

UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

SUMARIZAÇÃO AUTOMÁTICA DE CENAS  
FORENSES

ERICK VAGNER CABRAL DE LIMA BORGES

JOÃO PESSOA - PB  
Fevereiro - 2015

UNIVERSIDADE FEDERAL DA PARAÍBA  
CENTRO DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

## **SUMARIZAÇÃO AUTOMÁTICA DE CENAS FORENSES**

**ERICK VAGNER CABRAL DE LIMA BORGES**

JOÃO PESSOA - PB  
Fevereiro - 2015

**ERICK VAGNER CABRAL DE LIMA BORGES**

**SUMARIZAÇÃO AUTOMÁTICA DE CENAS FORENSES**

DISSERTAÇÃO APRESENTADA AO CENTRO DE INFORMÁTICA DA UNIVERSIDADE FEDERAL DA PARAÍBA, COMO REQUISITO PARCIAL PARA OBTENÇÃO DO TÍTULO DE MESTRE EM INFORMÁTICA (SISTEMAS DE COMPUTAÇÃO).

Orientador: Prof. Dr. Leonardo Vidal Batista

JOÃO PESSOA - PB  
Fevereiro - 2015

## CATALOGRAFIA NA BIBLIOTECA CENTRAL

## **AGRADECIMENTOS**

A Deus, por proporcionar a realização deste projeto.

Aos meus pais e avós por me darem as bases para que meus esforços fossem válidos.

Aos meus irmãos Vinicius e Victor por sempre estarem ao meu lado me incentivando.

Ao meu orientador, o professor doutor Leonardo Vidal Batista, pela sua confiança, seu apoio e seus ensinamentos durante todos estes anos de vida acadêmica.

Aos amigos do laboratório de P&D da empresa Vsoft e da UFPB pelas oportunidades e ensinamentos que me foram concedidos durante a pesquisa.

Aos amigos de pós-graduação por toda experiência que passamos juntos.

A todos que um dia me ajudaram nesta trajetória acadêmica.



## **RESUMO**

A presença crescente de dispositivos de gravação de vídeo nas mais diversas áreas vêm proporcionando um aumento no uso destas imagens principalmente para fins investigativos. Isto faz com que a utilização de métodos e ferramentas que realizem a análise e o monitoramento automático de ambientes seja cada vez mais necessária para dar suporte técnico e de conhecimento aos investigadores, possibilitando que os resultados alcançados sejam os mais eficientes e eficazes possíveis. Este trabalho descreve o desenvolvimento de métodos de visão computacional que têm como objetivo extrair aspectos relevantes de cenas – imagens individuais, ou quadros ou sequências de quadros de vídeo - e utilizar a informação obtida com o propósito de sumarização. Os métodos propostos visam a detectar e analisar movimentação, detectar faces classificando-as por gênero, efetuar reconhecimento de faces, realizar o rastreamento de faces humanas e reconhecer a cor predominante no vestuário de indivíduos. O sistema desenvolvido efetua a extração de informações relevantes, o que auxilia na redução do tempo necessário à inspeção por seres humanos, na interpretação e argumentação de casos e na documentação dos casos. Ao fim do trabalho, os métodos desenvolvidos apresentaram resultados compatíveis com os da literatura.

Palavras-Chave: Computação forense, Sumarização, Visão computacional.

## **ABSTRACT**

The growing presence of video recording devices in several areas are providing an increase in use these images mainly to investigative purposes. This makes the use of methods and tools that perform the analysis and the automatic monitoring of environments are increasingly needed to provide technical support and knowledge to investigators, enabling obtain efficient and effective results. This work describe the development of computational vision methods that aim extract some features of scenes. At the end of this extraction, a summarization tool of forensic scenes through the developed methods is proposed. The methods proposed aim to detect and analyze motion in scenes, detect faces classifying them through the gender recognition, recognize people through facial recognition, perform the tracking of human faces and pattern recognition of predominant color in the clothing of individuals. At the end of this work, developed methods presented results comparable to the ones found in the literature and may contribute to the fast extraction of information needed for human analysis, to assist in the interpretation and argumentation of cases and documenting the results.

Key-words: Forensic computing, Video summarization, Computational vision.

## LISTA DE FIGURAS

Figura 1. Visão geral da proposta .....	19
Figura 2. Ilustração do cone que as cores são representadas no modelo HSV.....	21
Figura 3. Ilustração do cubo que as cores são representadas no modelo RGB. ....	22
Figura 4. Operador LBP básico .....	23
Figura 5. Operador $LBP_{P,R}$ com vizinhanças (a) (8, 1), (b) (16, 2) e (c) (8, 2) .....	24
Figura 6. LBP com três níveis de informação: código LBP do pixel, histograma local de cada região e histograma concatenado.....	24
Figura 7. Face com pontos de referência do ASM .....	25
Figura 8. Ilustração do fluxo utilizado no reconhecimento de faces. ....	29
Figura 9. Exemplo de Transformação não-linear .....	30
Figura 10. <i>Screenshot</i> da ferramenta proposta por Thornton <i>et al.</i> (2011).....	33
Figura 11. Banco de dados Yale A: pessoa 1 em suas 11 poses.....	35
Figura 12. Banco de dados Extended Yale B .....	36
Figura 13. Amostras do banco de dados GTAV .....	37
Figura 14. Amostras do banco de faces Vsoft-Faces.....	37
Figura 15. Uma sequência de quadros do vídeo gerado para testes .....	38
Figura 16. Amostras do banco FERET utilizadas no trabalho proposto.....	39
Figura 17. Amostra de quadros dos vídeos <i>indoor sequence</i> e <i>outdoor sequence</i> . ....	39
Figura 18. Amostra do banco Dataset 2012 (Goyette <i>et al.</i> , 2012) que mostra a imagem de um quadro e seu respectivo gabarito de movimentação.....	40
Figura 19. Ilustração da detecção de objetos em movimento.....	42
Figura 20. Ilustração de um quadro que possui objeto com características procuradas .	43
Figura 21. Ilustração do quadro com maior movimentação e do gráfico de movimentação. Em vermelho estão destacadas as áreas que sofreram alteração em relação ao <i>background</i> . ....	44
Figura 22. Comparação entre movimentação gerada pelo método proposto e gabarito de movimentação do banco de dados (2012 <i>Dataset</i> ) (Goyette <i>et al.</i> , 2012). ....	44
Figura 23. Ilustração do reconhecimento de cor predominante no vestuário.....	47
Figura 24. Ilustração do ângulo e da distância entre os olhos. (a) Imagem original (b) Imagem normalizada geometricamente .....	48

Figura 25. a) Imagem original Métodos de normalização fotométrica: b) Equalização de histograma .....	49
Figura 26. Histogramas de Padrão Binário Local.....	49
Figura 27. Fluxo para o reconhecimento de faces e gênero .....	51
Figura 28. Visão geral da abordagem proposta .....	52
Figura 29. Captura de tela do sistema desenvolvido .....	53
Figura 30. Quadro original, quadro processado pelo método e que possui ruídos e presença de falsos positivos e quadro de gabarito, respectivamente. ....	56
Figura 31. Imagens normalizadas geometricamente pelos pontos extraídos do ASM ...	57
Figura 32. Imagens normalizadas geometricamente pelos pontos extraídos do ASM ...	63
Figura 33. Imagens normalizadas geometricamente pelos pontos extraídos manualmente .....	63
Figura 34. Ilustração da falha no método de rastreamento .....	67
Figura 35. Pós-processamento para segmentação de objetos em quadros de movimentação.....	68
Figura 36. Transformada de Hough sobre região dos olhos.....	70
Figura 37. Padrões encontrados após aplicação de filtros de pré-processamento .....	70
Figura 38. Exemplo onde o método encontra 3 círculos e verifica qual é mais predominantemente branco .....	71

## LISTA DE TABELAS

Tabela 1. Tabela construída para atribuição de cor no método de reconhecimento de cor predominante no vestuário de um indivíduo. ....	46
Tabela 2. Resultados das métricas para categoria <i>baseline</i> do banco de dados <i>Change Detection Benchmark Dataset (2012 Dataset)</i> (Goyette <i>et al.</i> , 2012).....	55
Tabela 3. Reconhecimento de cor predominante no vestuário.....	57
Tabela 4. Taxas de acerto para o método de classificação de gênero.....	58
Tabela 5. Comparação entre as taxas de acerto para o método de classificação de gênero com normalização geométrica por marcação manual e normalização geométrica por pontos do ASM. ....	59
Tabela 6. Taxas de acerto para o método de classificação de gênero um-para-um com normalização geométrica por marcação manual e com comparação entre todas as imagens.....	60
Tabela 7. EER para reconhecimento facial (1x1) com detecção utilizando pontos do ASM.....	61
Tabela 8. Taxa de acerto do reconhecimento facial com apenas uma imagem de teste por indivíduo .....	61
Tabela 9. Taxa de acerto do reconhecimento facial com apenas uma imagem de treinamento por indivíduo .....	62
Tabela 10. Reconhecimento facial com etapa de normalização geométrica baseada em marcação manual dos olhos.....	64
Tabela 11. Taxa de acerto par reconhecimento facial em vídeo .....	64
Tabela 12. Rastreamento de faces x faces isoladas .....	65
Tabela 13. Taxa de acertos para reidentificação de faces no rastreamento.....	66
Tabela 14. Tabela comparativa entre o ASM e o método proposto em desenvolvimento .....	71

## LISTA DE EQUAÇÕES

(1) Conversão RGB para HSV .....	22
(2) Equação do $LBP_{p,r}$ .....	23
(3) Sensibilidade.....	26
(4) Especificidade.....	26
(5) Taxa de falsos positivos.....	26
(6) Taxa de falsos negativos .....	26
(7) Porcentagem de classificações erradas.....	27
(8) Precisão.....	27
(9) Medida-F.....	27
(10) Função de distribuição acumulativa....	29
(11) Probabilidade no rastreamento .....	30
(12) Subtração do background.....	41
(13) Função de atualização do background.....	41
(14) Medida da quantidade de movimento.....	43
(15) Taxa de acerto.....	47
(16) Fator de redimensionamento para normalização geométrica.....	48
(17) Correlação.....	50

## LISTA DE ABREVIATURAS E SIGLAS

ASM	<i>Active Shape Models</i> (Modelos de Forma Ativa)
CDF	<i>Cumulative Distribution Function</i> ( Função de Distribuição Acumulada)
EER	<i>Equal Error Rate</i> (Taxa de Erro Igual)
ESP	Especificidade
FAR	<i>False Acceptance Rate</i> ( Taxa de Falsa Aceitação)
FN	<i>False Negative</i> ( Falso Negativo)
FNR	<i>False Negative Rate</i> ( Taxa de Falsos Negativos)
FP	<i>False Positive</i> ( Falso Positivo)
FPR	<i>False Positive Rate</i> ( Taxa de Falsos Positivos)
FRR	<i>False Rejection Rate</i> ( Taxa de Falsa Rejeição)
ICA	<i>Independent Component Analysis</i> ( Análise de Componentes Independentes)
IPC	Instituto de Polícia Científica
KNN	<i>K-Nearest Neighbor</i> ( K-vizinhos mais próximos)
LBP	<i>Local Binary Patterns</i> ( Padrão Binário Local)
LBPH	<i>Local Binary Patterns Histogram</i> ( Histograma de Padrões Binários Locais)
NC	<i>Normalized Correlation</i> ( Correlação Normalizada)
OPENCV	<i>Open Computer Vision Library</i>
ORL	<i>Olivetti Research Lab</i>
PCA	<i>Principal Component Analysis</i> ( Análise de Componentes Principais)
PDI	Processamento Digital de Imagens
PNG	<i>Portable Network Graphics</i>
Prec	Precisão
PWC	<i>Percentage of Wrong Classifications</i> ( Porcentagem de Classificações Erradas)
RE	<i>Recall</i> ( Revocação)
TN	<i>True Negative</i> ( Verdadeiro Negativo)
TP	<i>True Positive</i> ( Verdadeiro Positivo)

## SUMÁRIO

<b>1. Introdução .....</b>	<b>16</b>
1.1 Objetivos e Contribuições .....	17
1.2 Visão Geral do Sistema Proposto .....	19
1.3 Estrutura da Dissertação .....	19
<b>2. Fundamentação Teórica.....</b>	<b>20</b>
2.1 Ciência Forense e Computação Forense .....	20
2.2 Processamento Digital de Imagens .....	20
2.2.1 Modelo de Representação de Cores .....	21
2.3 Visão Computacional.....	22
2.4 Padrões Binários Locais.....	23
2.5 Modelos de Forma Ativa ( Active Shape Models , ASM) .....	24
2.6 Detecção e Análise de Movimentação .....	25
2.7 Biometria.....	27
2.8 Detecção e Reconhecimento de Faces .....	27
2.8.1 Normalização da Face.....	29
2.8.2 Rastreamento de Faces.....	30
<b>3. Trabalhos Correlatos .....</b>	<b>32</b>
<b>4. Materiais e Métodos .....</b>	<b>35</b>
4.1 Bancos de dados utilizados .....	35
4.2 Método de Detecção e Análise de Movimentação.....	40
4.3 Extração de Características a Partir de uma Face .....	45
4.3.1 Reconhecimento da Cor Predominante no Vestuário de um Indivíduo .....	45
4.3.2 Reconhecimento de Faces e de Gênero.....	47
4.3.3 Rastreamento de Faces.....	51
4.4 Sumarização de Cenas Através dos Métodos Propostos .....	52
<b>5. Resultados e Discussão .....</b>	<b>55</b>

5.1	Detecção e Análise de Movimentação .....	55
5.2	Extração de Características a partir da face .....	56
<b>6.</b>	<b>Considerações Finais e Trabalhos futuros.....</b>	<b>68</b>
	<b>Referências .....</b>	<b>73</b>

## 1. Introdução

Cada vez mais comuns em setores públicos e privados, dispositivos e sistemas de gravação de vídeo frequentemente são utilizados em análises e investigações criminais. A habilidade para obter informações destes vídeos tem um grande potencial para auxiliar investigações, ajudar a solucionar possíveis questionamentos que possam surgir quando é necessário analisar e resolver casos em questões cíveis, criminais e de comportamento social.

