum pós-processamento a fim de eliminar da classificação elementos pouco significativos. Em certos casos, adota-se a divisão dos quadros em sub-blocos; em outros, faz-se o uso de operações morfológicas – erosão e dilatação - ou de filtros, como os de suavização. De todo modo, o intuito é reduzir o esforço computacional, evitando, por exemplo, que ruído seja considerado relevante pelo sistema.

4.6 MÉTRICAS DE AVALIAÇÃO

Em problemas envolvendo aprendizagem de máquina, é de praxe que o resultado da classificação produzida seja avaliado conforme algumas medidas de desempenho. Na segmentação de imagens e na detecção de objetos, igualmente, a proposição de um método não pode deixar de acompanhar índices de qualidade padronizados. Dentre eles, dois são utilizados por este trabalho - o F_1 (também chamado de *F-measure* ou *F-score*) e o *MCC* (Matthews Correlation Coefficient). A literatura da segmentação de fogo em imagens, mais especificamente, aborda – embora não tão extensivamente – a comparação entre métodos de segmentação com base no desempenho obtido segundo esses índices: “A fim de comparar a performance de métodos de segmentação, podemos usar métricas padronizadas que comparam a imagem segmentada com uma imagem segmentada manualmente (o *ground truth*)” (TOULOUSE et al., 2015, p. 1067, tradução nossa). A aplicabilidade dessas ferramentas, no entanto, não está restrita à avaliação final de desempenho, sendo úteis também no processo de desenvolvimento e calibração das metodologias. Toulouse et al. (2015), por exemplo, estimam

os melhores limiares dos métodos investigados com base na otimização de F_1 . Já Santana, Gomes e Barata (2012), na proposição de seu sistema de detecção de fogo baseado em vídeo, avaliam diferentes combinações de modelos de cor comparando-as por meio do *MCC*.

Ambas as métricas são funções de TP (*true positive*), TN (*true negative*), FP (*false positive*) e FN (*false negative*), os números de verdadeiros positivos, verdadeiros negativos, falsos positivos e falsos negativos, respectivamente, definidos no quadro a seguir:

Quadro 1 - Conceitos de TP, TN, FP e FN

Verdadeiro positivo (TP):

Indicação correta de que uma condição é verdadeira.

Exemplo: indicar fogo quando realmente houver.

Verdadeiro negativo (TN):

Indicação correta que uma condição é falsa.

Exemplo: não indicar fogo quando realmente não houver.

Falso positivo (FP):

Indicação errada de que uma condição é verdadeira.

Exemplo: indicar fogo quando não houver.

Falso negativo (FN):

Indicação errada de que uma condição é falsa.

Exemplo: não indicar fogo quando houver.

4.6.1 F_1 , F-score ou F-measure

A métrica F_1 envolve as medidas de precisão (Pr) e revocação (Re), definidas por

$$Pr = \frac{TP}{TP + FP} \text{ e} \tag{44}$$

$$Re = \frac{TP}{TP + FN}, \tag{45}$$

e corresponde à sua média harmônica:

$$F_1 = 2 \times \frac{Pr \times Re}{Pr + Re} \tag{46}$$

A precisão (*precision*, mais usualmente) é definida nesse contexto como a fração de instâncias relevantes dentre as instâncias recuperadas, ou seja, a taxa de acertos dentre as indicações. Enquanto a revocação, (*recall*, mais usualmente), também chamada de sensibilidade, é definida como a fração de instâncias relevantes recuperadas dentre todas as instâncias relevantes, ou seja, a taxa de instâncias que devem ser indicadas e realmente são.

4.6.2 MCC (Matthews Correlation Coefficient)

Introduzido por Matthews (1975), o MCC é essencialmente uma medida de correlação entre a previsão e a observação em um processo de classificação binária, originalmente proposto como métrica na investigação da estrutura molecular de uma proteína.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TN + FN)(TN + FP)(TP + FN)(TP + FP)}} \quad (47)$$

Definido na equação (47), o MCC admite valores entre -1 e 1, uma escala que varia do total desacordo entre a previsão e a observação, no limite inferior, à predição perfeita, no limite superior. O ponto intermediário indica uma correlação meramente aleatória.

4.6.3 Outras métricas de avaliação

Tanto o MCC quanto o F_1 são utilizados por este trabalho na avaliação da performance dos métodos de segmentação de fogo em imagens. Além das medidas expostas, outras duas surgem adiante no desenvolvimento: a *taxa de verdadeiros positivos* (TPR) e a *taxa de falsos positivos* (FPR). A primeira é equivalente à revocação (*recall*), já apresentada na equação (45); a segunda representa a razão entre o número de instâncias falsas consideradas verdadeiras e o número total de instâncias realmente falsas. Em geral, um aumento da taxa de verdadeiros positivos faz crescer também a taxa de falsos positivos, em consequência de atributos de classificação mais brandos. Portanto, a otimização da classificação por ser encarada como uma busca sistematizada pela relação ótima entre TPR e PFR.

$$FPR = \frac{FP}{FP + TN} \quad (48)$$

5 DESENVOLVIMENTO

O projeto detalhado nas seções seguintes - e aqui resumido - foi guiado pela intenção de satisfazer um objetivo em particular: realizar a detecção de fogo em vídeo de maneira rápida e consideravelmente precisa. Em outras palavras, buscou-se encontrar em cada etapa do desenvolvimento um equilíbrio entre complexidade computacional e acurácia, com o propósito de atenuar a inconveniência dos falsos positivos e de possibilitar a operação do programa em tempo real. É certo, entretanto, que não foi dada a mesma ênfase a todas as etapas do projeto. Por se acreditar que a cor atua com protagonismo na caracterização do fogo, esse atributo recebeu maior destaque, motivando a criação de um *dataset* e a realização de um extensivo estudo comparativo de métodos de segmentação. Este estudo, por sua vez, embasou a escolha do método de segmentação por cor adotado no algoritmo proposto, que, em conjunto com a segmentação do movimento, resultou na extração das regiões candidatas a fogo em uma série de vídeos selecionados para treinamento. Utilizando varredura por blocos espaço-temporais, janelas contendo fogo em potencial foram rotuladas manualmente, com auxílio de um utilitário desenvolvido especificamente para essa função, a fim de se criar um novo *dataset* composto pelas classes “fogo” e “semelhante a fogo”. Por fim, a classificação, que utilizou o modelo SVM, foi ajustada por meio de validação cruzada e testada sobre duas combinações de descritores. O algoritmo final ainda conta com uma regra simples de detecção que avalia a localização e a recorrência das janelas contendo “fogo”; espera-se, dessa forma, evitar que classificações errôneas esporádicas acionem de imediato o alarme de detecção.

5.1 ALGORITMO PROPOSTO

O algoritmo proposto pode ser dividido em três etapas fundamentais: (1) segmentação das regiões candidatas a fogo, (2) varredura e classificação de blocos espaço-temporais e (3) tomada de decisão baseada em regras. Na primeira etapa, as regiões candidatas a fogo são segmentadas com base na cor e na ocorrência de movimento; a varredura extrai, então, as informações das janelas espaço-temporais apenas nas regiões segmentadas, obtendo o vetor de descritores que é submetido ao processo de classificação por SVM. Se houver detecção de uma janela contendo fogo, sua posição é salva e comparada à de possíveis novas ocorrências. A decisão final de se indicar a presença de fogo depende da satisfação de uma regra que tem como finalidade verificar a detecção recorrente naquela coordenada do vídeo. Por clareza, o panorama

de funcionamento do algoritmo proposto é apresentado após a exposição das etapas do projeto, na seção 5.8.

5.2 ESTUDO DA SEGMENTAÇÃO PELA COR

A escolha do método de segmentação por cor utilizado no desenvolvimento deste trabalho parte da avaliação de desempenho dos métodos apresentados da seção 4.1. Para isso, são adotados dois critérios de maior relevância: tempo de execução e valor do MCC. Além dos dados obtidos nessa avaliação, as seções a seguir apresentam uma breve exposição dos *datasets* utilizados e os dados estatísticos das regiões de interesse, levantados com o objetivo de verificar a procedência das relações inseridas nos métodos avaliados.

5.2.1 Datasets

Os métodos de segmentação implementados são aplicados às imagens de dois *datasets*; um criado especificamente para este trabalho e outro disponibilizado por Chino et al. (2015) - “*BoWFire*”. Enquanto este último é utilizado exclusivamente na etapa de avaliação, o *dataset* proposto participa também da otimização. Em ambos, cada imagem acompanha uma máscara binária correspondente contendo as anotações manuais dos pixels que apresentam fogo (o *ground-truth*), conforme pode ser visto na Figura 9. Onde há fogo, a máscara binária apresenta o valor *um* (branco); onde não há, apresenta o valor *zero* (preto).

Figura 9 - Exemplo de imagem do *dataset* proposto



Fonte (imagem original, à esquerda): New Mexico State University Fire Department – disponível em <https://fire.nmsu.edu/photos/>. À direita, *ground-truth* delineado manualmente.

5.2.1.1 Dataset 1 (proposto)

Com o objetivo deste trabalho em vista, o *dataset* proposto reúne somente imagens associadas ao conceito de incêndios “estruturais”. É limitado, portanto, a cenas envolvendo o interior ou o exterior de estruturas prediais capturadas a curta distância. Ao todo, são 100 imagens com qualidades variadas, extraídas de vídeos e fotos disponibilizados na internet. Dessas, 70 realmente contêm fogo e 30 possuem apenas cores que podem ser confundidas com fogo.

As anotações do *ground-truth* foram realizadas com o auxílio de uma mesa digitalizadora e um software de edição de imagens.

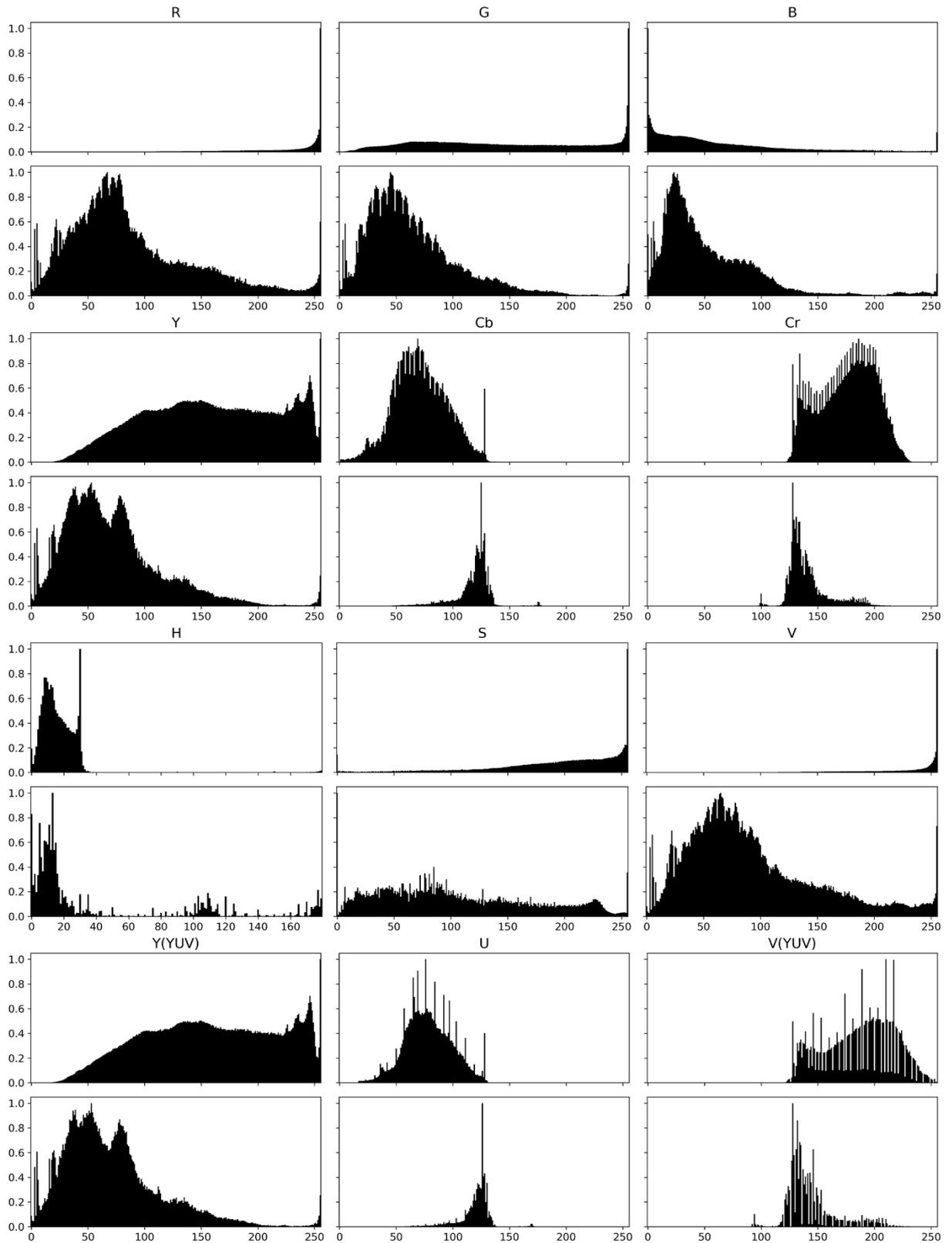
5.2.1.2 Dataset 2 (“BoWFire”)