A aplicação de um amplo espectro dos conhecimentos e técnicas, que podem abranger diversos ramos da ciência, utilizados para obter informações de um sistema legal, é denominada Ciência Forense (Andaló, 2012). Um ramo da ciência que pode contribuir para que os questionamentos existentes possam ser solucionados é o da Ciência da Computação. O domínio emergente que une os campos de Ciência Forense e Ciência da Computação é denominado Computação Forense.

Segundo Delp *et al.* (2009), podemos definir a Computação Forense como o conjunto de técnicas científicas para a coleta, validação, identificação, análise, interpretação, documentação e preservação de evidências derivadas de fontes digitais com a proposta de facilitar ou promover a reconstrução de eventos, na maioria das vezes, de natureza criminal.

Através dos métodos computacionais, a Computação Forense pode se relacionar com a Ciência Forense por três caminhos (Franke *et al.*, 2008):

- Através do fornecimento de ferramentas para melhorar a análise dos vestígios, superando algumas limitações da habilidade cognitiva humana. Também pode ser utilizada como auxílio ao examinador forense;

- Através do fornecimento de bases científicas para disciplinas ou procedimentos forenses que necessitam da análise de grande volume de dados, tarefa que pode ser impossível utilizando exclusivamente recursos humanos;

- Através da representação do conhecimento de especialistas pela implementação, em máquina, de suas habilidades de reconhecimento e raciocínio.

O principal desafio atrelado à Computação Forense é que as provas muitas vezes estão escondidas em um ambiente caótico, fazendo com que a dificuldade de extrair a informação seja maior do que quando se tem um ambiente controlado. Outro desafio é que os vídeos de dispositivos de circuitos fechados são frequentemente de baixa qualidade, o que se deve principalmente a questões práticas ou econômicas. A extração

adequada de informações em um vídeo depende de sua qualidade, e é importante que se planeje uma melhoria na qualidade de gravação destes. A análise dos dados torna-se um grande desafio quando feita manualmente, então, a criação de uma ferramenta automática pode tornar o trabalho dos examinadores humanos viável. A extração das características em uma cena forense pode ser realizada ou auxiliada por métodos computacionais de processamento digital de imagens, análise de imagens, visão computacional<sup>1</sup>, computação gráfica, visualização de dados, reconhecimento de padrões, aprendizagem de máquina, mineração de dados, robótica etc.

Propõe-se a pesquisar, desenvolver e avaliar métodos de extração de informações relevantes de cenas forenses, realizando de forma antecipada e automatizada algumas tarefas que seriam realizadas pelo examinador. Os métodos abordados neste trabalho objetivam extrair informações associadas à face humana, como o gênero e a identidade, à detecção e análise de movimentação, à cor predominante no vestuário de um indivíduo.

## **1.1 Objetivos e Contribuições**

O objetivo desta dissertação foi pesquisar, desenvolver e avaliar métodos, baseados em Visão Computacional, que possam realizar a análise de vídeos ou imagens, extraindo características das informações visuais com o intuito de auxiliar procedimentos investigativos na Ciência Forense ou, de forma mais genérica, o levantamento de dados acerca de um ambiente. Os métodos desenvolvidos estão associados aos seguintes temas

- Detecção e análise de movimentação em vídeo: compara quadros sequenciais distintos e detecta quando ocorrem diferenças consideradas substanciais, de acordo com algum critério. Além disto, o método analisa e extrai informações a partir das alterações encontradas, como tamanho e cores dos objetos que se movimentam e estatísticas de movimentação nos quadros. A especificação exclusiva das partes em que houve movimentação ou mudança reduz a quantidade de informação que precisa ser examinada.

---

<sup>1</sup> O termo 'visão computacional' será utilizado ao longo desta dissertação para designar a análise automática por computador de informações extraídas de uma cena (Marques Filho e Vieira Neto, 1999).

- Extração de Características a partir de Faces: extrai informações relevantes de uma cena, a partir da imagem da face. Para isto, alguns métodos foram desenvolvidos.
  - Detecção de faces: está ligada à determinação da presença e posição espacial de cada face existente em uma cena, sendo esta etapa utilizada como a que antecede a extração de características.
  - Reconhecimento da cor predominante no vestuário de um indivíduo: sabendo-se a posição espacial da face na cena, utiliza-se de uma estimativa da posição do corpo do indivíduo para extrair informação de cor predominante no vestuário, para isto, utiliza-se a transformação da imagem para o sistema de cor HSV extraindo apenas a informação de cor.
  - Reconhecimento de Faces: consiste no processo de identificar um indivíduo analisando a imagem de sua face e comparando-a com outras imagens de faces pré-rotuladas. Geralmente, este processo é dividido em extração de características e classificação. Para extrair as características da face, foi utilizado o algoritmo de Padrão Binário Local ( *Local Binary Patterns* , LBP) e posteriormente, na etapa de classificação, foi escolhida a métrica de correlação para comparar os LBP's gerados.
  - Reconhecimento de Gênero: uma característica importante para diversos fins, como investigações e observação de pessoas, e que pode ser fornecida pelo rosto humano. Propõe-se um método que determine o gênero de um indivíduo através da imagem de sua face e do uso de LBP.
  - Rastreamento de Faces: está ligado a rastrear o indivíduo ao longo do tempo em que este permanece na cena. Apresenta grande importância tanto para diferenciar múltiplos agentes, quanto para reidentificar uma face previamente identificada.

Através dos métodos implementados, uma sumarização de cenas foi proposta, onde o examinador pode escolher o que deseja extrair de uma cena, dispensando dados que ele considera desnecessários.

O sistema desenvolvido permite:

- Sumarizar cenas separando apenas o que é importante para análise do examinador, sendo útil, por exemplo, para possibilitar que resultados de uma investigação possam ser obtidos mais rapidamente;
- Pesquisar grandes volumes de dados de forma eficiente, podendo reduzir custos e carga horária dos examinadores;
- Visualizar e documentar os resultados da análise;
- Auxiliar na interpretação dos resultados e sua argumentação;

Destaca-se que o trabalho proposto despertou o interesse de empresas públicas, inclusive do Instituto de Polícia Criminal (IPC) do Estado da Paraíba, e também da empresa privada Vsoft Tecnologia Ltda., empresa local especializada em segurança, empregando o trabalho em seu setor de pesquisa e desenvolvimento e auxiliando com toda infraestrutura necessária.

## 1.2 Visão Geral do Sistema Proposto

Uma visão geral do sistema proposto neste trabalho pode ser observada na Figura 1. Ao receber uma cena a ser examinada, uma sumarização é aplicada através dos métodos adotados e parâmetros escolhidos, resultando em um conjunto de dados sumarizados da cena, que auxiliam no processo de decisão.



**Figura 1.** Visão geral da proposta

## 1.3 Estrutura da Dissertação

Este capítulo introduziu o escopo de pesquisa e apresentou os objetivos desta dissertação. O Capítulo 2 apresenta a Fundamentação Teórica necessária para embasar o trabalho, incluindo as definições gerais abordadas. No Capítulo 3 são elencados os trabalhos correlatos. No Capítulo 4 serão descritos todos os materiais utilizados e o desenvolvimento dos métodos propostos. No Capítulo 5 são mostrados os resultados obtidos seguidos de uma discussão sobre eles. O Capítulo 6 demonstra as conclusões finais desta dissertação, assim como os trabalhos futuros.

## 2. Fundamentação Teórica

Neste capítulo são apresentados os conceitos gerais necessários para o entendimento do trabalho proposto.

### 2.1 Ciência Forense e Computação Forense

Ciência Forense é a aplicação metodológica de um amplo espectro de disciplinas científicas para responder perguntas importantes relacionadas ao sistema legal, ou seja, consiste em abordagens multidisciplinares para realizar tarefas como investigar e reconstruir cenas de crime ou acidente, coletar e investigar evidências, identificar, classificar, quantificar, individualizar pessoas, objetos e processos, estabelecer vínculos, associações e reconstruções, e por fim, usar essas informações para chegar às conclusões que auxiliem os processos (Franke *et al.*, 2007).

Computação Forense pode ser entendida como o domínio que une os campos de Ciência da Computação e Ciência Forense, e engloba estudo de métodos de investigação, mediante o uso de computadores e métodos computacionais. Os métodos computacionais podem assistir em uma pesquisa básica e aplicada, apoiar o examinador em tratamento de casos diários e também fazer com que a união do conhecimento e inteligência dos seres humanos possa ser somado com estes métodos, formando uma espécie de inteligência híbrida (Franke *et al.*, 2008). Na prática, a Computação Forense pode ser entendida como o uso de métodos científicos para preservação, coleta, restauração, identificação, documentação e apresentação de evidências digitais.

### 2.2 Processamento Digital de Imagens

Uma imagem pode ser definida como uma função bidimensional  $f(x, y)$ , que mapeia coordenadas nos eixos  $x$  e  $y$  para determinado valor, chamado de intensidade, proporcional ao brilho da imagem naquele ponto. Quando estes valores de intensidade são finitos e discretos, dizemos que a imagem é digital. Uma imagem digital é composta por uma quantidade finita de elementos, cada um com localização e valor próprio. Estes elementos são mais conhecidos como *pixels* (Gonzalez e Woods, 1992).

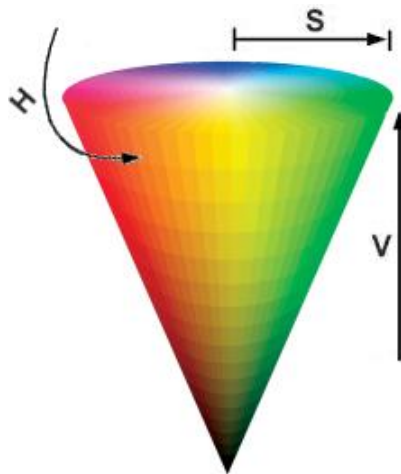
O Processamento Digital de Imagens (PDI) é o campo da Ciência da Computação que consiste no aprimoramento de informações pictóricas para interpretação humana (Marques Filho e Vieira Neto, 1999). Ao contrário do tratamento

de imagens, que preocupa-se somente com a manipulação de figuras para sua representação final, o PDI pode ser utilizado para novos processamentos de dados tais como aprendizagem de máquina ou reconhecimento de padrões. Estas técnicas de processamento são utilizadas para resolver vários problemas, melhorando a informação visual, para a interpretação humana e para o processamento de dados de cena para percepção automática através das máquinas.

### 2.2.1 Modelo de Representação de Cores

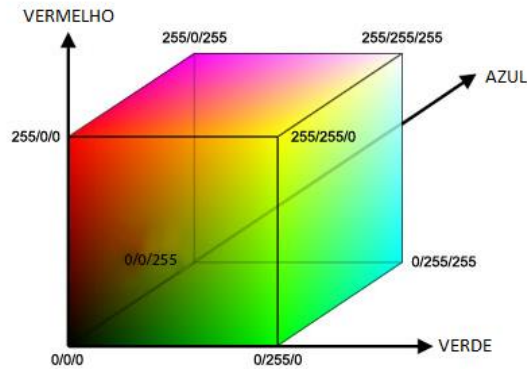
O objetivo dos modelos de cores é permitir a especificação de cores em um formato padronizado e aceito por todos. Em linhas gerais, um modelo de cores é uma representação tridimensional na qual cada cor é representada por um ponto no sistema de coordenadas 3-D (Marques Filho e Vieira Neto, 1999).

O modelo HSV (Alvy, 1978) é um modelo perceptual que representa as cores em matiz (*hue*, H), que pode ser vista como a tonalidade dominante da cor, saturação (*saturation*, S), que pode ser considerado como o grau de pureza da cor quanto à adição de branco (da esmaecida a intensa), e valor (*value*, V), o brilho da cor (mais clara ou mais escura). Estas cores estão representadas no espaço em um formato de cone, como mostra a Figura 2.



**Figura 2.** Ilustração do cone que as cores são representadas no modelo HSV.  
fonte: Adaptado de (Costa, 2013)

O modelo RGB é o mais utilizado em câmeras e monitores de vídeo. Este modelo é baseado em um sistema de coordenadas cartesianas, que pode ser visto como um cubo onde três de seus vértices são as cores primárias, outros três as cores secundárias, o vértice junto à origem é o preto e o mais afastado da origem corresponde à cor branca, conforme ilustra a Figura 3.



**Figura 3.** Ilustração do cubo que as cores são representadas no modelo RGB.

Caso seja necessário converter valores do modelo RGB para o HSV, pode-se utilizar a conversão descrita abaixo (Koschan e Abidi, 2008):

$$\begin{aligned}
 V &= \max(R, G, B) \\
 S &= \begin{cases} \frac{\max(R,G,B) - \min(R,G,B)}{\max(R,G,B)} & , \max(R, G, B) \neq 0 \\ 0 & , \max(R, G, B) = 0 \end{cases} \quad (1) \\
 H &= a \cos \frac{0,5 ((R - G) + (R - B))}{\sqrt{(R - G)(R - G) - (R - G)(G - B)}}
 \end{aligned}$$

### 2.3 Visão Computacional

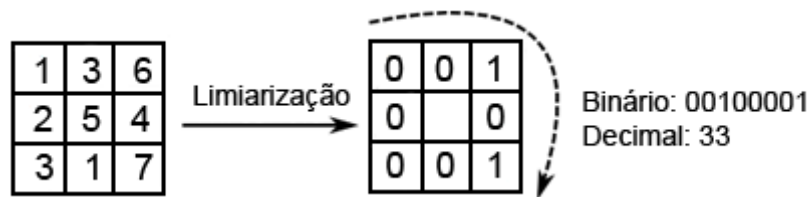
Visão Computacional pode ser descrita como o conjunto de métodos e técnicas através dos quais sistemas computacionais podem ser capazes de interpretar imagens. Inclui métodos para aquisição, restauração, segmentação e extração de atributos ou características para produzir informações. As técnicas empregadas podem ser utilizadas para permitir reconhecer, manipular e pensar sobre os objetos que compõem uma imagem.

A interpretação de uma imagem pode ser definida em termos computacionais como a transformação de um conjunto de dados digitais representados em uma imagem, em uma estrutura de dados descrevendo a semântica deste conjunto de dados em um contexto qualquer.

## 2.4 Padrões Binários Locais

O operador de análise de textura denominado Padrões Binários Locais (*Local Binary Patterns, LPB*) foi introduzido por Ojala *et al.* (1996) e pode ser definido como uma medida de textura invariante a escala de cinza, sendo considerado um poderoso meio de descrição de texturas 2D, devido a seu poder de discriminação, simplicidade computacional e tolerância contra mudanças monotônicas em escala de cinza (Ylioinas *et al.*, 2011).

A ideia básica do LBP é resumir a estrutura local em uma imagem por meio da comparação de cada pixel com os seus vizinhos. Inicialmente, o LBP foi proposto através do deslocamento de uma janela 3x3 na imagem onde para cada um dos oito pixels vizinhos ao pixel central, é aplicada uma limiarização com relação ao valor de intensidade do pixel central, caso o valor do vizinho seja maior, atribui-se 1, caso contrário atribui-se 0. Ao término da varredura, cada pixel terá um valor binário formando um dentre  $2^8 = 256$  possíveis resultados. A Figura 4 ilustra este processo.

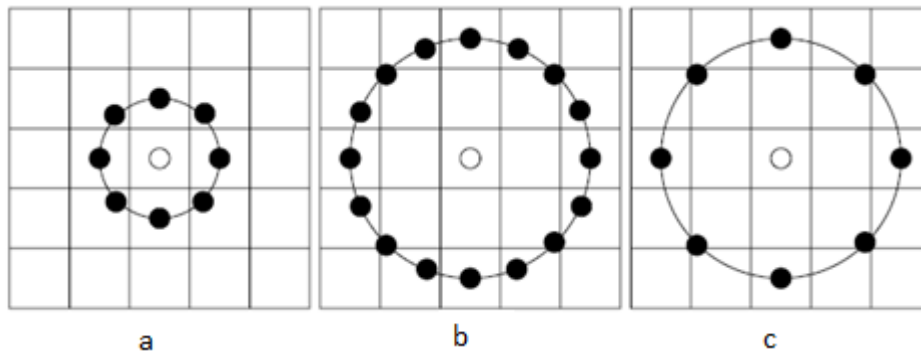


**Figura 4.** Operador LBP básico

Posteriormente, o operador foi estendido para utilizar vizinhos de tamanhos diferentes. Para isto, foi utilizada a notação  $LBP_{P,R}$ , que representa uma vizinhança de  $P$  pontos igualmente espaçados e com distância  $R$  do centro. O cálculo dos códigos LBP pode ser dado pela Equação 2.

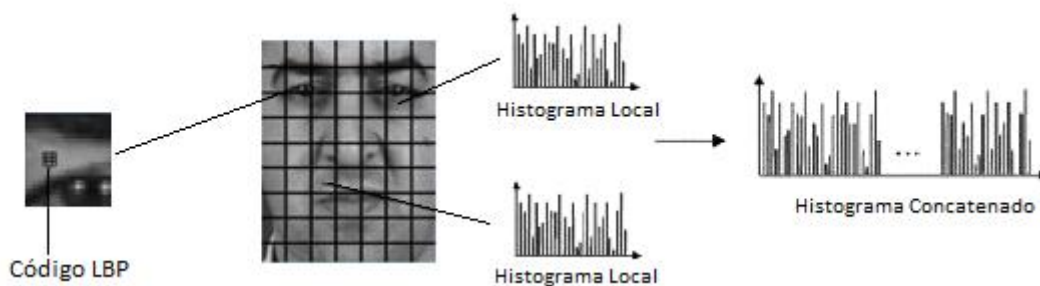
$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c) 2^p \quad (2)$$

Onde  $g_c$  corresponde ao nível de cinza do *pixel*  $(x_c, y_c)$ ,  $g_p$  se refere ao nível de cinza de  $P$  *pixels* igualmente espaçados num círculo de raio  $R$ , e  $s$  define uma função de limiarização que atribui 1 se  $x > 0$  e atribui 0, caso contrário. Na Figura 5 é exemplificado como a vizinhança pode ser determinada para um determinado conjunto de valores de  $P$  e  $R$ .



**Figura 5.** Operador  $LBP_{P,R}$  com vizinhanças (a) (8, 1), (b) (16, 2) e (c) (8, 2)  
 Fonte: (Pietikäinen *et al.*, 2011)

Com o objetivo de incorporar informação espacial no LBP, Ahonen *et al.* (2004) propuseram uma variação no LBP. Para isto, a imagem é dividida em R pequenas regiões locais de blocos não sobrepostos e de mesmo tamanho. Posteriormente, o Histograma de códigos LBP  $H^r$ , onde  $r \in \{1,2,\dots,R\}$ , é calculado para cada bloco e concatenado em um único histograma que representará a face. A Figura 6 ilustra todo este processo.



**Figura 6.** LBP com três níveis de informação: código LBP do pixel, histograma local de cada região e histograma concatenado.  
 Fonte: Adaptado de (Rodríguez 2006)

## 2.5 Modelos de Forma Ativa ( *Active Shape Models* , ASM)

*Active Shape Models* (ASM) é uma popular ferramenta estatística para localizar exemplos de objetos conhecidos em uma imagem e foi proposto inicialmente por Cootes *et al.* (1995).

O ASM pode ser utilizado para detectar a posição da face e das suas características peculiares, como olhos, bocas, etc. em uma imagem. Para isto, ele se vale do modelo prévio, com típicas variações da face, para tentar encontrar a melhor combinação do modelo de forma que se assemelhe com uma variação de face

(Rodriguez, 2006). O modelo de forma da face é representado por pontos de referência (Figura 7).



**Figura 7.** Face com pontos de referência do ASM

## 2.6 Detecção e Análise de Movimentação

Detecção de movimentação é um importante problema em Visão computacional e é comumente utilizada em etapas iniciais para diversas aplicações relacionadas a vigilância, contagem de objetos, monitoramento de tráfego, reconhecimento de linguagem gestual, entre outros (Lu *et al.*, 2008). Esse método tem como objetivo relatar as diferenças entre quadros de um vídeo, pois geralmente em vídeos de câmeras de vigilância, uma grande porção da área das imagens permanece inalterada ao longo de uma sequência.

A detecção e análise de movimentação auxilia o examinador a reduzir a quantidade de informação à ser examinada, reduzindo de forma significativa a quantidade de dados a processar, assim como, concentra-o nas mudanças da cena que podem ser resultado da atividade de um indivíduo suspeito.

Existem diversas abordagens para detecção de movimento, desde abordagens mais comuns, a abordagens mais robustas como as que utilizam modelos gaussianos para extrair movimentação através da informação de estrutura e de cor (Stauffer e Grimson, 1999) (Ren *et al.*, 2001) (Chakraborty e Fuentes, 2009). Alguns autores (Lu *et al.*, 2008)(Huwer e Niemann, 2000) buscam uma abordagem híbrida para detectar regiões de alteração. Em geral, existem três abordagens representativas para o

problema: diferença temporal, subtração do plano de fundo e o fluxo ótico (Lu *et al.*, 2008).

Para efeito comparativo com outros métodos de detecção de movimentação presentes na literatura, algumas métricas comumente são utilizadas. Estas métricas baseiam-se nos seguintes conceitos:

- verdadeiro positivo (*True Positive*, TP): número de identificações corretas, ou seja, é identificada movimentação no *pixel* quando a movimentação está presente no respectivo *pixel* na imagem de gabarito.
- falso positivo (*False Positive*, FP): número de identificações incorretas, ou seja, é identificada movimentação no *pixel* quando a movimentação não está presente no respectivo *pixel* na imagem de gabarito.
- falsos negativos (*False Negative*, FN): número de rejeições incorretas, ou seja, não é identificada movimentação no *pixel* quando a movimentação está presente no respectivo *pixel* na imagem de gabarito.
- verdadeiros negativos (*True Negative*, TN) - número de rejeições corretas, ou seja, não é identificada movimentação no *pixel* quando a movimentação não está presente no respectivo *pixel* na imagem de gabarito.

As métricas geradas são valores de especificidade (Esp), sensibilidade ou revocação (*recall*, Re), taxa de falsos positivos (*False Positive Rate*, FPR), taxa de falsos negativos (*False Negative Rate*, FNR), porcentagem de classificações erradas (*Percentage of Wrong Classifications*, PWC), precisão (Prec) e medida F (*F-Measure*). As equações abaixo demonstram como essas métricas podem ser calculadas (Goyette *et al.*, 2012).

$$\text{Re} = \frac{\text{TP}}{(\text{TP} + \text{FN})} \quad (3)$$

$$\text{Esp} = \frac{\text{TN}}{(\text{TN} + \text{FP})} \quad (4)$$

$$\text{FPR} = \frac{\text{FP}}{(\text{FP} + \text{TN})} \quad (5)$$

$$\text{FNR} = \frac{\text{FN}}{(\text{TP} + \text{FN})} \quad (6)$$

$$PWC = \frac{100 * (FN + FP)}{(TP + FN + TN + FP)} \quad (7)$$

$$Prec = \frac{TP}{(TP + FP)} \quad (8)$$

$$Medida F = \frac{2 * Prec * Re}{(Prec + Re)} \quad (9)$$

## 2.7 Biometria

A biometria é um ramo da ciência associada ao estudo estatístico de características comportamentais ou físicas dos seres vivos. Este termo tem sido aplicado em um contexto mais específico para representar o estudo das características físicas e comportamentais de um indivíduo visando identificá-lo unicamente. Estas características analisadas podem ser relacionadas com características físicas (retina, íris, face, geometria da palma da mão, impressão digital, etc.) ou com características comportamentais (assinatura manuscrita, voz, maneira de andar, etc.) (Kazienko, 2003).

A utilização da biometria é cada vez mais comum em situações que exigem a autenticação, uma vez que os identificadores biométricos não podem ser facilmente extraviados, forjados, ou compartilhados. Métodos de identificação biométricos são considerados mais confiáveis do que métodos baseados em *tokens* (como *smart cards*) ou senhas (Maltoni *et al.*, 2009).

O reconhecimento do usuário em sistemas biométricos pode ser abordado de duas maneiras: verificação ou identificação. A verificação é uma abordagem dita 1-1 (um para um), exige menos processamento e é mais utilizada em aplicações que atrelam a informação de uma característica biométrica ao uso de *logins* por parte do usuário. A identificação é uma abordagem dita 1-N (um para muitos), pode ser feita com todos os dados armazenados e no fim, pode ser escolhido o considerado dado idêntico ao apresentado inicialmente, levando-se em consideração uma margem de erro (Alves, 2007).

## 2.8 Detecção e Reconhecimento de Faces

Detecção e Reconhecimento de faces são dois conceitos distintos, sendo importante diferenciá-los. O conceito de detecção de faces está ligado à determinação da presença e posição espacial de cada face existente em uma imagem (Omaia, 2009). Geralmente, esta etapa é utilizada como a que antecede o reconhecimento. O

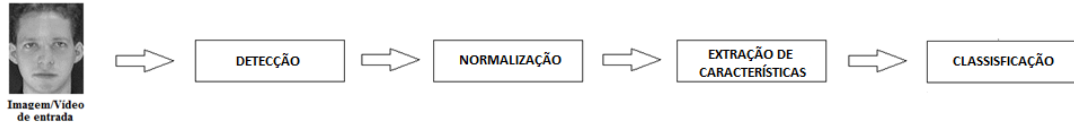
reconhecimento de faces consiste no processo de identificar um indivíduo analisando sua imagem de face e comparando-a com outras imagens de faces pré-rotuladas.

O processo de reconhecimento de faces pode ser abordado de duas maneiras: identificação e verificação (Abate *et al.*, 2007). Na identificação, a imagem de face de uma pessoa é comparada a uma galeria de imagens de faces tendo como objetivo determinar a pessoa correspondente à face investigada. Na verificação, a imagem de face investigada e a suposta pessoa que ela representa são previamente informados e o objetivo é verificar a informação declarada, comparando a imagem de face investigada com uma galeria de imagens de faces para confirmar que a mesma corresponde à pessoa declarada (Matos, 2008).

Para avaliar a abordagem um-para-um, que avalia o desempenho do método em relação à verificação, é necessário a escolha de um limiar e a cada comparação entre duas faces é gerado um valor, caso este valor seja menor que o limiar as faces são consideradas pertencentes a um mesmo indivíduo, caso contrário, as faces são consideradas de indivíduos distintos. Variando o limiar escolhido previamente, pode-se gerar curvas que representam a Taxa de Falsa Rejeição (*False Rejection Rate*, FRR), quando o método rejeita um indivíduo quando deveria aceitá-lo a Taxa de Falsa Aceitação (*False Acceptance Rate*, FAR), quando o método aceita um indivíduo quando deveria rejeitá-lo. O ponto de encontro entre essas curvas é denominado a Taxa de Erro Igual (*Equal Error Rate*, EER) e é comumente utilizada para avaliar os métodos.