O *dataset BoWFire* consiste de 226 imagens com resoluções variadas, das quais 119 contêm fogo e 107 não. As que contêm fogo retratam situações de emergência, como incêndios em prédios e indústrias, protestos e acidentes de carro, enquanto as demais apresentam situações de emergência sem fogo visível ou apenas regiões semelhantes a fogo (CHINO et al., 2015).

5.2.2 Dados estatísticos das regiões de interesse

As instâncias fundamentais (pixels) dos conjuntos de dados avaliados estão divididas em apenas duas categorias – “fogo” e “não fogo” -, ou seja, a segmentação aqui abordada é um problema de “classificação” binária, com a qual se pretende discernir uma categoria da outra. A fim de compreender melhor essa separação, são computados os dados presentes nas tabelas a seguir. Constam nelas a média, a mediana, o desvio padrão e a obliquidade dos pixels do *dataset* proposto em cada canal de cor contemplado pelos métodos de segmentação discutidos na seção 4.1. Os dados da Tabela 2 são calculados sobre todos os pixels da classe “fogo”, enquanto os da Tabela 3 são obtidos de um conjunto amostrado da classe oposta, de maneira que o número de pixels presentes em cada classe é equivalente (aproximadamente 3,2 milhões de pixels por classe).

Figura 10 - Histogramas das regiões anotadas como "fogo" e "não fogo"



Histogramas das regiões classificadas como “fogo” e “não fogo”. A primeira linha de cada modelo ilustra os histogramas, por canal, das regiões contendo fogo; imediatamente abaixo seguem os histogramas das regiões que não contêm fogo. Constam os modelos RGB (linhas 1 e 2), YCbCr (3 e 4), HSV (5 e 6) e YUV (7 e 8).

Tabela 2 - Dados estatísticos dos pixels anotados como fogo

	Média	Mediana	Desvio P.	Obliquidade
R	225,29	248	42,95	-1,671
G	144,52	138	73,43	0,069
B	58,50	40	59,39	1,405
Y _{YCbCr}	158,87	158	58,00	-0,116
Cb	71,37	70	23,74	0,094
Cr	175,37	178	24,77	-0,199
H*	16,86	14	14,78	6,728
S	191,63	207	59,65	-1,242
V _{HSV}	225,41	248	43,01	-1,670
Y _{YUV}	158,87	158	58,00	-0,116
U	78,62	77	20,70	0,096
V _{YUV}	186,26	189	30,48	-0,199

*Canais normalizados entre 0 e 255, com exceção de H (entre 0 e 180°)

Tabela 3 - Dados estatísticos dos pixels anotados como não-fogo

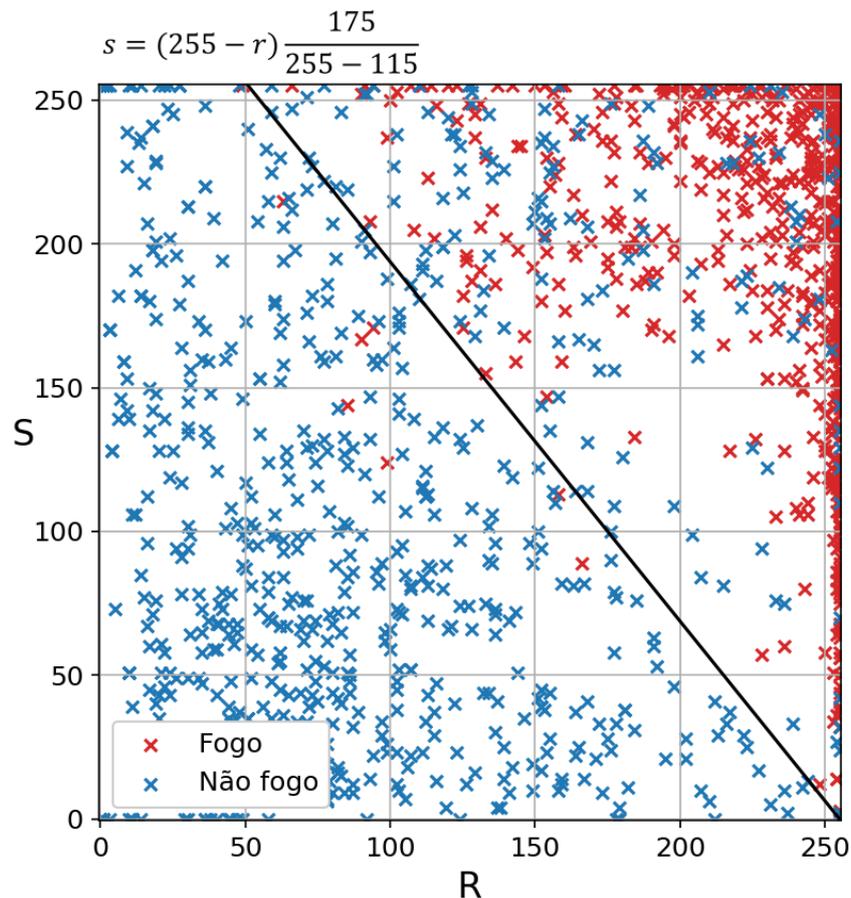
	Média	Mediana	Desvio P.	Obliquidade
R	101,41	89	67,52	0,565
G	77,38	63	60,03	1,013
B	61,71	43	59,82	1,389
Y _{YCbCr}	82,78	70	59,18	0,886
Cb	116,11	121	18,57	-1,431
Cr	141,30	136	19,00	1,039
H*	40,08	14	51,08	1,460
S	115,27	102	79,38	0,280
V _{HSV}	105,33	93	68,49	0,534
Y _{YUV}	82,78	70	59,18	0,886
U	117,68	122	16,20	-1,439
V _{YUV}	144,34	138	23,34	1,033

*Canais normalizados entre 0 e 255, com exceção de H (entre 0 e 180°)

A partir de uma simples comparação entre os canais de cor, ou entre uma tabela e outra, podem ser compreendidas algumas das condições impostas pelos métodos de segmentação discutidos anteriormente. A condição (12), segundo a qual os pixels de fogo são descritos por $R > G > B$, talvez seja a mais óbvia. Outra propriedade facilmente observável é a que diz respeito à diferença entre os canais Cb e Cr, muito maior na Tabela 2. Nota-se também a maior concentração de H em torno de sua média para os pixels de fogo, fato evidenciado por um desvio padrão reduzido e por uma obliquidade positiva acentuada.

Dito isso, não seria razoável deixar de destacar o fato de que as estatísticas descritas possuem limitações, uma vez que qualquer relação entre pixels formulada a partir delas está condicionada ao pressuposto de que as componentes de cor aqui presentes são independentes entre si. Em consequência disso, condições como a definida em (13) - na qual S é dada em função de R - não podem ser deduzidas. Para tal, devem ser realizados cruzamentos entre componentes de modelos de cor distintos, como ilustra a Figura 11, na qual a condição (13) se torna evidente: pixels anotados como fogo podem ser separados da classe oposta por uma reta decrescente que relaciona as componentes R e S.

Figura 11 - Separação das classes "fogo" e "não fogo" no plano R-S



O gráfico exibe, por amostragem, os pixels do *dataset* proposto no plano R-S, sendo R o canal vermelho do modelo RGB e S a saturação do modelo HSV. A reta traçada ilustra a separação das classes “fogo” e “não fogo” segundo a relação (13), com $R_T = 115$ e $S_T = 175$ (valores obtidos pelo processo de otimização).

5.2.3 Estudo comparativo dos métodos de segmentação por cor

O processo de avaliação e escolha do método de segmentação por cor mais adequado exige antes algumas considerações. Em primeiro lugar, é importante lembrar que cada método aqui discutido é decorrente de uma aplicação específica. Isso significa que, na sua origem, outros critérios foram adotados na sua formulação e otimização. Em certos casos, a complexidade computacional - ou o tempo de execução - nem sequer é relevante. Por essa razão, não se faz aqui um juízo da qualidade geral de cada método; em vez disso, pretende-se apenas escolher aquele que melhor se alinha aos objetivos deste trabalho.

Em síntese, o que se espera é que o método de segmentação por cor seja rápido e capaz de recortar fogo em imagens sem ignorar uma quantidade considerável de regiões importantes. Acredita-se que numa etapa preliminar de segmentação o reconhecimento das áreas de interesse seja mais relevante do que o descarte de falsos positivos; nesta aplicação, principalmente, a não detecção implica em consequências mais graves do que a ocorrência de detecções equivocadas. Por outro lado, uma segmentação excessivamente branda também é prejudicial à eficiência do método nas etapas subsequentes, visto que as verificações seguintes ocorrem por eliminação e que, portanto, uma segmentação exagerada levaria mais tempo a ser percorrida.

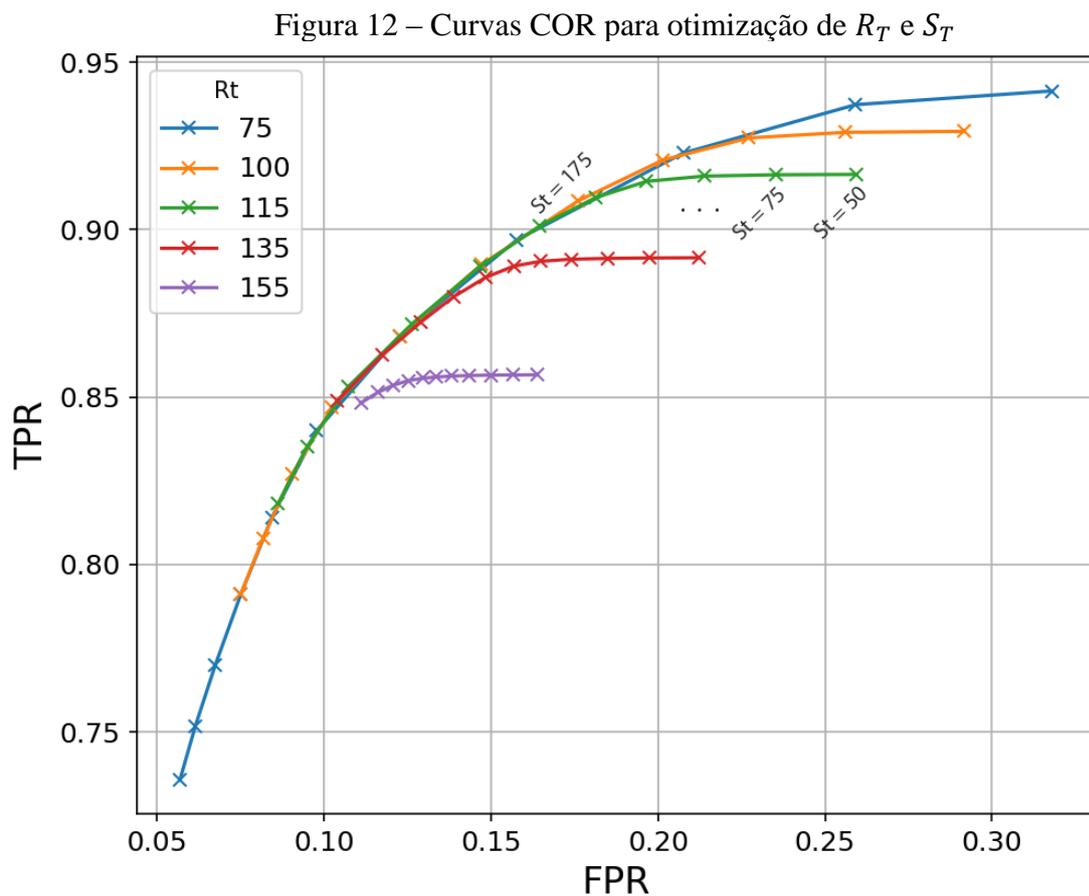
Considera-se também a possibilidade de que os métodos aqui tratados não tenham sido implementados ou otimizados da melhor maneira. A qualidade da implementação, especificamente, muitas vezes depende da eficiência das funções contidas nas bibliotecas utilizadas. A otimização de um dado método, por sua vez, além de ser condicionada aos critérios do autor, muitas vezes carece de melhores instruções em sua origem. De todo modo, um amplo esforço foi empreendido na tentativa de explorar o potencial máximo de cada metodologia.

Acrescenta-se, também, que nenhum dos métodos implementados foi submetido a qualquer processo de filtragem espacial ou transformação morfológica, dado que o interesse está voltado ao resultado não manipulado de cada segmentação.

5.2.3.1 Processo de otimização

A otimização dos métodos de segmentação por cor foi realizada priorizando-se a taxa de verdadeiros positivos, também conhecida como *recall*. Essa escolha se baseia na premissa, já citada, de que a taxa de detecção, nesta primeira etapa, é mais importante do que a eliminação de falsos positivos. Para isso, os valores ótimos foram determinados a partir de curvas COR

(Característica de Operação do Receptor), como as ilustradas a seguir, nas quais se relacionam as taxas de verdadeiros positivos (TPR) e de falsos positivos (FPR). Nos casos em que a otimização dependia de mais de uma constante ou limiar, um conjunto de curvas foi traçado. Sempre que possível, os valores foram ajustados de forma que a taxa de verdadeiros positivos (TPR) fosse de 90%; respeitado esse primeiro critério, optou-se pela combinação de valores que resultou no melhor MCC.

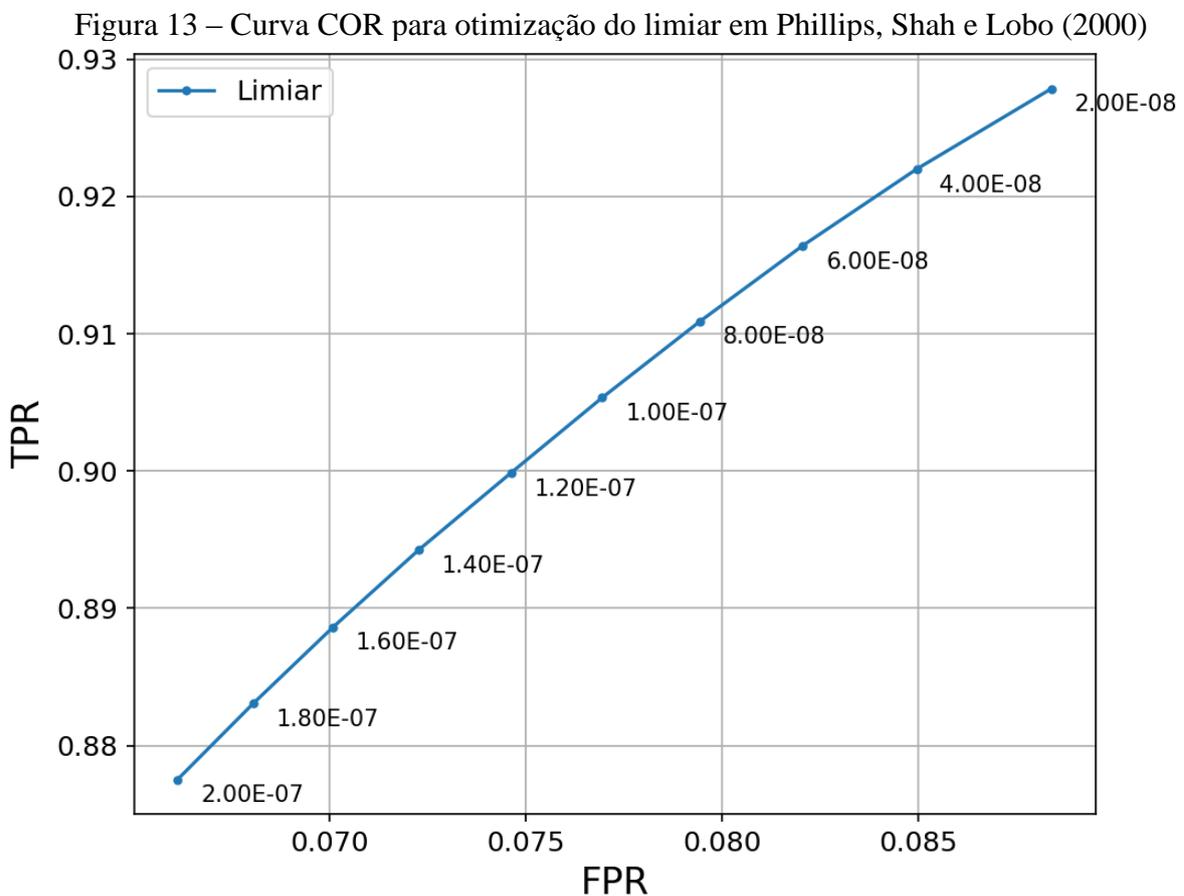


Relação entre TPR e FPR para diferentes valores de R_T e S_T (CHEN; KAO; CHANG, 2003). Observa-se que para $R_T = 115$ e $S_T = 175$ o valor de FPR é o menor possível dentre os pontos com TPR maior que 90%.

Para o método proposto por Chen, Kao e Chang (2003) (Figura 12), as melhores constantes encontradas foram $R_T = 115$ e $S_T = 175$. Para Horng, Peng e Chen (2005), $[h_1, h_2] = [0, 30]$, $[s_1, s_2] = [55, 255]$ e $[i_1, i_2] = [120, 255]$.

A segmentação sugerida por Phillips, Shah e Lobo (2000) exige a construção prévia de um histograma 3D de 256^3 bins no modelo de cores RGB. Conforme explicado na seção 4.1.3, isso é feito pela soma de distribuições gaussianas centradas em cada pixel do *dataset* (no presente caso, o *dataset 1*). A princípio, a tentativa de efetuar esse cálculo com a distribuição

completa no espaço compreendido pelo cubo de lado 256 resultou num processo computacionalmente custoso – para cada pixel, de um conjunto de aproximadamente 6,5 milhões de pixels, seria executada a soma de uma distribuição gaussiana de 256^3 valores. Esse problema foi solucionado truncando-se a distribuição em cinco desvios padrões, com a segurança de que essa aproximação não comprometeria o desempenho. Por fim, a limiarização do histograma foi otimizada para $\tau = 1.1 \times 10^{-7}$ (Figura 13); *bins* abaixo desse valor são zerados e *bins* acima são transformados em 255. Dessa forma, o histograma é tratado como uma *look-up table* que indica os valores do espaço RGB (Figura 14) traduzidos como fogo.

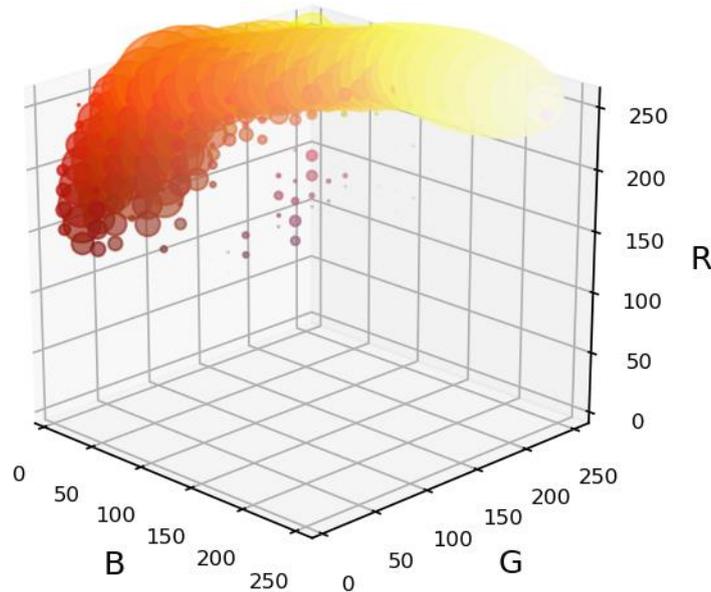


Relação entre TPR e FPR para diferentes limiares de corte do histograma utilizado na implementação do método proposto por Phillips, Shah e Lobo (2000). Observa-se que para um limiar logo abaixo de $1,2 \times 10^{-7}$ o valor de FPR é o menor possível dentre os pontos com TPR maior que 90%.

A retroprojeção de histograma foi implementada nos modelos de cores HSV e YCbCr. Em ambos, o histograma modelo foi gerado conforme a receita proposta por Phillips, Shah e Lobo (2000), mas com um número de *bins* reduzido, dado que Wirth e Zaremba (2010) sugerem a utilização de histogramas de apenas 32 *bins* por canal. Embora se tenha respeitado essa sugestão, em outro ponto a implementação sofreu uma alteração crucial: na proposta original,

somente os canais de cores de cada modelo foram utilizados, resultando em histogramas de no máximo duas dimensões; na implementação avaliada por este trabalho, no entanto, todos os canais foram utilizados, resultando em histogramas de 32^3 bins.

Figura 14 - Representação da "nuvem" de pixels cor de fogo



Fonte: autoria própria. Representação gráfica do histograma gerado a partir dos pixels registrados como fogo no *dataset 1*. O diâmetro de cada círculo indica a frequência de ocorrência dos níveis de intensidade na faixa correspondente. Para efeito de visualização, são utilizados apenas 16 bins por canal.