No geral, os sistemas de reconhecimento de faces podem ser classificados em dois grupos: baseados na aparência ou baseados nos pontos característicos da face. Os métodos baseados na aparência utilizam as características holísticas e são aplicados na face inteira ou em regiões específicas. Quanto aos métodos baseados em características, a maioria usa o comprimento da cabeça, a distância entre os olhos e dos olhos com a boca ou as distâncias e ângulos entre os cantos dos olhos, os extremos da boca, nariz no processo de reconhecimento (Zhao *et al.*, 2003).

Geralmente, o fluxo para a resolução do problema de reconhecimento de faces envolve a detecção desta a partir de uma cena que pode conter várias pessoas (Etapa de Detecção), o processamento da imagem (Normalização Fotométrica, Recorte, Escala, etc.), a extração de características e por fim, a classificação (Etapa de Reconhecimento). A Figura 8 demonstra este processo.



**Figura 8.** Ilustração do fluxo utilizado no reconhecimento de faces.

### 2.8.1 Normalização da Face

Em um processo de reconhecimento, as imagens geralmente passam por um processo de normalização geométrica e fotométrica logo após a etapa de detecção. O processo geométrico consiste em normalizar as imagens de modo que as imagens, no final do processo, possuam a mesma posição e escala. O processo fotométrico consiste em melhorar as condições da imagem de modo que ruídos, sombras e variações de contraste ou de brilho não afetem o processo de reconhecimento.

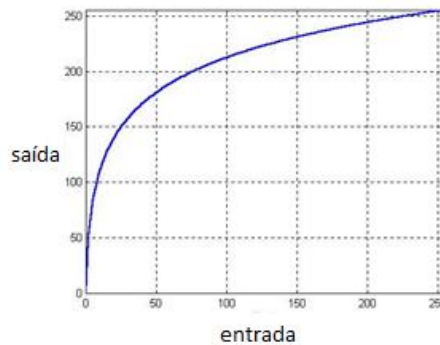
A iluminação do ambiente é um dos fatores mais importantes para determinar o sucesso ou o fracasso de um método, logo, é muito importante se preocupar com a compensação das variações do ambiente. Algumas técnicas de compensação de iluminação mais utilizadas são a Equalização de Histograma, a Transformação Não-Linear e o *LogAbout* (Liu *et al.*, 2002).

O objetivo da equalização de histograma é redistribuir os valores de tons de cinza dos *pixels* em uma imagem, de modo a obter um histograma mais uniforme. Para obter este resultado, utiliza-se uma função auxiliar denominada função de transformação. A forma mais comum de se equalizar um histograma é utilizar a Função de Distribuição Acumulada (*Cumulative Distribution Function*, CDF) da distribuição de probabilidades que pode ser expressa por Marques e Vieira (1999):

$$S_k = T(r_k) = \sum_{j=0}^k \frac{n_j}{n} = \sum_{j=0}^k P_r(r_j) \quad (10)$$

em que,  $0 \leq r_k \leq 1$ ,  $n$  é o número total de *pixels* na imagem,  $K = 0, 1, \dots, L-1$ , onde  $L$  é o número de níveis de cinza da imagem digitalizada,  $n_j$  é o número de *pixels* cujo nível de cinza corresponde à  $j$  e  $P_r(r_j)$  é a probabilidade do  $j$ -ésimo nível de cinza. Transformações não-lineares são muito utilizadas em visão computacional. Transformações baseadas na função logarítmica podem estender os níveis de cinza mais baixos e comprimir os níveis de cinza mais altos, o que pode melhorar a iluminação deficiente. É muito útil para reduzir sombras e para imagens com iluminação não-

uniformes (Liu *et al.*, 2003). A curva de uma transformação logarítmica pode ser observada na Figura 9.



**Figura 9.** Exemplo de Transformação não-linear  
Fonte: Adaptado de (Liu *et al.*, 2002)

O método *LogAbout* foi proposto por Liu *et al.* (2002) com o objetivo de tentar resolver problemas de compensação de iluminação em detecção de face. Este método consiste na aplicação de um filtro passa-alta na imagem original, seguida da função logarítmica.

### 2.8.2 Rastreamento de Faces

Apesar dos avanços na área, o reconhecimento de faces ainda pode ser considerada uma tarefa desafiadora principalmente em ambientes dinâmicos, como vídeos, onde as condições da imagem podem variar significativamente de um quadro para o outro. Realizar o reconhecimento de faces em quadros individuais pode gerar evidências fracas, entretanto, a integração das informações de vários quadros consecutivos pode levar a uma extração de informação mais precisa. Um conceito importante que auxilia na obtenção de informações integradas dos quadros de um vídeo é o de rastreamento de faces (Kim *et al.*, 2008).

As etapas de detecção de faces e de reconhecimento de faces são passos críticos que impactam diretamente o rastreamento de indivíduos em vídeo. O rastreamento de faces pode ser visto como uma filtragem temporal que estima a probabilidade da imagem de entrada está em um determinado estado como demonstra a Equação 11.

$$P(u_t | F_{0...t}), t = 1, 2, \dots \quad (11)$$

em que  $F_t$  é a imagem de entrada e  $u_t$  é o estado do rastreamento no tempo  $t$ .

O rastreamento de faces apresenta grande importância, principalmente quando múltiplos agentes, imagens de faces, por exemplo, são encontrados na mesma cena, pois permite a aplicação de algoritmos para analisar a interação entre agentes, a redução do espaço de busca para métodos de detecção baseados em janela deslizante e o reconhecimento de ações individuais (Yilmaz *et al.*, 2006). Dado que o sistema conheça a identificação do indivíduo e saiba como rastrear individualmente cada indivíduo, as ações efetuadas por este poderão ser reconhecidas. Diversas abordagens vêm sendo utilizadas para o rastreamento de pessoas, como modelos baseados em aparência (Black *et al.*, 1996), modelos de contorno (Isard e Blake, 1996), modelos 3D (Cascia *et al.*, 2000) e representação de *kernel* (Comaniciu *et al.*, 2003).

Em aplicações para vigilância e monitoramento, o rastreamento de face vem sendo comumente utilizado junto à reidentificação das faces, que visa à identificação de indivíduos que já foram previamente identificados em quadros anteriores.

### 3. Trabalhos Correlatos

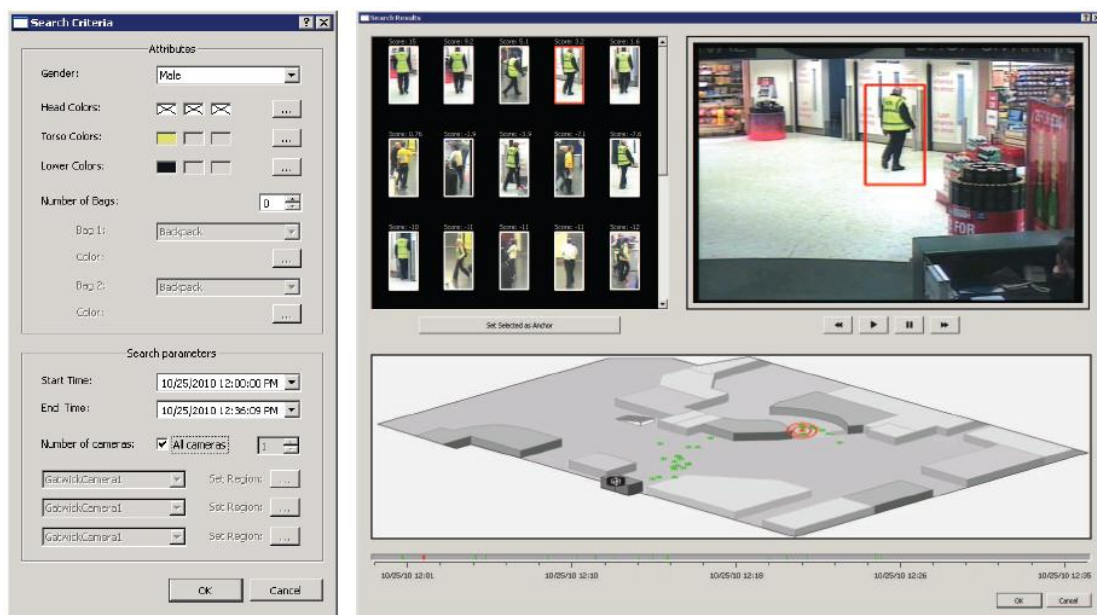
Neste capítulo são descritos os trabalhos relacionados ao uso de métodos computacionais com a finalidade de obter informações que possam ajudar na análise forense em vídeos de sistemas de vigilância. O foco desta revisão bibliográfica foi em extração de características forenses e extração de características em vídeos de câmeras de vigilância.

A baixa qualidade dos vídeos gravados por câmeras de circuito fechado pode esconder informações úteis, baseado nisto Jerian *et al.* (2007) apresentam um ambiente de *software*, denominado Ambiente de Processamento de Imagem Modular (*Modular Image Processing Environment*, MIPE), desenvolvido para ser usado como uma ferramenta forense com o intuito de investigar vídeos de vigilância. O principal objetivo do trabalho é proporcionar ao examinador o conhecimento e o controle completo sobre o que os métodos de tratamento da imagem podem fazer nas imagens degradadas. As principais contribuições deste trabalho são melhorar a qualidade das imagens dos vídeos forenses, que normalmente possuem qualidade baixa, para que detalhes de interesse possam ser melhor encontrados, se tornar uma aplicação acessível para restauração de imagens, usando validação de imagens, destacando bordas e sendo aberto para aceitar outros algoritmos que extraíam informações forense relevantes. Alguns exemplos de métodos utilizados neste trabalho são métodos para corrigir as linhas de deslocamento, que podem aparecer nas imagens, e também a correção de efeitos de embassamento ocasionados por movimentações na cena.

Uma tendência nos sistemas de vigilância é a introdução de uma inteligência segura e ativa através do desenvolvimento de técnicas computacionais capazes de extrair informações das cenas de forma automática. Uma técnica prática e desafiadora é a detecção e localização de anomalias, eventos com baixa probabilidade de ocorrência, em vídeos de vigilância de maneira que o sistema possa avisar que algumas atividades inesperadas estão ocorrendo. Xiao *et al.* (2015) e Sodemann *et al.* (2012) realizam uma revisão de diversas técnicas capazes de detectar anomalias de forma automática, algumas técnicas bastantes empregadas são o HOG(Dalal e Triggs, 2005) (Felzenszwalb *et al.* 2010) e o fluxo ótico(Adam *et al.* 2008)( Mehran *et al.* 2010). São empregadas diversas técnicas para extrair as mais diversas informações de vídeos de câmeras de vigilância, como por exemplo, movimentação de objetos intrusos (Chen *et al.*, 2014),

classificação de objetos abandonados (Otoom *et al.*, 2008), classificação do tráfego de veículos (Katkar *et al.*, 2015), rastreamento de pessoas (Li *et al.*, 2015), entre outras.

Algumas ferramentas que utilizam métodos de extração de características e com objetivos similares ao deste trabalho foram propostas na literatura. Thornton *et al.* (2011) descrevem uma nova abordagem para análise de vídeos, que permite uma busca através de grandes volumes de dados de vídeos de vigilância, com o objetivo de encontrar pessoas que correspondem a um perfil determinado. Através da ferramenta proposta, os autores esperam localizar rapidamente um indivíduo de interesse em tempo real, ou acelerando o processo de investigação. A técnica proposta é voltada para vigilância de grandes áreas, ou seja, o perfil que se deseja encontrar nas cenas é composto de atributos que são observáveis à distância (incluindo informações de roupas, cor do cabelo, etc.) ao invés de identificar informações ao nível de face. A Figura 10 apresenta um *screenshot* da ferramenta desenvolvida pelos autores.



**Figura 10.** Screenshot da ferramenta proposta por Thornton *et al.* (2011)  
Fonte: (Thornton *et al.*, 2011)

Aramvith *et al.* (2009) destacaram a necessidade de sistemas de monitoramento e vigilância inteligentes que auxiliem no processamento e entendimento de informações e desenvolveram um sistema denominado *Video Information Processing and Analysis for Surveillance System*. O sistema desenvolvido é composto por 6 partes que são reconstrução de super resolução de imagens, detecção de suspeitos através da subtração de *background*, rastreamento de suspeitos, extração de aparência de suspeitos como altura, velocidade, segmentação de partes do corpo, detecção de acessórios e cor da

pele, análise da atividade de suspeitos e nível de decisão consciente. A saída do sistema são características dos suspeitos, padrão e direção das atividades suspeitas e nível de incidentes suspeitos.

Attard e Farrugia (2011) sugeriram a necessidade de sistemas de vigilância de vídeo com contínuo monitoramento em substituição a sistemas ineficazes tradicionais, pois, o quantidade de dados geradas por estes sistemas é muito grande comparada ao efetivo humano. Os autores apresentaram um sistema autônomo de vigilância de vídeo capaz de detectar e rastrear tanto atividades humanas quanto atividades relacionadas a veículos. Para realizar isto, os autores trabalharam focados em 3 áreas. Inicialmente o método de subtração de *background* para detecção de objetos seguida de uma extração de regiões de interesse foi utilizado para detectar objetos que se movimentaram na cena. Posteriormente, a classificação de objetos foi utilizada nestas regiões de interesse para reconhecer humanos e veículos na cena. Por fim, um rastreamento destes objetos foi proposto com o objetivo de armazenar a trajetória destes na cena.