A otimização do método proposto por Rudz et al. (2013) requer a determinação de várias constantes e valores de referência. Por essa razão, devido à complexidade, os melhores valores foram encontrados heurísticamente, da seguinte forma: com auxílio do *ground-truth*, as regiões segmentadas na etapa de agrupamento foram divididas entre dois grupos, um de regiões contendo realmente fogo e outro de falsos positivos. Calculou-se, então, a média e o desvio padrão de cada grupo, dados que serviram na escolha do limiar que melhor os separa. Dessa forma, ficou definido que $\tau_R = 0,317$, $\tau_G = 0,108$, $\tau_B = 0,191$, $k_R = 0,844$, $k_G = 1,09$ e $k_B = 0,828$. Os demais dados de referência foram gerados diretamente do *dataset 1*.

O outro método de agrupamento, de Rossi e Akhloufi (2009), em contrapartida, requer uma única constante, cujo valor ótimo encontrado foi $k = 4$, também por intermédio de uma curva COR.

É importante destacar que todos os histogramas de referência utilizados foram normalizados dividindo-os por sua soma, isto é, são equivalentes a funções densidade das suas respectivas variáveis.

5.2.3.2 Desempenho dos métodos de segmentação

A Tabela 4 exibe os tempos relativos de execução, em ordem crescente, tendo como referência o método mais rápido - Horng, Peng e Chen (2005).

Tabela 4 - Tempo relativo de execução sobre o *dataset 1*

	Tempo Relativo de Execução
1. Horng, Peng e Chen (2005)	1,0000
2. Chen, Kao e Chang (2003)	1.1347
3. Phillips, Shah e Lobo (2000)	1.5871
4. Wirth e Zaremba YCrCb (2010) ¹	4.0899
5. Wirth e Zaremba HSV (2010)	4.4969
6. Çelik e Demirel (2009)	4.8246
7. Rossi e Akhloufi (2009)	7.3122
8. Rudz et al. (2013)	108.22

As tabelas a seguir expõem o desempenho de cada método segundo quatro das métricas previstas – precisão (*precision*), recuperação (*recall*), F_1 e MCC. Em negrito estão destacados os melhores resultados obtidos em cada métrica de avaliação.

Tabela 5 - Dados de desempenho por método de segmentação (*dataset 1*)

	Precision	Recall	F_1	MCC
1. Horng, Peng e Chen (2005)	0,221	0,903	0,354	0,379
2. Chen, Kao e Chang (2003)	0,278	0,901	0,425	0,446
3. Phillips, Shah e Lobo (2000)	0,456	0,903	0,606	0,609
4. Wirth e Zaremba YCrCb (2010)	0,424	0,902	0,576	0,582
5. Wirth e Zaremba HSV (2010)	0,441	0,904	0,593	0,598
6. Çelik e Demirel (2009)	0,338	0,640	0,442	0,414
7. Rossi e Akhloufi (2009)	0,203	0,751	0,320	0,314
8. Rudz et al. (2013)	0,639	0,363	0,463	0,456

Tabela 6 - Dados de desempenho por método de segmentação (*dataset 2*)

	Precision	Recall	F_1	MCC
1. Horng, Peng e Chen (2005)	0,1594	0,8472	0,2684	0,3091
2. Chen, Kao e Chang (2003)	0,2074	0,8487	0,3333	0,3718
3. Phillips, Shah e Lobo (2000)	0,2744	0,8373	0,4133	0,4412
4. Wirth e Zaremba YCrCb (2010)	0,3131	0,8549	0,4583	0,4838
5. Wirth e Zaremba HSV (2010)	0,3330	0,8339	0,4760	0,4949

¹ Os métodos 4 e 5 não foram implementados exatamente como proposto no artigo original.

6. Çelik e Demirel (2009)	0,2306	0,6707	0,3432	0,3471
7. Rossi e Akhloufi (2009)	0,1200	0,7493	0,2069	0,2257
8. Rudz et al. (2013)	0,4202	0,2837	0,3387	0,3202

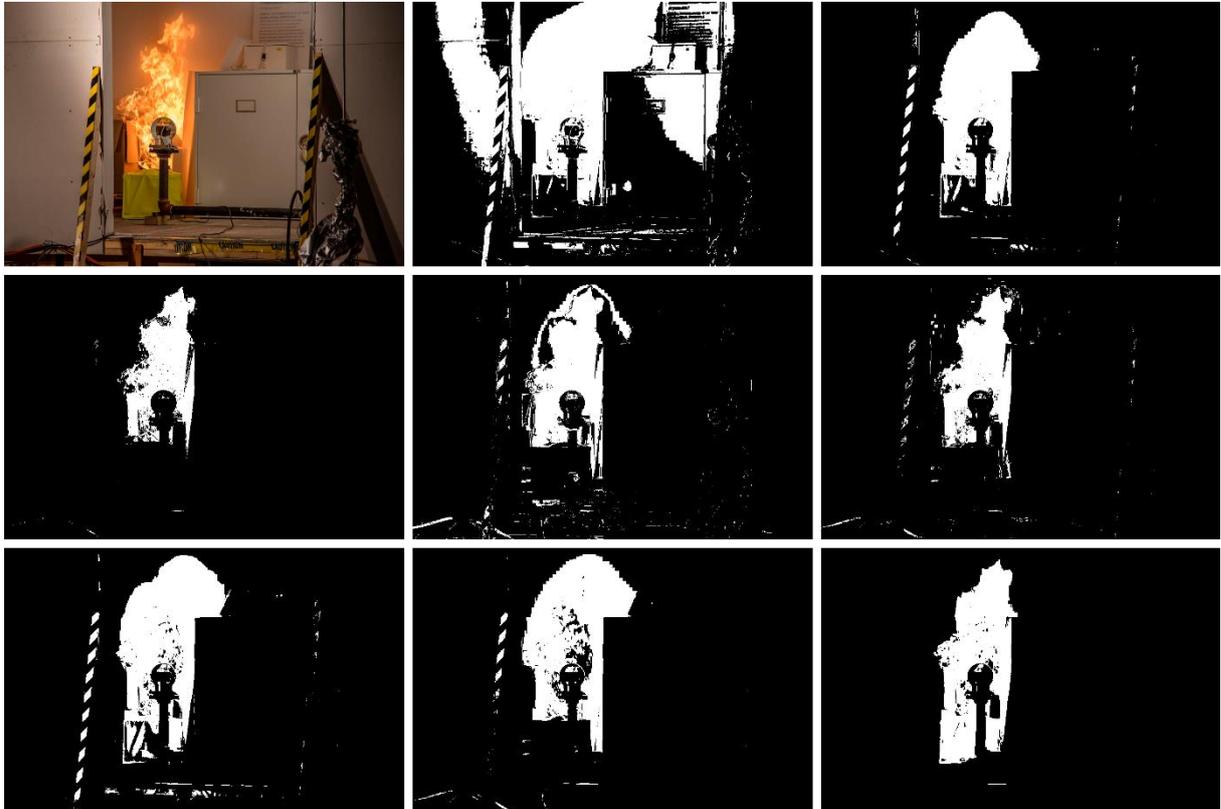
Quanto ao tempo de execução, são levantadas as seguintes conclusões: os métodos de segmentação baseados em regras são, em geral, mais rápidos que os demais. Os que realizam indexação via histograma são suficientemente rápidos, porém a necessidade de se calcular um novo histograma a cada repetição torna os métodos de retroprojeção mais demorados que o método apresentado por Phillips, Shah e Lobo (2000), no qual o histograma modelo é utilizado como uma simples *look-up table*. Os métodos que envolvem agrupamento de dados são mais lentos, não somente pelo processo de agrupamento, mas principalmente pela etapa de filtragem subsequente.

Quanto ao desempenho, fica evidente que os métodos de indexação via histograma apresentam os melhores índices. Há uma perda geral de desempenho na etapa de teste (*dataset 2*), mas as taxas de *recall* permanecem razoavelmente altas, com exceção de Rudz et al. (2013), que se destaca mais por sua precisão. Essa diferença substancial se deve à filtragem dos falsos positivos, que, embora torne a segmentação mais certa, descarta erroneamente regiões importantes. Dos métodos de segmentação baseados em regras, o de Chen, Kao e Chang (2003) é o que mais se destaca, especialmente se o tempo de execução for um fator determinante. Outra conclusão importante, ainda sobre os métodos baseados na indexação via histograma, é referente à influência dos modelos de cores utilizados. A retroprojeção, tanto em HSV quanto YCrCb é mais constante que a indexação no modelo RGB, fato evidenciado por uma queda de desempenho menor entre os resultados de treinamento e teste.

Os dados de desempenho e tempo de execução obtidos por meio da análise exposta não indicam uma superioridade óbvia. De acordo com a rapidez e os valores de *recall*, os três primeiros apresentam resultados satisfatórios. No entanto, tendo em vista que o custo computacional da etapa seguinte é diretamente proporcional ao tamanho da segmentação obtida nesta etapa, é sensato que a escolha leve em conta também a capacidade de eliminação de falsos positivos. Por essa razão, o método de Phillips, Shah e Lobo (2000) parece se adequar melhor aos objetivos deste trabalho, ainda que seja bastante sensível a variações de cor e iluminação, por ter seu desempenho intimamente relacionado ao conjunto de dados utilizado em sua implementação. De fato, a observação mostra que a segmentação obtida por meio deste método

é, em certas situações, excessivamente restritiva, enquanto o método de Chen, Kao e Chang (2003) se mostra mais constante.

Figura 15 – Exemplos dos resultados obtidos por cada método de segmentação avaliado



Fonte (imagem original): J. Stoughton/NIST – disponível em <https://www.nist.gov/blogs/taking-measure/rescuing-our-history-ashes>. Na primeira linha, da esquerda para a direita: imagem original; Horn; e Chen. Na segunda linha: Phillips; Wirth YCbCr; e Wirth HSV. Na terceira linha: Celik; Rossi; e Rudz.

5.3 SEGMENTAÇÃO DO MOVIMENTO

Dentre os métodos de segmentação do movimento discutidos na revisão bibliográfica, muitos satisfazem o critério do baixo esforço computacional esperado nesta etapa do projeto. Os usuais métodos adaptativos de subtração de fundo, por exemplo, são especialmente eficientes e servem muito bem ao propósito deste trabalho. Diante de um conjunto de opções, a escolha do método de segmentação do movimento acabou por se basear numa especificidade da abordagem proposta: em virtude da seleção de descritores, os quais são elaborados a seguir, sabe-se, de antemão, que um dos requisitos do método é a geração da imagem do histórico do movimento para cada quadro da sequência. A partir dela, obtém-se facilmente a imagem da energia do movimento, por limiarização. Uma vez que essa nova imagem representa justamente a silhueta do primeiro plano, considerou-se desnecessária e ineficiente a adição de outro

algoritmo independente de segmentação do movimento. Além disso, acredita-se que essa abordagem atenua eficientemente os falsos negativos causados pela saturação que frequentemente ocorre nas imagens de fogo: conforme apontam Liu e Ahuja (2004), o fogo costuma exibir uma estrutura de anéis concêntricos; na periferia, onde a temperatura é menor e as chamas contrastam com o fundo, é possível observar a sua dinâmica com mais clareza; no centro, em contrapartida, onde a temperatura é maior e as bordas não são visíveis, costuma haver forte saturação, o que torna a região aparentemente estática. Essa “ausência” de movimento torna inevitável que uma fração considerável das chamas seja ignorada. Entretanto, devido à propagação do fogo e ao seu comportamento caótico, as regiões agora “estáticas” outrora apresentaram movimento. A fim de possibilitar a detecção desse movimento mais antigo, o método de segmentação deve possuir memória considerável, o que pode ser obtido sem muito esforço por meio da imagem do histórico do movimento.

Em detalhes, o procedimento de segmentação do movimento segue os seguintes passos:

- 1) Por diferenciação (*frame differencing*), extrai-se o movimento entre quadros consecutivos, considerando um limiar mínimo de variação de intensidade de 15;
- 2) Segundo a equação (40), constrói-se a tMHI, com duração de dez quadros e τ variando em incrementos de 0,05 s (convencionou-se uma taxa de 20 quadros por segundo);
- 3) Obtém-se a MEI – admitida como a máscara binária do movimento - pela limiarização da MHI, de forma que todo pixel maior que zero é transformado em 255.

5.4 VARREDURA E ROTULAÇÃO

Com a finalidade de produzir um *dataset* composto por um número satisfatório de exemplos, 43 vídeos foram selecionados. Desses, 27 foram utilizados para treinamento e 16 para testes. O requisito para inclusão foi simples: cada vídeo deveria conter fogo ou objetos semelhantes a fogo registrados a curta distância com uma câmera estacionária. Preocupou-se também com a diversidade de cenários e condições de iluminação.

Assim como no algoritmo proposto, cujo panorama de funcionamento é dado ao fim deste capítulo, os vídeos foram redimensionados para a resolução de 320 x 240 pixels.

Posteriormente, cada quadro foi submetido a uma varredura por blocos de 16 x 16 pixels – um total de 300 blocos por quadro – em busca de regiões contendo pixels candidatos a fogo. Aqui, a *intersecção* entre a máscara do movimento e a obtida pela segmentação da cor foi utilizada como critério de marcação, conforme ilustra a sequência da Figura 16.

Figura 16 - Marcação das regiões candidatas a fogo



Fonte (imagem original): captura de tela de um dos vídeos do *dataset* apresentado por Steffens, Rodrigues e Botelho (2015). Na primeira linha, da esquerda para a direita: imagem original; segmentação da cor; segmentação do movimento. Na segunda linha, da esquerda para a direita: intersecção das máscaras anteriores; região de interesse; marcações automáticas.

Quadros contendo fogo em potencial foram marcados automaticamente e depois supervisionados. Considerou-se haver fogo somente nas janelas em que ao menos 25% dos pixels foram anotados como candidatos, a fim de se reduzir o número de marcações provocadas por ruído.

A marcação automática presumiu a presença de fogo e foi manualmente desfeita, quando equivocada, com o auxílio de uma interface desenvolvida especificamente para essa função (Figura 17). Por fim, as janelas foram divididas em dois grupos de amostras para treinamento – “fogo” e “semelhante a fogo”. Cada exemplo do *dataset* consiste de um grupo de cinco imagens (ou matrizes): (1) imagem original em RGB; (2) intersecção das máscaras de cor e movimento; (3) matriz do gradiente de movimento; (4) máscara binária do gradiente de movimento; e (5) imagem do histórico do movimento. As imagens e máscaras extraídas de cada janela possuem funções específicas relacionadas aos descritores utilizados posteriormente no treinamento e classificação. Por simplificação, a intersecção das máscaras de cor e movimento,

que resulta na máscara binária com a anotação dos pixels candidatos a fogo, será referida adiante como $M_{cor \cap mov}$. Já a máscara do gradiente de movimento, que aponta os pixels onde o gradiente é não nulo, será referida como M_{mg} . Em ambas, o valor *um*, representado pela cor *branca*, indica as regiões de interesse. As demais posições, representadas pela cor *preta*, têm valor *zero*.

Figura 17 - Interface de rotulação

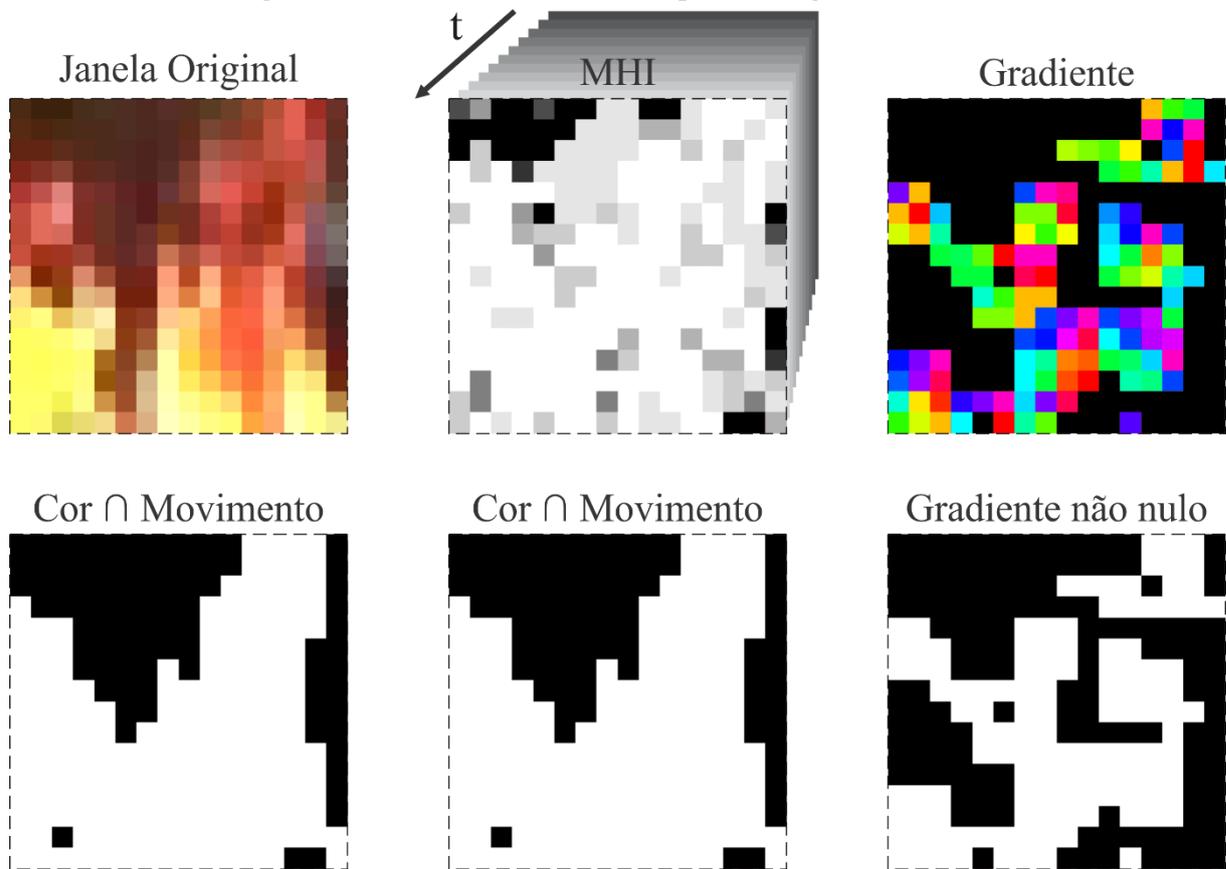


Fonte (imagem original): captura de tela de um dos vídeos do *dataset* apresentado por Steffens, Rodrigues e Botelho (2015). Os blocos contornados em verde indicam as anotações automáticas das regiões candidatas a fogo. A segunda janela, que ilustra a etapa de marcação manual, apresenta os blocos que realmente contém fogo, mantidos em verde, e os que foram clicados por não apresentarem fogo, contornados em vermelho. Esses grupos formam as duas classes do conjunto de dados utilizado para treinamento – “fogo” e “semelhante a fogo”.

5.5 DESCRITORES

O custo computacional também foi uma preocupação na escolha dos descritores utilizados na classificação. Por essa razão, optou-se por uma abordagem estatística na caracterização das grandezas envolvidas. O grupo de descritores se divide entre os de cor (média e desvio padrão de cada canal do modelo YCbCr, desvio padrão do canal G e obliquidade do canal R) e os de movimento (média e desvio padrão do histórico e do gradiente do movimento), sendo 12 no total. Alguns já tiveram sua validade verificada pela literatura, outros, até onde se pôde verificar, carecem de avaliação de eficácia.

Figura 18 - Instâncias de um exemplo do conjunto de dados



Fonte: autoria própria. Na primeira linha, matrizes do conjunto de dados obtidas a partir da rotulação das janelas na interface gráfica; cada exemplo é composto por três matrizes e duas máscaras distintas; a segunda linha ilustra as máscaras associadas às matrizes correspondentes.

5.5.1 Dados estatísticos da cor

A segmentação preliminar da cor, discutida na seção 5.2, é realizada segundo o método proposto por Phillips, Shah e Lobo (2002), cujo mapeamento, pixel a pixel, utiliza o modelo de cores RGB. Na presente etapa, tendo em vista que as imagens são analisadas bloco a bloco, a cor é empregada novamente como característica discriminante, sob uma nova ótica, que leva em conta os valores de toda uma região. Para isso, no entanto, a fim de evitar elementos redundantes, optou-se por um modelo de cores distinto: o YCbCr. Uma vez que sua representação cromática se dá por meio das componentes diferença-azul e diferença-vermelho, é particularmente apropriado à tarefa de distinguir as cores comumente envolvidas num incêndio. Essa propriedade discriminativa da representação cromática do modelo de cores YCbCr é observada por Rudz et al. (2013), cuja pesquisa demonstra que o canal Cb produz o melhor resultado de segmentação do fogo em imagens quando utilizado, sozinho, numa abordagem baseada em agrupamento de dados. Çelik e Demirel (2009), por outro lado, apontam

para a importância das relações entre os canais do modelo YCbCr, traduzidas pelas equações (16) a (20), dentre as quais se destaca a que explora a distância exibida entre os valores de Cb e Cr (19).

Outros dois importantes elementos são propostos por Borges e Izquierdo (2010), conforme exposto na seção 4.3: a variância do canal G e a obliquidade do canal R. A importância do primeiro se deve ao fato de as regiões de fogo apresentarem grande variação de intensidade no canal verde; a segunda, por sua vez, explora a tendência de saturação do canal vermelho. Por serem valores estatísticos da distribuição de cores numa região, ambos podem ser encarados como descritores de textura.