## 4. Materiais e Métodos

Os materiais utilizados e os métodos propostos estão descritos nas seções seguintes. Em cada seção são expostas as técnicas utilizadas, assim como os materiais necessários para cada método.

### 4.1 Bancos de dados utilizados

Para desenvolver os métodos e realizar os testes foi necessária a utilização de alguns bancos de dados de imagens existentes na literatura, descritos a seguir.

#### 4.1.1 Banco de Faces Yale A

O Banco de Faces da Universidade de Yale (Georghiades *et al.*, 2001) foi utilizado neste trabalho para os testes de reconhecimento de faces. O banco possui imagens de 15 pessoas em 11 situações diferentes para representar cada pessoa, totalizando 165 imagens ao todo. As imagens possuem diferentes expressões faciais ou configurações como: diferentes centros de iluminação, óculos, expressão de felicidade, expressão de tristeza, luz à esquerda, luz à direita, expressão de sono, etc. As imagens encontram-se em tons de cinza. A Figura 11 mostra um indivíduo presente no banco de dados e suas diversas variações.



**Figura 11.** Banco de dados Yale A: pessoa 1 em suas 11 poses.  
Fonte: Banco de Dados Yale Face Database A (Georghiades *et al.*, 2001)

#### 4.1.2 Banco de Faces Yale B

O Banco de faces *Extended Yale B* (Georghiades *et al.*, 2001) foi utilizado neste trabalho para os testes de reconhecimento de faces. Possui 16128 imagens de 28 indivíduos diferentes, sendo 9 posições diferentes sobre 64 formas de iluminação diferentes. Como o reconhecedor de faces proposto foi desenvolvido visando reconhecimento frontal de faces, foram selecionadas 10 imagens frontais de cada indivíduo, ou com pouca inclinação, e com alguma variação de iluminação para serem

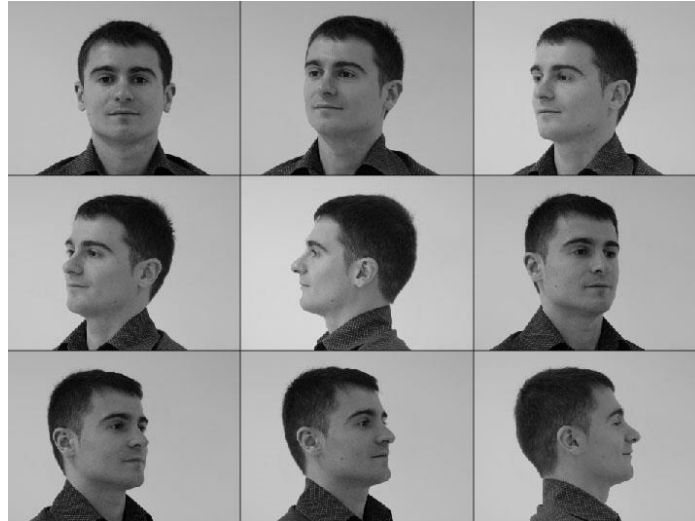
utilizadas nesse trabalho. As imagens selecionadas foram as das posições 01, 02, 03, 04 e 05, com a iluminação frontal e a imagem da posição 00 com iluminação frontal e com as variações de iluminação identificadas pelas seguintes nomenclaturas: A+005E+10, A+005E-10, A-005E+10 e A-005E-10. Através desta seleção criou-se um banco de 256 imagens, todas com dimensão igual a 640x480, já em níveis de cinza. Algumas imagens selecionadas podem ser visualizadas na Figura 12.



**Figura 12.** Banco de dados Extended Yale B  
Fonte: Banco de Dados Extended Yale B (Georghiades *et al.*, 2001)

#### 4.1.3 Banco de Faces GTAV

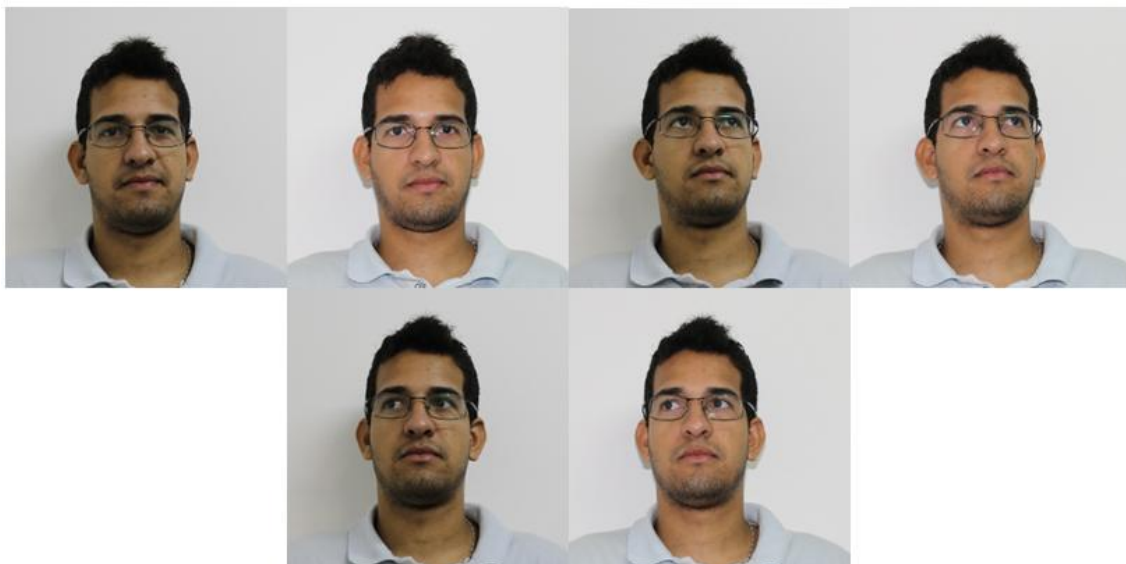
O banco de faces GTAV (Tarrés e Rama) foi criado para testar a robustez dos algoritmos de reconhecimento de face com relação à variação da pose e da iluminação. É composto por 44 indivíduos com 27 imagens de cada indivíduo. As imagens apresentam variações de expressão facial e acessórios. A dimensão de cada foto é de 320 x 240 e elas estão no formato *bmp*. Para este trabalho, foram utilizadas apenas 414 imagens frontais escolhidas manualmente. Esta escolha foi realizada para retirar imagens em que os indivíduos aparecem obstruindo boca ou olhos, assim como fotos de perfil. A Figura 13 ilustra algumas imagens do banco.



**Figura 13.** Amostras do banco de dados GTAV  
Fonte: (Tarrés e Rama)

#### 4.1.4 Banco de Faces VSOFT-Faces

O banco de faces Vsoft-Faces (Marinho, 2012) possui imagens de 14 sujeitos em seis poses, sob duas fontes de iluminação e contra um fundo branco uniforme. As fotos originais possuem 1500 x 1500 *pixels*. Entretanto, para este trabalho, foi utilizada uma variação do banco com imagens de tamanho 250 x 250 *pixels*, com o intuito de utilizar uma resolução mais próxima da resolução dos bancos utilizados na literatura. Na Figura 14, são apresentadas algumas imagens que fazem parte das fotos originais do banco.



**Figura 14.** Amostras do banco de faces Vsoft-Faces  
Fonte: Banco de dados Vsoft-Faces (Marinho, 2012)

Com o intuito de realizar testes de reconhecimento facial utilizando faces capturadas de vídeos, foi construída uma base de vídeos com os mesmos indivíduos que estão presentes no banco de faces construído por Marinho (2012). Os vídeos gravados simulam câmeras de vigilância e possuem resolução de 1920 x 1080. Uma peculiaridade desses vídeos é o fato dos indivíduos aparecerem com camisas de cores distintas e, em alguns momentos, com determinados acessórios, como óculos, e em outros momentos, sem acessórios. Outros vídeos com os mesmos indivíduos do banco de faces Vsoft-Faces foram extraídos das câmeras de vigilância do sistema de segurança da empresa Vsoft Tecnologia Ltda., com o intuito de utilizá-los em testes com câmeras de vigilância reais, entretanto, a baixa qualidade destes vídeos (350x240) não permitiu a sua utilização nos testes de reconhecimento facial. A Figura 15 mostra uma sequência de quadros dos vídeos gravados.



**Figura 15.** Uma sequência de quadros do vídeo gerado para testes

#### **4.1.5 Banco de Faces FERET (The Facial Recognition Technology)**

O banco de faces FERET (Phillips *et al.*, 2000) foi construído para desenvolver capacidades relacionadas ao reconhecimento automático de faces. O banco original possui 14051 imagens em tom de cinza que vão desde imagens frontais a imagens de perfil. Neste trabalho, o banco de dados FERET foi utilizado para testar o método de reconhecimento de gênero. Para efeito comparativo com a literatura, foram selecionadas aleatoriamente 540 imagens frontais de indivíduos distintos. A Figura 16 mostra algumas imagens do banco de faces utilizadas no trabalho.



**Figura 16.** Amostras do banco FERET utilizadas no trabalho proposto  
Fonte: banco de faces FERET (Phillips *et al.*, 2000)

#### 4.1.6 *Incremental Learning for Visual Tracking Dataset*

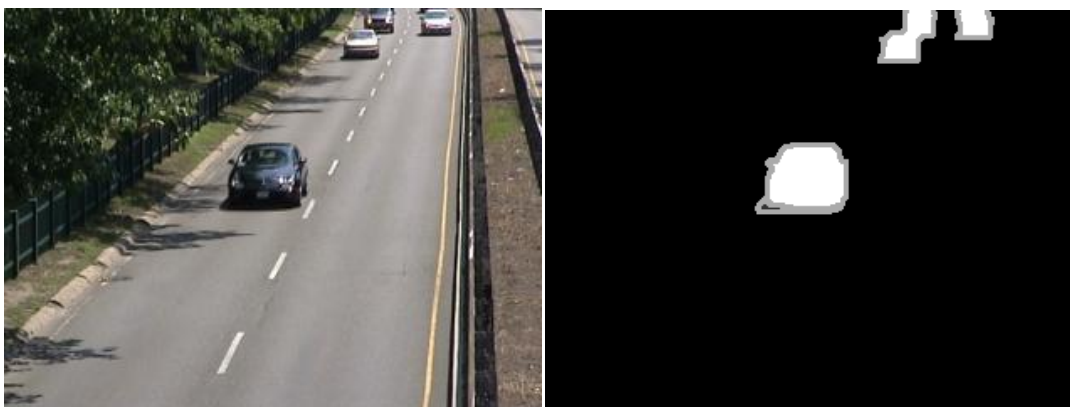
O banco de dados utilizado em Ross *et al.* (2007) foi criado para realizar testes sobre um algoritmo de rastreamento de objetos. O banco de dados é composto por quatro vídeos com objetos se movimentando. Para realizar os testes sobre o método de rastreamento de faces proposto, foram utilizados dois vídeos deste banco contendo indivíduos em movimentação. O primeiro é uma sequência gravada em um ambiente interno onde um indivíduo se move de uma região escura para uma região com grande iluminação, este vídeo é denominado *indoor sequence*. O segundo vídeo utilizado neste trabalho, denominado *outdoor sequence*, possui um indivíduo que se move em um ambiente externo proporcionando assim grandes mudanças de iluminação e sombras nas mudanças de pose. A Figura 17 ilustra os dois vídeos utilizados.



**Figura 17.** Amostra de quadros dos vídeos *indoor sequence* e *outdoor sequence*.  
Fonte: Banco de dados Incremental Learning for Visual Tracking Dataset (Ross *et al.*, 2007).

#### 4.1.7 Banco de Referência para Detecção de Mudanças (*Change Detection Benchmark Dataset 2012*)

O banco de referência para detecção de mudanças em quadros "Dataset 2012" (Goyette *et al.*, 2012) contém 6 categorias de vídeos com 4 à 6 seqüências de vídeos em cada categoria. Os vídeos vêm acompanhados de subdiretórios com todos os quadros do vídeo e com o gabarito de movimentação para cada quadro. O gabarito de movimentação contém alguns tipos de marcações, para regiões estáticas, o valor é 0, para regiões com ocorrência de sombras, o valor é 50, para regiões fora da região de interesse, o valor é 85 e para regiões com movimentação, o valor é 255. A Figura 18 demonstra um quadro e seu respectivo gabarito.



**Figura 18.** Amostra do banco Dataset 2012 (Goyette *et al.*, 2012) que mostra a imagem de um quadro e seu respectivo gabarito de movimentação.

Fonte: banco Dataset 2012 (Goyette *et al.*, 2012)

Para este trabalho, os testes foram realizados sobre apenas uma categoria de vídeos do banco contendo 4 vídeos de amostra denominada *baseline*. Esta categoria foi escolhida, pois os vídeos que ela contém possuem características que se adéquam aos propósitos dos métodos desenvolvidos.

#### 4.2 Método de Detecção e Análise de Movimentação

Para extrair informações sobre movimentação de uma cena, o método desenvolvido foi a subtração de *background*, que é implementado através da utilização de um modelo de plano de fundo, sendo efetuada assim a diferença espacial entre este modelo do ambiente e a imagem do quadro atual. Esta abordagem foi escolhida, pois geralmente câmeras de vigilância estão em uma posição estática, permitindo que o

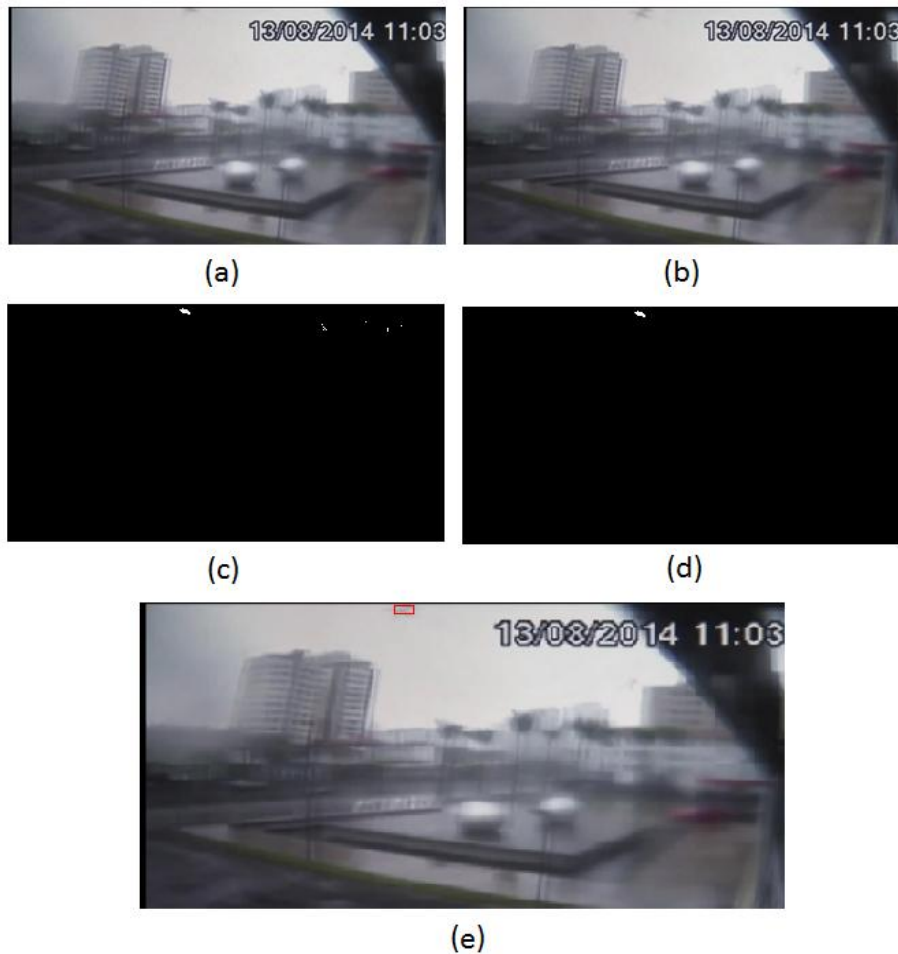
fundo seja sempre bem definido. A diferença resultante da subtração de *background* é obtida através da aplicação da Equação 12 (Haupt, 2002).

$$\Delta I = |f(x, y, t) - f(x, y)| \quad (12)$$

Uma característica peculiar do método implementado é que ele tenta evitar a falsa detecção de movimento através da atualização do *background*, caso isto não fosse feito e o fundo fosse tomado como estático, sempre que houvesse uma variação permanente do cenário, esta sempre seria detectada como movimentação. Portanto, o método atualiza o *background* sempre que não for detectado movimento de um quadro para o outro. Para efetuar a atualização do *background* o método proposto avalia a Equação 13 descrita abaixo:

$$f(x, y) = \left\{ \begin{array}{ll} f(x, y, t) & , \Delta I > 0 \\ f(x, y) & , \Delta I = 0 \end{array} \right\} \quad (13)$$

Com o intuito de tornar a detecção ainda mais específica, foram implementados parâmetros para determinar o limiar de tamanho dos objetos a serem pesquisados, assim como a cor predominante no objeto. Para encontrar objetos de um tamanho específico, os *pixels* do quadro atual que sofreram diferenciação em relação ao *background* passam por um processo de detecção de *blobs* (*blob detection*), que consiste em agrupar estes *pixels* em objetos e diferenciá-los de outros conjuntos de *pixels* no quadro. A Figura 19 ilustra esse processo.



**Figura 19.** Ilustração da detecção de objetos em movimento.  
 (a) *background*, B (b) quadro atual, A (c) diferença das imagens (A-B) (d) imagem após filtragem de objetos com tamanho mínimo (12x12) (e) resultado final do método com o objeto encontrado indicado por um retângulo vermelho

Outro parâmetro que identifica de forma mais específica objetos na imagem, é a escolha de cor predominante no objeto. Para isto, o método proposto após detectar objetos que se movimentaram e separá-los através do processo de *blob detection*, utiliza-se do sistema de cor HSV para detectar qual a cor predominante naquele objeto.

Como os valores de cores RGB variam muito de acordo com as condições de iluminação, sombras, etc, o método transforma a imagem original para o sistema de cores HSV e através dos canais *h* (*hue*), *s* (*saturation*) e *v* (*value*) extrai a informação de qual cor está presente em cada *pixel* do objeto, e atribui como cor predominante a que está representada em mais *pixels*.

A Figura 20 ilustra um quadro no vídeo onde foi encontrado um objeto que se movimentou e que possui características desejadas como cor predominante, cor laranja, por exemplo, largura mínima de 70 *pixels* e altura mínima de 70 *pixels*.



**Figura 20.** Ilustração de um quadro que possui objeto com características procuradas (objeto marcado em verde) e os em vermelho são objetos encontrados, porém que não satisfazem as características desejadas.

Além de poder escolher encontrar objetos com características específicas, o método proposto também possibilita gerar um gráfico sobre a movimentação do vídeo, que pode ser sobre todos os objetos em cena ou apenas os que possuem as características especificadas. Dessa maneira, é possível obter certas informações, como o quadro onde houve maior quantidade de movimento. A quantidade de movimento é medida como a quantidade de *pixels* que sofreram alteração quando comparado o quadro atual com o *background* e pode ser expressa na Equação 14.

$$Q = \frac{p'}{p} \quad (14)$$

em que  $p'$  é a quantidade de *pixels* alterados e  $p$  é a quantidade total de *pixels* da imagem.

Esta característica é muito importante, pois possibilita ao examinador concentrar-se apenas em momentos do vídeo em que o gráfico mostre que houve movimentação, ou movimentação acima de um determinado limiar. Uma ilustração de um gráfico gerado e do quadro com maior movimentação pode ser visualizada na Figura 21.



**Figura 21.** Ilustração do quadro com maior movimentação e do gráfico de movimentação. Em vermelho estão destacadas as áreas que sofreram alteração em relação ao *background*.

Os vídeos utilizados para avaliar este método foram vídeos de câmeras de vigilância extraídos do site de domínio público *youtube*, vídeos do banco de referência para detecção de mudanças em quadros (*Change Detection Benchmark Dataset*) (Goyette *et al.*, 2012), assim como vídeos das câmeras de vigilância cedidos pela empresa Vsoft Tecnologia Ltda.

Para efeito comparativo com outros métodos presentes na literatura, foram calculados valores especificados nas equações 3 a 9, descritas na seção 2.6, que medem a sensibilidade, a especificidade, a taxa de falsos positivos, a taxa de falsos negativos, a porcentagem de classificações erradas, a precisão e a medida-F do método proposto sobre os vídeos do banco de referência para detecção de mudanças em quadros (*Dataset 2012*) (Goyette *et al.*, 2012). Para calcular essas métricas foi necessário comparar cada *pixel* do quadro do gabarito de movimentação apresentado no banco de dados com os *pixels* do quadro de movimentação resultante do método proposto. A Figura 22 compara visualmente o gabarito de um determinado quadro com o resultado do método proposto.



**Figura 22.** Comparação entre movimentação gerada pelo método proposto e gabarito de movimentação do banco de dados (2012 *Dataset*) (Goyette *et al.*, 2012).

### **4.3 Extração de Características a Partir de uma Face**

Com o objetivo de extrair características peculiares à seres humanos, alguns métodos foram desenvolvidos como: detecção de faces, reconhecimento da cor predominante no vestuário de um indivíduo, reconhecimento de gênero, rastreamento e reconhecimento de faces.

Para desenvolver os métodos e realizar os testes foi necessária a utilização de alguns bancos de faces existentes na literatura. Os bancos de faces escolhidos para os testes de reconhecimento de faces foram o Yale A (Georghiades *et al.*, 2001), Extended Yale B (Georghiades *et al.*, 2001), o GTAV (Tarrés e Rama) e o banco de faces VSOFT-Faces (Marinho, 2012). Para o reconhecimento de gênero foi utilizado o banco de faces FERET (Phillips *et al.*, 2000).

Como todos os métodos propostos, que serão descritos a seguir, dependem da localização de uma face na imagem, inicialmente foi utilizado o algoritmo de detecção de faces de Viola e Jones (2001), que utiliza conceitos de reconhecimento de padrões e Adaboost, já implementados na biblioteca multiplataforma *OpenCv* (Zelinsky, 2008). Alguns testes foram realizados para determinar o melhor conjunto de parâmetros para detecção de faces, como fator de escala e tamanho das faces.

#### **4.3.1 Reconhecimento da Cor Predominante no Vestuário de um Indivíduo**

A informação de cor predominante no vestuário de um indivíduo pode ser utilizada em diversas aplicações, como controle de acesso e robótica, ou em apoio para otimizar outros métodos. Na análise de cenas forenses, encontrar esta cor pode ser importante, por exemplo, para casos onde o examinador deseja procurar um suspeito que de acordo com testemunhas, ou outras informações, estava com certo padrão de cor em sua vestimenta.

Para encontrar a cor predominante no vestuário de um indivíduo, o método proposto utiliza a informação de localização de faces para pressupor a localização do vestuário. Para encontrar a região onde supostamente existe informação de vestimenta, o método pressupõe que os indivíduos estão em uma pose de pé, portanto, ao ter a informação da localização da face, o método procura por uma região abaixo desta que, proporcionalmente ao seu tamanho, possa representar a vestimenta do indivíduo. Ao encontrar essa região, o método converte cada pixel para componentes de cor segundo o padrão HSV.

Esta conversão é importante, pois é invariante à alta intensidade na luz branca, luz ambiente e orientações de superfície relativas à fonte de luz, portanto, pode formar uma boa escolha para métodos de detecção de cor.

Ao término da conversão, o método realiza uma varredura na região verificando qual a cor predominante na vestimenta do indivíduo de respectiva face anteriormente encontrada. Para determinar qual cor no sistema HSV o *pixel* representa, uma tabela de cores foi previamente construída contendo informações do cone HSV, inicialmente compara-se o valor do *hue* (H), que varia de 0° a 180° no método proposto, para descobrir qual a tonalidade dominante da cor, e em seguida verifica-se a saturação, grau de pureza, e por fim o brilho. Caso o brilho seja menor que 40, a cor é considerada a cor preta, caso o brilho seja maior que 190 e a saturação menor que 27 a considerada a cor branca e caso o brilho seja maior que 185 e a saturação maior que 53 é considerada a cor cinza . A Tabela construída e utilizada no método proposto pode ser visualizada abaixo:

**Tabela 1.** Tabela construída para atribuição de cor no método de reconhecimento de cor predominante no vestuário de um indivíduo.

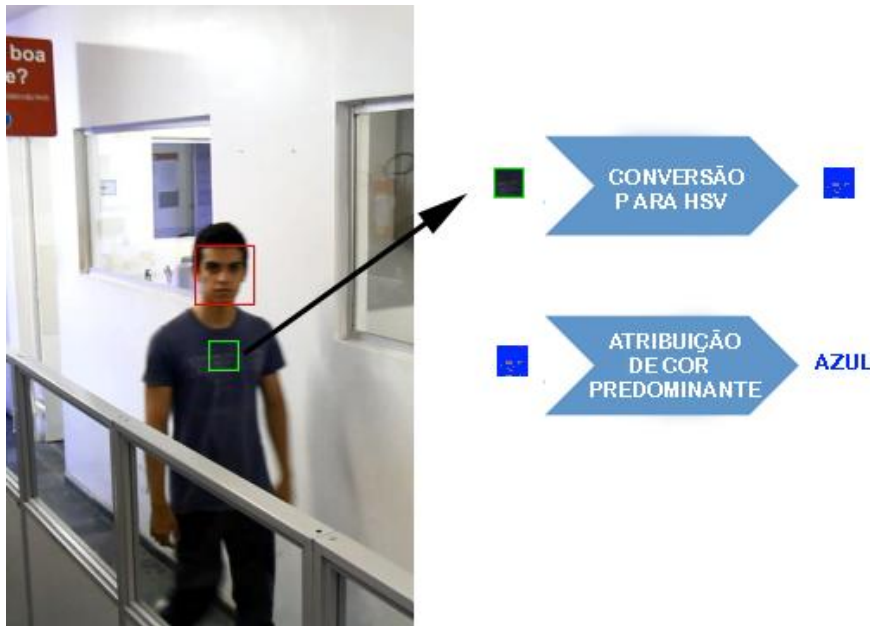
Cor	Hue (H)
Vermelho	$h \leq 14$
Laranja	$14 < h \leq 25$
Amarelo	$25 < h \leq 34$
Verde	$34 < h \leq 73$
Aqua	$73 < h \leq 102$
Azul	$102 < h \leq 127$
Roxo	$127 < h \leq 149$
Rosa	$149 < h \leq 175$
Vermelho	$h > 175$

Para realizar os testes, foram utilizados os quadros de vídeos gravados em uma empresa local, assim como utilizados vídeos encontrados em bancos na literatura como o *Incremental Learning for Visual Tracking* (Ross *et al.*, 2007), extraído-se as imagens onde pelo menos uma face estava presente. Para calcular a taxa de acerto do método proposto, a cor predominante no vestuário de cada indivíduo presente nos vídeos foi marcada manualmente. Em seguida, a cada atribuição do método proposto, a cor predominante contida no arquivo de gabarito é comparada com a cor que o método

atribuiu ao indivíduo, caso sejam iguais, é considerado um acerto, caso contrário, é considerado um erro. A taxa de acerto pode ser expressa pela Equação 15:

$$\text{taxa de acerto} = \frac{\text{atribuições corretas}}{\text{total de atribuições}} \quad (15)$$

A Figura 23 ilustra o método proposto.