Com a segurança que provém dessas conclusões, delinea-se o primeiro conjunto de descritores, D_{cor} :

$$D_{cor} = [\mu_Y, \mu_{Cb}, \mu_{Cr}, \sigma_Y, \sigma_{Cb}, \sigma_{Cr}, \sigma_G, \gamma_R] \quad (49)$$

sendo μ a média, σ o desvio padrão e γ a assimetria (ou obliquidade) do canal subscrito correspondente. Nos cálculos são considerados somente os pixels p onde $M_{cor \cap mov} = 1$, de modo que para um dado canal c :

$$\mu_c = \frac{1}{n} \sum_{x,y} p_c(x,y) \quad (50)$$

$$\sigma_c = \sqrt{\frac{1}{n} \sum_{x,y} [p_c(x,y) - \mu_c]^2} \quad (51)$$

$$\gamma_c = \frac{1}{n\sigma_c^3} \sum_{x,y} [p_c(x,y) - \mu_c]^3 \quad (52)$$

para x, y onde $M_{cor \cap mov} = 1$, sendo n a contagem de posições nas quais $M_{cor \cap mov} = 1$. Essa medida, que também pode ser entendida como a área da região que contém os pixels candidatos a “fogo”, é dada por:

$$n = \sum_{x,y} M_{cor \cap mov}(x, y) \quad (53)$$

5.5.2 Dados estatísticos do movimento

Conforme indica a revisão bibliográfica, a dinâmica do fogo pode ser explorada de diversas formas, seja por análise temporal, espectral ou espacial. De todo modo, o que se pretende é diferenciá-lo de outros objetos ordinários em movimento fazendo uso de sua natureza caótica e intermitente. Em termos mais específicos, essa característica das chamas pode ser traduzida por uma ausência de orientação evidente – enquanto se observa uma ordenação (tendência) no movimento ordinário, nas chamas o deslocamento aparenta seguir, simultaneamente, múltiplas direções. Apostando na representação desse comportamento obtida pela imagem do histórico do movimento (MHI) e por seu gradiente (MG), concebeu-se o segundo grupo de descritores, D_{mov} :

$$D_{mov} = [\mu_{MHI}, \mu_{MG}, \sigma_{MHI}, \sigma_{MG}] \quad (54)$$

sendo μ a média e σ o desvio padrão dos subscritos correspondentes, calculados de maneira análoga à já apresentada na seção anterior:

$$\mu_{MHI} = \frac{1}{n} \sum_{x,y} MHI(x, y) \quad (55)$$

$$\sigma_{MHI} = \sqrt{\frac{1}{n} \sum_{x,y} [MHI(x, y) - \mu_{MHI}]^2} \quad (56)$$

para x, y onde $M_{cor \cap mov} = 1$, sendo n novamente a contagem de posições nas quais $M_{cor \cap mov} = 1$. Os cálculos das medidas que envolvem o gradiente de movimento empregam uma máscara binária distinta, M_{mg} :

$$\mu_{MG} = \frac{1}{n} \sum_{x,y} MG(x, y) \quad (57)$$

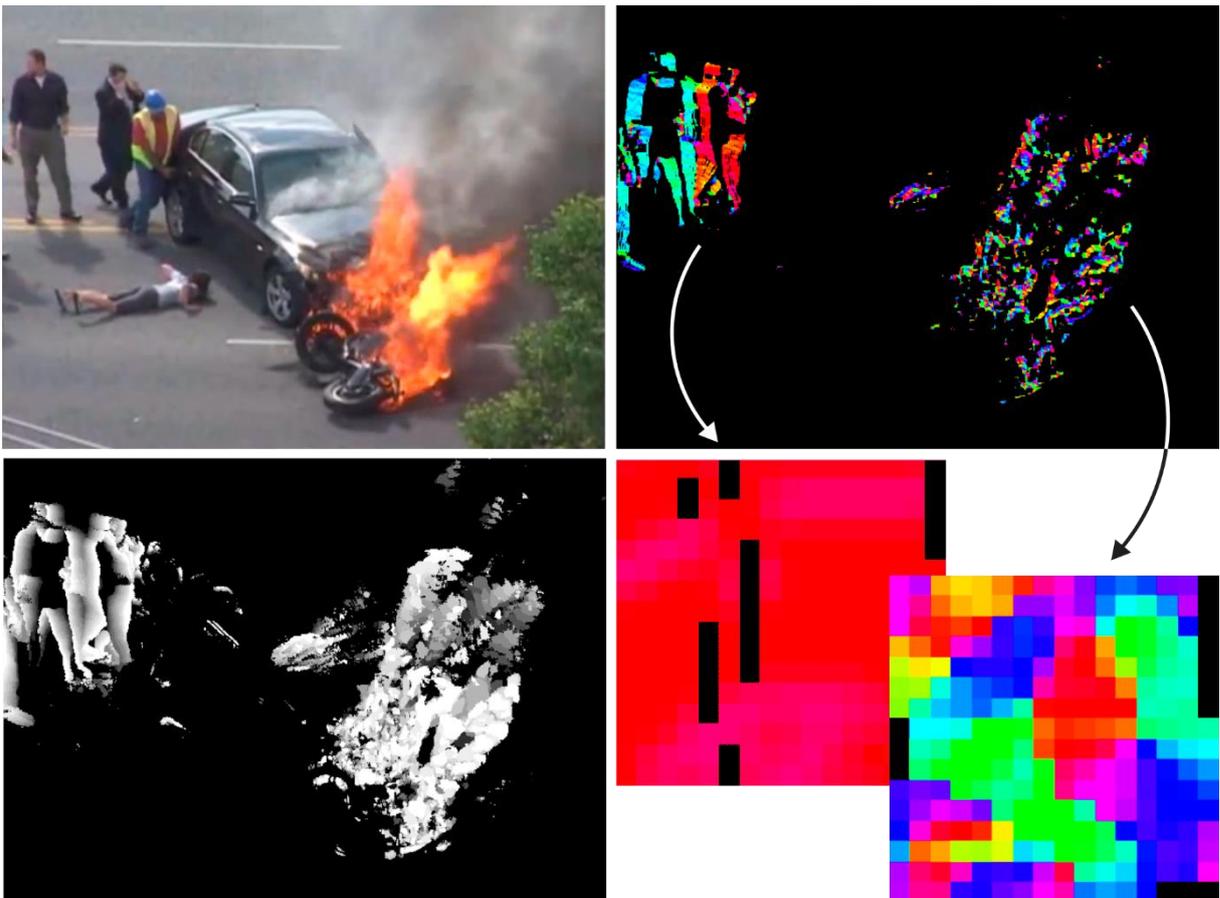
$$\sigma_{MG} = \sqrt{\frac{1}{n} \sum_{x,y} [MG(x,y) - \mu_{MG}]^2} \quad (58)$$

para x, y onde $M_{mg} = 1$, sendo n a contagem de posições nas quais $M_{mg} = 1$. Essa medida, que também pode ser entendida como a área da região que compreende as posições onde o gradiente de movimento é válido, é dada por:

$$n = \sum_{x,y} M_{mg}(x,y) \quad (59)$$

A imagem do histórico do movimento e o gradiente de movimento foram computados de acordo com Bobick e Davis (2001) e Bradski e Davis (2002) (ver seções 4.2.3 e 4.2.4).

Figura 19 - Diferença entre o movimento ordinário e o desempenhado pelas chamas



Fonte (imagem original): captura de tela de um dos vídeos do *dataset* apresentado por Steffens, Rodrigues e Botelho (2015). Diferentes perfis de movimento evidenciados pelas imagens do histórico e do gradiente de movimento.

A ilustração do gradiente do movimento é apenas uma representação que mapeia um ângulo a uma cor correspondente; o descritor, de fato, é uma matriz com valores entre 0 e 359 – um para cada orientação angular.

5.6 VALIDAÇÃO E TREINAMENTO

O *dataset* para treinamento, elaborado conforme apresentado na seção 5.4, é composto por 8.402 amostras - 4047 positivas e 4.355 negativas. Por meio de validação cruzada, um modelo SVM (*support vector machine*) foi otimizado sobre esse *dataset*, adotando como métrica de avaliação o MCC. Além da alternância dos parâmetros do modelo, também foram testadas diferentes combinações de descritores, separados por categoria, a fim de se verificar a influência de cada uma no resultado da classificação. Os resultados desse processo são exibidos na seção seguinte.

As combinações de descritores seguem a divisão da seção anterior, resultando em três cenários. No primeiro, são considerados somente os descritores de cor, no segundo, somente os descritores de movimento; por fim, considera-se a união dos vetores de descritores. Com base no conjunto completo de dados, a cada descritor foi atribuída uma função de normalização, de forma que todos fossem escalonados e transladados para que ocupassem uma faixa de valores entre -1 e 1. Essa medida é fundamental para um bom desempenho do modelo SVM (HSU; CHANG; LIN, 2003). Evidentemente, a função de normalização computada na etapa de treinamento é a mesma adotada na etapa de testes.

5.7 RESULTADOS DE OTIMIZAÇÃO E CLASSIFICAÇÃO

A qualidade da classificação é aferida em duas etapas: validação e teste. A primeira medida, obtida por validação cruzada, provém do próprio processo de otimização sobre o *dataset* para treinamento mencionado na seção anterior; a segunda é o resultado alcançado pelo modelo ótimo de classificação sobre um *dataset* para teste, criado especificamente para essa função a partir de um conjunto distinto de vídeos. A seguir são exibidos os resultados da classificação por categoria. Em cada um dos três casos particulares a otimização do modelo foi realizada por um processo de validação cruzada de dez dobras (*10-fold*).

5.7.1 Treinamento e otimização da classificação

No primeiro cenário, o vetor de descritores é composto somente pelo grupo de descritores de cor: $\mu_Y, \mu_{Cb}, \mu_{Cr}, \sigma_Y, \sigma_{Cb}, \sigma_{Cr}, \sigma_G, \gamma_R$. Do treinamento resultaram os seguintes parâmetros ótimos para o modelo SVM: $C = 100$, $gamma = 1$ e $kernel = \text{“rbf”}$, com os quais se obteve um MCC médio de 0,579.

No segundo cenário, compôs-se o vetor somente com os descritores de movimento: $\mu_{MHI}, \mu_{GM}, \sigma_{MHI}, \sigma_{GM}$. Neste caso, os parâmetros ótimos resultantes, para o mesmo modelo, foram: $C = 100$, $gamma = 0,5$ e $kernel = \text{“rbf”}$, com os quais se obteve um MCC médio de 0,530.

Por fim, a união dos dois conjuntos anteriores foi utilizada no treinamento e otimização do modelo SVM, resultando nos seguintes parâmetros: $C = 0,5$, $gamma = 1$ e $kernel = \text{“rbf”}$, com os quais se obteve um MCC médio de 0,691.

Tabela 7 - Parâmetros ótimos do modelo SVM por cenário

Cenário	Parâmetros do modelo SVM			MCC
	C	Gamma	Kernel	
Descritores de cor	100	1,0	“rbf”	0,579
Descritores de movimento	100	0,5	“rbf”	0,530
Descritores de cor e movimento	0,5	1,0	“rbf”	0,691

5.7.2 Resultados da classificação sobre o dataset de teste

A medida de qualidade (MCC) da classificação apresentada na seção anterior pode ter se dado sob uma condição de *sobreajuste*, isto é, uma condição na qual há forte adaptação do modelo aos dados de treinamento sem que, no entanto, esse possa ser replicado em situações distintas. Essa possibilidade implica na exigência de um conjunto independente de dados para avaliação correta do modelo. Por essa razão, um novo *dataset* foi construído a partir de um grupo distinto de vídeos, análogo ao *dataset* utilizado no treinamento, porém com um número reduzido de amostras – 2.489 positivas e 2.833 negativas. É com base no resultado da classificação sobre esse novo *dataset* que se pretende, por fim, verificar a robustez do método de detecção proposto.

Os dados obtidos nesta etapa e apresentados na Tabela 8 referem-se à classificação obtida sobre o *dataset* de teste quando os parâmetros ótimos correspondentes a cada cenário

são utilizados no treinamento (ver Tabela 7). Sobre eles, são feitas as seguintes observações: comparando os valores do MCC entre as etapas de treinamento e teste, verifica-se que os descritores de movimento são mais robustos que os de cor, dado que seus respectivos cenários apresentam menor queda de desempenho. Em outras palavras: a cor parece estar mais intimamente relacionada aos dados de treinamento. Por outro lado, os descritores de cor conferem ao sistema maior precisão, isto é, possibilitam que o sistema elimine mais falsos positivos do que quando se considera somente o movimento. Este segundo cenário, por sua vez, destaca-se por uma baixa taxa de falsos negativos, evidenciada pelo melhor índice de relocação entre os três. Por fim, os resultados do último cenário indicam que a união dos dois conjuntos de descritores causa também a soma de suas qualidades: como se vê, as melhores métricas de desempenho são alcançadas quando tanto a cor quanto o movimento são empregados na classificação.