**Figura 23.** Ilustração do reconhecimento de cor predominante no vestuário

### 4.3.2 Reconhecimento de Faces e de Gênero

Os métodos de reconhecimento de faces e de gênero foram desenvolvidos seguindo raciocínio semelhante. Para reconhecer o gênero de uma face, a etapa de detecção de faces foi aplicada nas imagens do banco de faces FERET (Phillips *et al.*, 2000), com o objetivo de retirar dados desnecessários para etapas posteriores. De forma semelhante, para realizar o reconhecimento de faces, a etapa de detecção de faces foi aplicada nas imagens do banco de faces YALE A (Georghiades *et al.*, 2001), YALE B (Georghiades *et al.*, 2001) e o GTAV (Tarrés e Rama). Estes bancos foram escolhidos por serem utilizados na literatura. Foi aplicada a etapa de detecção de faces ainda sobre o banco de dados VSOFT-Faces (Marinho, 2012) e nos vídeos criados com indivíduos presentes neste mesmo banco.

Em seguida, algumas técnicas de normalização foram aplicadas com o intuito de melhorar as condições das faces extraídas dos bancos. Primeiramente, foi aplicada a

normalização de escala e rotação, também chamada de normalização geométrica, que visa retirar possíveis variações de posição, rotação e escala entre as faces. Para realizar a etapa de normalização geométrica, o algoritmo desenvolvido utiliza-se dos pontos detectados pelo ASM. Ao detectar os pontos da face, extrai-se a coordenada do olho esquerdo e a coordenada do olho direito, analisa-se o ângulo de rotação e a distância entre eles.

Para normalizar as imagens quanto à rotação, a imagem é rotacionada até que o ângulo entre os pontos dos olhos fique com valor zero. Para normalizar quanto à escala, inicialmente um fator de redimensionamento  $f$  é calculado, este fator pode ser expressado na Equação 16.

$$f = \frac{d_b}{d} \quad (16)$$

em que  $d_b$  é a distância padrão escolhida previamente e  $d$  é a distância entre os olhos encontrados pelo ASM. Após encontrar este fator de escala, a imagem é redimensionada para que a distância dos olhos, encontrados pelo ASM, fique igual a distância dos olhos previamente encontrada.

A Figura 24 apresenta o fluxo e o resultado da normalização geométrica desenvolvida.



**Figura 24.** Ilustração do ângulo e da distância entre os olhos. (a) Imagem original (b) Imagem normalizada geometricamente

Após a etapa de normalização geométrica, alguns métodos de normalização fotométrica foram testados com o intuito de retirar sombras e variações de contraste ou de brilho. Os métodos de normalização utilizados foram Equalização de Histograma,

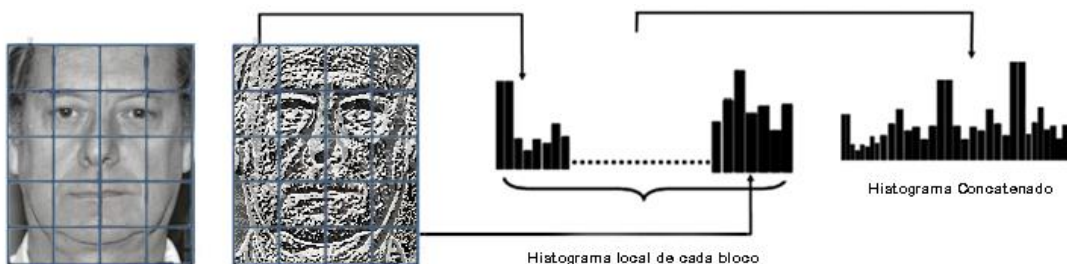
Transformação Não-Linear (*Log*) e o *LogAbout* (Liu *et al.*, 2002). A Figura 25 ilustra o efeito de cada um destes métodos, respectivamente.



**Figura 25.** a) Imagem original Métodos de normalização fotométrica: b) Equalização de histograma c) *Log* d) *LogAbout*

Após as etapas de normalização, primeiramente uma etapa para extração de características da face é proposta e posteriormente uma classificação destas utilizando arquivos rotulados anteriormente. Para extração foi utilizado o LBP que extrai as características locais da textura da imagem de face. Nesse trabalho foram utilizado os Histogramas de Padrões Locais Binários (*Local Binary Pattern Histograms*, LPBH), que divide a imagem em blocos, gera o histograma de LBP de cada bloco, e posteriormente, concatena os histogramas locais gerando assim o vetor de características. O vetor de características é então entendido como uma coleção de histogramas de LBP (Figura 26).

Inicialmente, os parâmetros utilizados para o  $LBP_{P,R}$  foram de 8 pontos igualmente espaçados e com distância 1 do centro, visando realizar testes com a proximidade do LBP padrão que utiliza uma máscara 3 x 3.



**Figura 26.** Histogramas de Padrão Binário Local.

Para a etapa de classificação, os vetores de características gerados foram comparados tanto na abordagem 1-1 (um para um), que compara um dado histograma com todos os outros pré-rotulados, como também na abordagem 1-n (um para n), onde é

utilizado o conceito de validação cruzada, onde divide-se o banco de imagens em grupos, e para cada rodada um grupo é escolhido para treinamento e os outros grupos são escolhidos para serem as imagens de testes. A validação cruzada acabará quando cada grupo tiver sido escolhido o de treinamento uma vez.

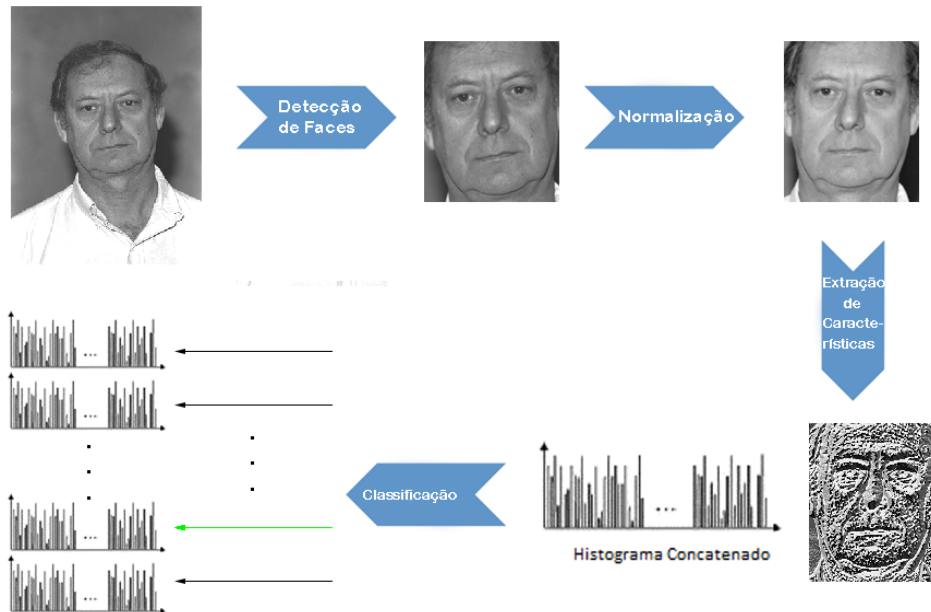
Para classificar os vetores de características, a etapa de classificação computa um *score* para cada par de vetores através da métrica de medida de similaridade. No método proposto, a métrica utilizada foi a correlação que pode ser medida pela Equação 17.

$$d(H_1, H_2) = \frac{\sum_i (H_1(i) - \bar{H}_1)(H_2(i) - \bar{H}_2)}{\sqrt{\sum_i (H_1(i) - \bar{H}_1)^2 \sum_i (H_2(i) - \bar{H}_2)^2}} \quad (17)$$

em que,  $H_1$  é o primeiro histograma a ser comparado,  $H_2$  é o segundo histograma a ser comparado e  $\bar{H}_k = \frac{1}{N} \sum_i H_k(j)$  onde  $N$  é o número total de valores do histograma.

Ao término da atribuição dos *scores* da rodada, para cada vetor de característica de entrada, será atribuído como seu gênero, o gênero do vetor onde o grau de similaridade foi maior, ou seja, a distância entre os histogramas de LBP foi menor. De modo semelhante, no reconhecimento de faces, o vetor de entrada será atribuído ao vetor de treinamento que apresentou maior similaridade com ele. Para calcular a taxa de acerto foi utilizada a Equação 15.

A Figura 27 ilustra o respectivo fluxo do método desenvolvido no reconhecimento de faces e de gênero.



**Figura 27.** Fluxo para o reconhecimento de faces e gênero

### 4.3.3 Rastreamento de Faces

Alguns métodos para rastreamento de faces através da reidentificação de pessoas utilizam-se de abordagens holísticas, como a Análise de Componentes Principais (PCA), que possuem a característica de não manter a escalabilidade para galerias de imagens faciais contendo muitos indivíduos. Para cada novo indivíduo, todos os modelos precisam ser reconstruídos devido à estimativa da capacidade de discriminação de cada descritor. Tendo em vista que para rastrear faces, é necessária a atualização constante das faces conhecidas, portanto, o método para reidentificação de faces foi desenvolvido utilizando-se a posição espacial da face e a textura de cada face encontrada (extraída pelo descritor LBP).

Para desenvolver o método para rastreamento de faces, inicialmente, para cada quadro de vídeo, foi realizada a detecção de faces. Ao encontrar uma face, aplicam-se as normalizações fotométrica e geométrica. O rastreamento proposto é aplicado somente às faces frontais.

Em seguida, o algoritmo compara esta face com as faces encontradas nos quadros anteriores. Esta comparação é feita inicialmente, comparando a posição espacial das faces do quadro anterior com a posição da face do quadro atual. Se a face estiver posicionada em uma distância menor que um limiar determinado, estas faces são consideradas do mesmo indivíduo, caso esteja em uma distância maior que o limiar, é considerada como uma nova face. Dado que a face seja considerada nova, pela etapa de

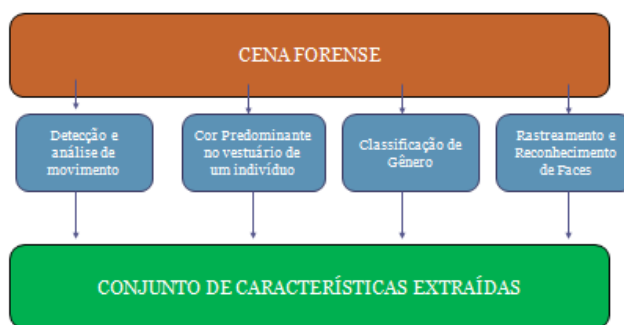
verificação de posição, o método então verificará se a face pertence ao banco de dados de faces conhecidas, caso ela pertença, é atribuído o identificador, caso contrário, o método adiciona esta face ao banco de dados de faces conhecidas, atribuindo a ela um novo identificador.

Para realizar a consulta ao banco de faces conhecidas, o método proposto utiliza-se do método de reconhecimento de faces anteriormente citado, ou seja, a cada nova face que necessite consultar o banco, é gerado um histograma LBP da imagem, e em seguida, procura-se no banco por um histograma similar.

Os testes foram realizados sobre vídeos do sistema de câmeras de vigilância da empresa Vsoft Tecnologia Ltda, gravados no ambiente interno da mesma empresa e também sobre vídeos do banco de dados *Incremental Learning for Visual Tracking Dataset* (Ross *et al.*, 2007).

#### 4.4 Sumarização de Cenas Através dos Métodos Propostos

Através dos métodos desenvolvidos uma sumarização pôde ser proposta, onde para cada método proposto, um conjunto de parâmetros pode ser utilizado para extrair da cena apenas informações relevantes. A sumarização realizada neste trabalho, ao processar uma cena, possibilita que o usuário realize uma análise em tempo real ou processe toda a cena sumarizando-a e armazenando os dados para posterior análise. A Figura 28 ilustra uma visão geral da abordagem proposta.



**Figura 28.** Visão geral da abordagem proposta

Quanto à detecção de movimentação, o sistema permite que o usuário escolha um conjunto de critérios de busca como tamanho dos objetos, cor predominante nos objetos e grau de movimentação que ocorreu na cena. Ao encontrar objetos com os parâmetros desejados, o sistema permite a gravação dos quadros em um novo vídeo, diminuindo o escopo do vídeo original e facilitando uma posterior análise, assim como o armazenamento do quadro que obteve maior movimentação. A ferramenta de

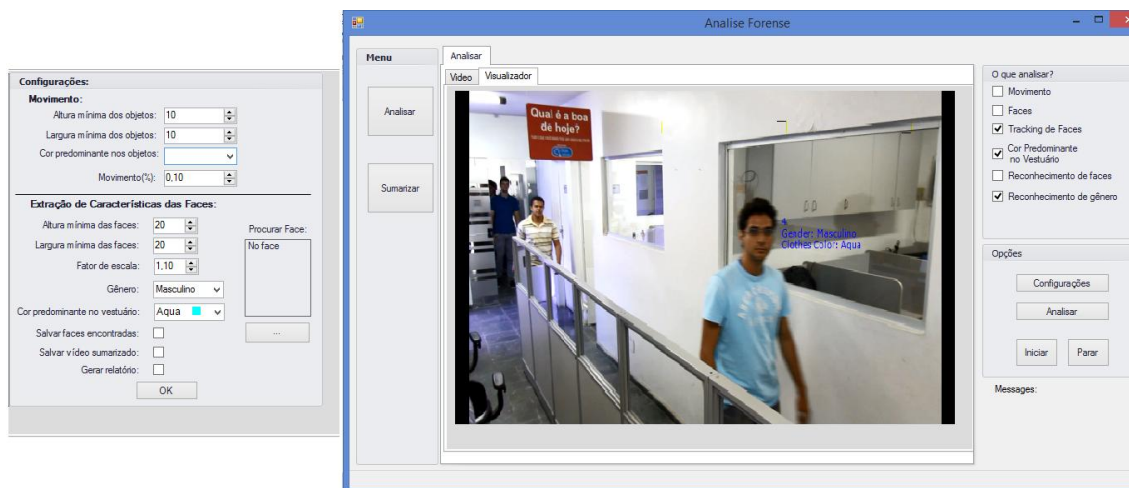
sumarização desenvolvida ainda permite a geração de um gráfico que relaciona a intensidade de movimentação na cena ao tempo de vídeo.