Tabela 8 - Resultado da classificação do modelo sobre o *dataset* de teste

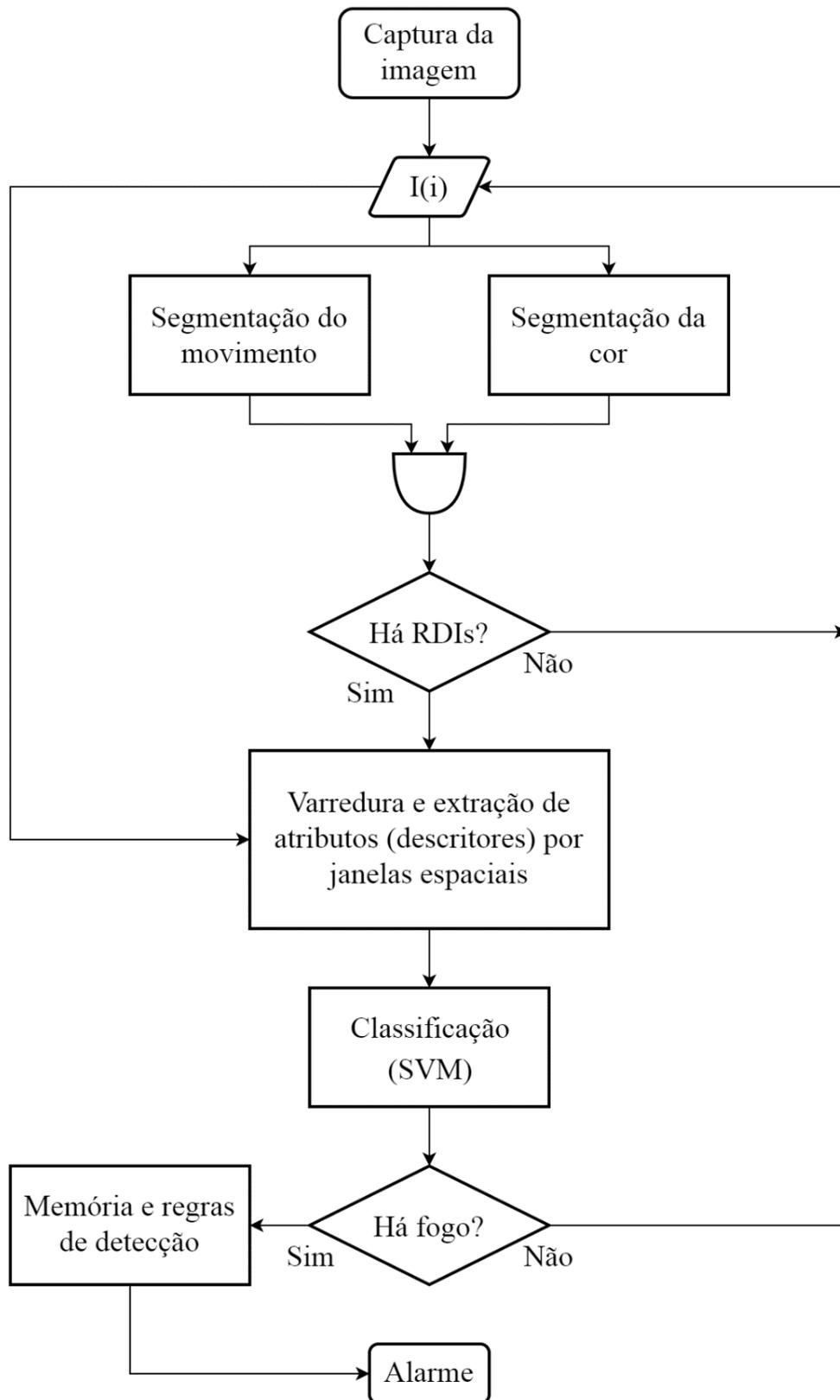
Cenário	Métricas de desempenho			
	Precisão	Relocação	F ₁	MCC
Descritores de cor	0,73	0,63	0,68	0,360
Descritores de movimento	0,69	0,81	0,75	0,413
Descritores de cor e movimento	0,75	0,77	0,76	0,472

Cada cenário foi implementado respeitando os parâmetros correspondentes apresentados na Tabela 7

5.8 PANORAMA DE FUNCIONAMENTO

Embora haja uma estrutura lógica na exposição das etapas do projeto, acredita-se que, por clareza, o algoritmo de detecção deva ser apresentado como um todo. Seu fluxograma, ilustrado na Figura 20, compreende os seguintes passos: (1) captura da imagem, onde o vídeo é lido e redimensionado; (2) obtenção das máscaras de cor e movimento por meio de segmentação; (3) intersecção das máscaras de cor e movimento - esta etapa resulta na segmentação das regiões candidatas a “fogo”; (4) se existirem regiões de interesse, a varredura por janelas espaciais realiza a extração dos atributos (descritores) da região encontrada; (5) o modelo SVM classifica o vetor de descritores extraído na etapa 4 e decide se aquela janela contém fogo ou não; (6) se houver, a posição é salva; (7) por fim, a regra de detecção é conferida: sempre que o modelo SVM retorna uma classificação positiva, a posição da ocorrência é salva; se, dentro de um intervalo de um segundo, a contagem de ocorrências for igual ou maior que cinco, então confirma-se a presença de fogo.

Figura 20 - Fluxograma



Fluxograma do algoritmo proposto

6 CONSIDERAÇÕES FINAIS

Os resultados com conclusões mais imediatas foram obtidos em etapas distintas do trabalho: a primeira, que também configura um projeto à parte, permitiu constatar os pontos fortes e fracos de uma série de métodos de segmentação por cor; a segunda, que é onde culmina todo o desenvolvimento, traz informações a respeito da utilidade de um conjunto de descritores e, por extensão, do desempenho do modelo de aprendizagem concebido a partir dele. Em paralelo, ainda podem ser feitas considerações sobre o conhecimento pessoal adquirido durante a pesquisa e sobre a utilidade prática do algoritmo completo de detecção de fogo em vídeo.

Diante da extensa lista de trabalhos acadêmicos contemplados, acompanhados de numerosas abordagens promissoras, reunir os conceitos e traçar um caminho coerente de ação não foi tarefa fácil. Quando ainda se flertava com a literatura relacionada, pouco se conhecia sobre aprendizagem de máquina. A falta de domínio da linguagem de programação foi também um impedimento recorrente. Nesse caso, a solução quase sempre era encontrada em meio à grande e solidária comunidade de programadores que dedicam seu tempo a fornecer uma luz aos mais desavisados. Espero que a conhecimento reunido neste trabalho sirva de agradecimento, ainda que nem todas as experiências tenham sido transformadas em texto.

O estudo de métodos de segmentação por cor compreende uma série de índices de desempenho. Além disso, muitas impressões a respeito do processo de implementação foram incorporadas. Essa experiência, útil na escolha do método de segmentação mais adequado às necessidades deste projeto, também pode ser de alguma utilidade a outros autores. De forma genérica, pode-se afirmar que os métodos apresentados por Phillips, Shah e Lobo (2002) e Chen, Kao e Chang (2003) configuram boas escolhas iniciais.

A segunda metade do desenvolvimento permitiu inferir sobre a validade dos descritores visuais analisados. Os de cor, que haviam sido propostos em outros trabalhos, já possuíam relevância confirmada. Os de movimento, por outro lado, foram incluídos apenas como hipóteses bem embasadas. Esperava-se verificar, desse modo, não só sua utilidade como forma única de classificação como também em conjunto com os descritores de cor. Essa foi a ideia por trás da decisão de repartir o vetor de descritores em dois. Em decorrência dos resultados obtidos, deduz-se que ambas as categorias podem fornecer bons resultados individualmente, mas que sua associação é preferível. Os dados sugerem também que não existe uma correlação entre esses grupos, visto que o desempenho destacado por seus índices é, de certo modo, complementar. A adequação de um modelo baseado em máquinas de vetores de

suporte também foi verificada, embora isso já estivesse bem consolidado na literatura. De todo modo, este trabalho figura como mais uma confirmação de sua aplicabilidade. Ainda sobre a qualidade dos atributos: é possível afirmar com segurança que o gradiente de movimento calculado a partir da imagem do histórico do movimento fornece ótimas pistas na discriminação entre o movimento ordinário e o desempenhado pelas chamas.

Na observação prática do funcionamento do algoritmo constatou-se a ocorrência de diversos falsos positivos, de acordo com o esperado: um MCC de 0,472 (valor final alcançado no conjunto de teste), embora satisfatório, sugere um grande número de falhas de classificação. Como medida de mitigação, a inclusão da regra de detecção após a etapa de reconhecimento provou-se bastante eficaz, eliminando uma boa parcela dos alarmes falsos. Confirmou-se também o funcionamento em tempo real, utilizando o hardware mencionado na seção 3.2, exceto nas situações em que o número de regiões candidatas a fogo era excessivo. Nesse caso, o alto número de classificações simultâneas reduziu a taxa de quadros significativamente.

Sobre a investigação de alternativas e complementos futuros, pode-se dizer o seguinte: como em toda aplicação de aprendizagem de máquina, o desempenho é diretamente proporcional à qualidade dos dados; neste trabalho, acredita-se que a amostragem tenha sido insuficiente. Além disso, o uso de uma única técnica de aprendizagem impossibilita julgamentos mais sólidos: ainda que os dados e descritores sejam mantidos os mesmos, é possível que o desempenho possa ser significativamente melhorado por outras abordagens, como uma baseada em redes neurais.

REFERÊNCIAS

BERTOLINI, Dener E. et al. Detecção de incêndios utilizando persistência espaço-temporal e segmentação por cor em vídeo. In: CONGRESSO BRASILEIRO DE AUTOMÁTICA, 20., 2014, Belo Horizonte. **Anais do XX Congresso Brasileiro de Automática**. Belo Horizonte: Ufmg, 2014. p. 157 - 164.

BISHOP, Christopher. **Pattern Recognition and Machine Learning**. New York: Springer-verlag, 2006. 738 p.

BOBICK, A.f.; DAVIS, J.w.. The recognition of human movement using temporal templates. **Ieee Transactions On Pattern Analysis And Machine Intelligence**, [s.l.], v. 23, n. 3, p.257-267, mar. 2001. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/34.910878>.

BORGES, Paulo Vinicius Koerich; IZQUIERDO, Ebroul. A Probabilistic Approach for Vision-Based Fire Detection in Videos. **Ieee Transactions On Circuits And Systems For Video Technology**, [s.l.], v. 20, n. 5, p.721-731, maio 2010. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/tcsvt.2010.2045813>.

BRADSKI, Gary R.; DAVIS, James W.. Motion segmentation and pose recognition with motion history gradients. **Machine Vision And Applications**, [s.l.], v. 13, n. 3, p.174-184, 1 jul. 2002. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s001380100064>.

BUKOWSKI, Richard W et al. Performance of home smoke alarms analysis of the response of several available technologies in residential fire settings. **Technical Note (nist Tn)**, [s.l.], p.1-396, 2007. National Institute of Standards and Technology. <http://dx.doi.org/10.6028/nist.tn.1455-1r2007>.

ÇELIK, Turgay et al. 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, abr. 2007. Elsevier BV. <http://dx.doi.org/10.1016/j.jvcir.2006.12.003>

ÇELIK, Turgay; DEMIREL, Hasan. Fire detection in video sequences using a generic color model. **Fire Safety Journal**, [s.l.], v. 44, n. 2, p.147-158, fev. 2009. Elsevier BV. <http://dx.doi.org/10.1016/j.firesaf.2008.05.005>.

ÇETIN, A. Enis et al. Video fire detection – Review. **Digital Signal Processing**, [s.l.], v. 23, n. 6, p.1827-1843, dez. 2013. Elsevier BV. <http://dx.doi.org/10.1016/j.dsp.2013.07.003>.

CHAMBERLIN, D.s.; ROSE, A.. The flicker of luminous flames. **Proceedings Of The Symposium On Combustion**, [s.l.], v. 1-2, p.27-32, jan. 1948. Elsevier BV. [http://dx.doi.org/10.1016/s1062-2888\(65\)80007-7](http://dx.doi.org/10.1016/s1062-2888(65)80007-7).

CHEN, Thou-ho; KAO, Cheng-liang; CHANG, Sju-ma. An intelligent real-time fire-detection method based on video processing. **Ieee 37th Annual 2003 International Carnahan**

Conference Onsecurity Technology, 2003. Proceedings., [s.l.], p.104-111, 2003. IEEE. <http://dx.doi.org/10.1109/ccst.2003.1297544>.

CHEN, Thou-ho; WU, Ping-hsueh; CHIOU, Yung-chuen. An early fire-detection method based on image processing. **2004 International Conference On Image Processing, 2004. Icip '04.**, [s.l.], p.1707-1710, 2004. IEEE. <http://dx.doi.org/10.1109/icip.2004.1421401>

CHINO, Daniel Y. T. et al. BoWFire: Detection of Fire in Still Images by Integrating Pixel Color and Texture Analysis. **2015 28th Sibgrapi Conference On Graphics, Patterns And Images**, [s.l.], p.95-102, ago. 2015. IEEE. <http://dx.doi.org/10.1109/sibgrapi.2015.19>.

COLLINS, Robert T.; LIPTON, Alan J.; KANADE, Takeo. A System for Video Surveillance and Monitoring. In: INTERNATIONAL TOPICAL MEETING ON ROBOTICS AND REMOTE SYSTEMS, 8., 1999, Pittsburgh Pennsylvania. **Proceedings of the American Nuclear Society (ANS)**. La Grange Park, Ill.: American Nuclear Society, 1999. p. 1 - 15.

DIMITROPOULOS, Kosmas; TSALAKANIDOU, Filareti; GRAMMALIDIS, Nikos. Flame Detection for Video-based Early Fire Warning Systems and 3D Visualization of Fire Propagation. **Signal Processing, Pattern Recognition And Applications / 779**: Computer Graphics and Imaging, [s.l.], 2012. ACTAPRESS. <http://dx.doi.org/10.2316/p.2012.779-011>.

FRIZZI, Sebastien et al. Convolutional neural network for video fire and smoke detection. **Iecon 2016 - 42nd Annual Conference Of The Ieee Industrial Electronics Society**, [s.l.], p.877-882, out. 2016. IEEE. <http://dx.doi.org/10.1109/iecon.2016.7793196>.

GONZALEZ, Rafael C.; WOODS, Richard E.. **Digital Image Processing**. 3. ed. Upper Saddle River, Nj, Usa: Pearson, 2007.

GÜNAY, Osman et al. Fire Detection in Video Using LMS Based Active Learning. **Fire Technology**, [s.l.], v. 46, n. 3, p.551-577, 16 set. 2009. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s10694-009-0106-8>.

HAFIANE, Adel et al. A New Supervised Evaluation Criterion for Region Based Segmentation Methods. **Advanced Concepts For Intelligent Vision Systems**, [s.l.], p.439-448, 2007. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-74607-2_40.

HEALEY, G. et al. A system for real-time fire detection. **Proceedings Of Ieee Conference On Computer Vision And Pattern Recognition**, [s.l.], p.605-606, 1993. IEEE Comput. Soc. Press. <http://dx.doi.org/10.1109/cvpr.1993.341064>.

HORNG, Wen-bing; PENG, Jian-wen; CHEN, Chih-yuan. A new image-based real-time flame detection method using color analysis. **Proceedings. 2005 Ieee Networking, Sensing And Control, 2005.**, [s.l.], p.100-105, 2005. IEEE. <http://dx.doi.org/10.1109/icnsc.2005.1461169>.

HSU, Chih-wei; CHANG, Chih-chung; LIN, Chih-jen. **A Practical Guide to Support Vector Classification**. Taipei: National Taiwan University, 2003. 16 p. Disponível em: <<https://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>>. Acesso em: 19 nov. 2019.

- INSTITUTO SPRINKLER BRASIL. **Estatísticas 2018**. 2019. Disponível em: <<http://sprinklerbrasil.org.br/instituto-sprinkler-brasil/estatisticas/estatisticas-2018/>>. Acesso em: 8 jan. 2019.
- JIANG, Bo et al. Towards a solid solution of real-time fire and flame detection. **Multimedia Tools And Applications**, [s.l.], v. 74, n. 3, p.689-705, 4 jul. 2014. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s11042-014-2106-z>.
- KJELDSEN, R.; KENDER, J.. Finding skin in color images. **Proceedings Of The Second International Conference On Automatic Face And Gesture Recognition**, [s.l.], p.312-317, 1996. IEEE Comput. Soc. Press. <http://dx.doi.org/10.1109/afgr.1996.557283>.
- KO, Byoungchul; CHEONG, Kwang-ho; NAM, Jae-yeal. Early fire detection algorithm based on irregular patterns of flames and hierarchical Bayesian Networks. **Fire Safety Journal**, [s.l.], v. 45, n. 4, p.262-270, jun. 2010. Elsevier BV. <http://dx.doi.org/10.1016/j.firesaf.2010.04.001>.
- LIU, Che-bin; AHUJA, Narendra. Vision based fire detection. **Proceedings Of The 17th International Conference On Pattern Recognition, 2004. Icpr 2004.**, [s.l.], p.134-137, 2004. IEEE. <http://dx.doi.org/10.1109/icpr.2004.1333722>.
- MARTIN, D. et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. **Proceedings Eighth Ieee International Conference On Computer Vision. Iccv 2001**, [s.l.], p.416-423, 2001. IEEE Comput. Soc. <http://dx.doi.org/10.1109/iccv.2001.937655>.
- 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, out. 1975. Elsevier BV. [http://dx.doi.org/10.1016/0005-2795\(75\)90109-9](http://dx.doi.org/10.1016/0005-2795(75)90109-9).
- MCFARLANE, N. J. B.; SCHOFIELD, C. P.. Segmentation and tracking of piglets in images. **Machine Vision And Applications**, [s.l.], v. 8, n. 3, p.187-193, maio 1995. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/bf01215814>.
- MOHRI, Mehryar; ROSTAMIZADEH, Afshin; TALWALKAR, Ameet. **Foundations of Machine Learning**. Cambridge, Ma: MIT Press, 2012. 412 p.
- PAULUZE, Thaiza. **Sistema de informações registra menos de 1% dos casos de incêndios no país**: Calcula-se que sejam 300 mil ocorrências por ano no Brasil com cerca de 1.000 mortes. 2019. Disponível em: <<https://www1.folha.uol.com.br/cotidiano/2019/02/sistema-de-informacoes-registra-menos-de-1-dos-casos-de-incendios-no-pais.shtml>>. Acesso em: 27 fev. 2019.
- PEDRINI, Hélio; SCHWARTZ, William Robson. **Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações**. São Paulo: Thomson, 2008.
- PHILLIPS, W.; SHAH, M.; LOBO, N. da Vitoria. Flame recognition in video. **Proceedings Fifth Ieee Workshop On Applications Of Computer Vision**, [s.l.], p.224-229, 2000. IEEE Comput. Soc. <http://dx.doi.org/10.1109/wacv.2000.895426>.

PHILLIPS, Walter; SHAH, Mubarak; LOBO, Niels da Vitoria. Flame recognition in video. **Pattern Recognition Letters**, [s.l.], v. 23, n. 1-3, p.319-327, jan. 2002. Elsevier BV. [http://dx.doi.org/10.1016/s0167-8655\(01\)00135-0](http://dx.doi.org/10.1016/s0167-8655(01)00135-0).

MUHAMMAD, Khan et al. Convolutional Neural Networks Based Fire Detection in Surveillance Videos. **Ieee Access**, [s.l.], v. 6, p.18174-18183, 2018. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/access.2018.2812835>.

ROSSI, L.; AKHLOUFI, M.. Dynamic Fire 3D Modeling Using a Real-Time Stereovision System. **Technological Developments In Education And Automation**, [s.l.], p.33-38, 15 dez. 2009. Springer Netherlands. http://dx.doi.org/10.1007/978-90-481-3656-8_8.

RUDZ, S et al. Investigation of a novel image segmentation method dedicated to forest fire applications. **Measurement Science And Technology**, [s.l.], v. 24, n. 7, p.075403-075403, 17 jun. 2013. IOP Publishing. <http://dx.doi.org/10.1088/0957-0233/24/7/075403>.

SANTANA, Pedro; GOMES, Pedro; BARATA, Jose. A vision-based system for early fire detection. **2012 Ieee International Conference On Systems, Man, And Cybernetics (smc)**, [s.l.], p.739-744, out. 2012. IEEE. <http://dx.doi.org/10.1109/icsmc.2012.6377815>.

STEFFENS, Cristiano Rafael; RODRIGUES, Ricardo Nagel; BOTELHO, Silvia Silva da Costa. An Unconstrained Dataset for Non-Stationary Video Based Fire Detection. **2015 12th Latin American Robotics Symposium And 2015 3rd Brazilian Symposium On Robotics (lars-sbr)**, [s.l.], p.25-30, out. 2015. IEEE. <http://dx.doi.org/10.1109/lars-sbr.2015.10>.

SWAIN, Michael J.; BALLARD, Dana H.. Indexing via color histograms. **[1990] Proceedings Third International Conference On Computer Vision**, [s.l.], p.390-393, 1990. IEEE Comput. Soc. Press. <http://dx.doi.org/10.1109/iccv.1990.139558>.

TÖREYIN, B.U.; DEDEOGLU, Y.; CETIN, A.E.. Flame detection in video using hidden Markov models. **Ieee International Conference On Image Processing 2005**, [s.l.], v. 2, p.1230-1233, 2005. IEEE. <http://dx.doi.org/10.1109/icip.2005.1530284>.

TÖREYIN, B. Uğur et al. Computer vision based method for real-time fire and flame detection. **Pattern Recognition Letters**, [s.l.], v. 27, n. 1, p.49-58, jan. 2006. Elsevier BV. <http://dx.doi.org/10.1016/j.patrec.2005.06.015>.

TOULOUSE, Tom et al. Benchmarking of wildland fire colour segmentation algorithms. **Iet Image Processing**, [s.l.], v. 9, n. 12, p.1064-1072, 1 dez. 2015. Institution of Engineering and Technology (IET). <http://dx.doi.org/10.1049/iet-ipr.2014.0935>.

WIRTH, Michael; ZAREMBA, Ryan. Flame Region Detection Based on Histogram Backprojection. **2010 Canadian Conference On Computer And Robot Vision**, [s.l.], p.167-174, 2010. IEEE. <http://dx.doi.org/10.1109/crv.2010.29>.

WREN, C.r. et al. Pfunder: real-time tracking of the human body. **Ieee Transactions On Pattern Analysis And Machine Intelligence**, [s.l.], v. 19, n. 7, p.780-785, jul. 1997. Institute of Electrical and Electronics Engineers (IEEE). <http://dx.doi.org/10.1109/34.598236>

XU, Zhengguang; XU, Jialin. Automatic Fire Smoke Detection Based on Image Visual Features. **2007 International Conference On Computational Intelligence And Security Workshops (cisw 2007)**, [s.l.], p.316-319, dez. 2007. IEEE.
<http://dx.doi.org/10.1109/cisw.2007.4425500>.

ZHAO, Jianhui et al. 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. National Library of Serbia. <http://dx.doi.org/10.2298/csis101012030z>.