Em relação à extração de características a partir de uma face, a sumarização desenvolvida propõe para o conjunto de critérios de busca alguns parâmetros para extrair apenas informações que são relevantes para o examinador. Estes critérios de busca são: tamanho das faces, cor predominante no vestuário de um indivíduo, gênero e identidade da face à ser reconhecida. Após a escolha dos parâmetros, o sistema permite que ao encontrar na cena quadros que possuam as características desejadas, estes sejam armazenados para uma análise posterior.

Ao término de cada sumarização, a ferramenta desenvolvida pode ter como resultado um relatório contendo todas as informações que foram extraídas da cena. Este relatório pode ser útil para facilitar o armazenamento de informações assim como para visualizar e documentar os resultados da análise, podendo auxiliar na interpretação dos resultados e na sua argumentação.

À medida que cada método proposto seja aperfeiçoado, a ferramenta conseguirá cada vez mais sumarizar e armazenar uma maior quantidade de dados, resumindo assim o escopo de observação e podendo gerar análises mais rápidas.

A escolha dos critérios de busca é realizada através de uma interface gráfica que pode ser analisada através na Figura 29 que mostra uma captura da tela do sistema desenvolvido.



**Figura 29.** Captura de tela do sistema desenvolvido

Os testes de sumarização foram realizados extraindo características dos métodos propostos a partir dos vídeos do sistema de vigilância da empresa Vsoft Tecnologia

Ltda., vídeos gravados no ambiente interno da mesma empresa, vídeos de câmeras de vigilância extraídos do *youtube*, vídeos do banco de dados *Incremental Learning for Visual Tracking Dataset* (Ross *et al.*, 2007) e vídeos do banco de referência para detecção de mudanças em quadros (Goyette *et al.*, 2012).

## 5. Resultados e Discussão

Foram realizados alguns testes de acordo com o objetivo de cada método proposto. Esta seção descreve os resultados obtidos por estes.

### 5.1 Detecção e Análise de Movimentação

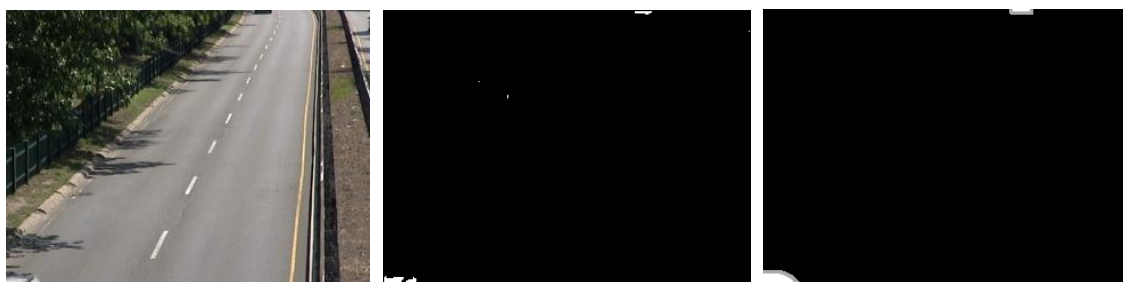
Para realizar os testes de detecção e análise de movimentação foram utilizados diversos vídeos com o intuito de observar o comportamento do método, porém para efeito comparativo com métodos existentes na literatura, métricas foram geradas a partir de testes com o banco de referência para detecção de mudanças em quadros (Goyette *et al.*, 2012). Foram utilizados todos os vídeos da categoria *baseline* presente no banco de dados. Para cada vídeo, foram encontrados os valores de verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN).

Após calcular os valores mencionados acima, foram calculadas as métricas utilizadas para efeito comparativo com outros métodos. Por intermédio das métricas geradas, observou-se que os resultados obtidos pelo método proposto estão compatíveis com outros métodos modernos (Yoshinaga *et al.*, 2013)(Strauss *et al.*, 2012) que foram avaliados com o mesmo banco de dados e seus resultados podem ser observados em Goyette *et al.* (2012). Os resultados podem ser observados na Tabela 2.

**Tabela 2.** Resultados das métricas para categoria *baseline* do banco de dados *Change Detection Benchmark Dataset (2012 Dataset)* (Goyette *et al.*, 2012)

Métricas	Método Proposto	Yoshinaga <i>et al.</i> (2013)	Strauss <i>et al.</i> (2012)
<i>Re</i>	0.6303	0.7551	0.7044
<i>Esp</i>	0.9932	0.9940	0.9923
<i>FPR</i>	0.0068	0.0060	0.0077
<i>FNR</i>	0.3697	0.2449	0.2956
<i>PWC</i>	2.3570	1.9154	2.2142
<i>Prec</i>	0.8001	0.7833	0.7009
<i>F-Measure</i>	0.7029	0.7554	0.6616

Um fato que pode ser observado na Tabela 2, é que as taxas de FNR e  $Re$  apresentam-se como uma deficiência do método proposto. O valor de falsos negativos influencia bastante nestas taxas, estes são encontrados quando o método não atribui movimento a uma área que houve movimentação. Outro fato observado é que algumas imagens apresentaram ruídos que foram contabilizados como falsos positivos e que podem ter influenciado nas taxas. A Figura 30 apresenta um quadro onde existe a presença de ruído e de falsos negativos em uma área que houve movimentação segundo o gabarito.



**Figura 30. Quadro original, quadro processado pelo método e que possui ruídos e presença de falsos positivos e quadro de gabarito, respectivamente.**

## **5.2 Extração de Características a partir da face**

Quanto aos métodos que extraem informação a partir da face, foi necessário utilizar uma etapa inicial que foi utilizar o detector de faces Viola e Jones (2001) nos bancos de faces Yale A, *Extended* Yale B, VSOF-T-Faces, FERET e GTAV.

### **5.2.1 Reconhecimento de cor predominante no vestuário de um indivíduo**

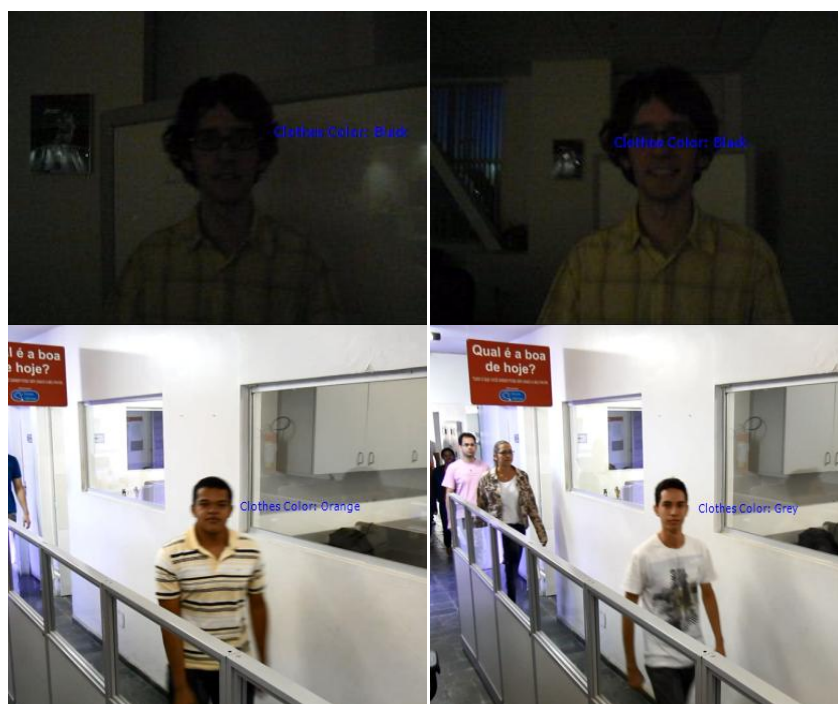
A partir da extração da face na cena, o método de reconhecimento de cor predominante no vestuário de um indivíduo foi testado em um conjunto de vídeos, onde para cada face encontrada no vídeo foi atribuído uma cor predominante ao vestuário. Para cada vídeo foi aplicado o rastreamento de faces, e para cada nova face detectada, a atribuição de cor predominante no vestuário foi realizada. Ao término, as atribuições foram analisadas e comparadas com a marcação realizada previamente da cor predominante no vestuário de cada indivíduo. Esta atribuição prévia foi realizada analisando os vídeos e armazenando a cor do vestuário em que cada indivíduo aparece.

Após comparar as atribuições com as marcações, seguindo a Equação 15, o método proposto obteve taxa de 81% de acerto. A Tabela 3 mostra os resultados obtidos.

**Tabela 3.** Reconhecimento de cor predominante no vestuário

Vídeo	Número de quadros	Número de indivíduos encontrados	Taxa de acerto (%)
vídeo Interno01	1604	13	93,00%
vídeo Interno02	4542	21	84,00%
<i>Outdoor sequence</i> (Ross et al. 2007)	569	6	50, 00%
<i>Indoor sequence</i> (Ross et al. 2007)	770	2	100,00%

Ao analisar o sistema proposto, algumas limitações puderam ser observadas em imagens com variações de iluminação no ambiente, onde o valor de saturação e brilho ocasionaram atribuições em que a pureza da cor não foi encontrada, e em camisas contendo alterações de cor como, listras horizontais ou com estampas. Algumas destas atribuições incorretas podem ser visualizadas na Figura 30.



**Figura 31.** Imagens normalizadas geometricamente pelos pontos extraídos do ASM

### 5.2.2 Reconhecimento de gênero

Para avaliar o método de classificação de gênero, foi realizado um estudo para analisar o desempenho do LBP, assim como o efeito das normalizações geométricas e fotométricas propostas. Para o teste de classificação de gênero, foram retiradas 540 imagens frontais de faces do banco de dados FERET (Phillips *et al.*, 2000), sendo 270 mulheres distintas e 270 homens distintos.

Para avaliar a classificação de gênero, os testes foram realizados através da técnica de validação cruzada. Inicialmente, nos testes de validação cruzada foram realizados testes com duas rodadas, onde cada rodada teve como treinamento um conjunto de 135 homens distintos e 135 mulheres distintas e o restante das imagens foi utilizado para os testes de classificação, e também com cinco rodadas, onde cada rodada teve como treinamento um conjunto de 54 homens distintos e 54 mulheres distintas, e o restante para os testes. Ao término de todas as rodadas a taxa de acerto média é calculada somando as taxas de acerto de cada rodada e dividindo pela quantidade de rodadas. A taxa de acerto de cada rodada pode ser medida através da Equação 15.

Para imagens onde a normalização geométrica não foi aplicada, para utilizar a comparação proposta, as imagens precisaram ser redimensionadas para possuírem o mesmo tamanho. Entretanto, esta transformação pode afetar os dados presentes na face.

Os testes foram realizados aplicando as diferentes técnicas de normalização, com o intuito de discutir o efeito destas sobre os métodos propostos. Quanto à classificação de gênero, o método obteve uma taxa de acerto máxima de 86,66% no grupo de imagens onde foi realizada a normalização de iluminação através da transformação não-linear (*Log*). A Tabela 4 demonstra as taxas de acerto do método proposto.

**Tabela 4.** Taxas de acerto para o método de classificação de gênero

<b>Detecção de Faces</b>	<b>Normalização Geométrica</b>	<b>Normalização Fotométrica</b>	<b>2 rodadas</b>	<b>5 rodadas</b>
Não	Não	Não	60,18%	63,51%
Sim	Não	Não	67,31%	73,14%
Sim	Não	Histograma	68,19%	73,70%
Sim	Não	Log	70,83%	77,03%
Sim	Não	LogAbout	64,86%	66,48%
Sim	Sim	Não	67,17%	71,66%

Sim	Sim	Histograma	67,87%	73,14%
Sim	Sim	Log	69,95%	75,00%
Sim	Sim	LogAbout	64,62%	65,03%

Ao analisar a Tabela 4, percebe-se que a normalização geométrica influenciou de forma negativa no reconhecimento de gênero. Isto pode ser explicado por que em alguns casos, o algoritmo ASM detectou de forma errada os olhos, influenciando assim a normalização geométrica. Para analisar o efeito correto da normalização geométrica, os pontos dos olhos foram marcados manualmente, substituindo os pontos encontrados pelo ASM e gerando uma nova normalização geométrica. A Tabela 5 mostra um comparativo dos resultados obtidos com esta nova normalização e com a normalização gerada através dos pontos encontrados pelo ASM.

**Tabela 5.** Comparação entre as taxas de acerto para o método de classificação de gênero com normalização geométrica por marcação manual e normalização geométrica por pontos do ASM.

<b>Deteccão de Faces</b>	<b>Normalização Fotométrica</b>	<b>Normalização Geométrica</b>	<b>2 rodadas</b>	<b>5 rodadas</b>
Sim	Não	ASM	67,17%	71,66%
Sim	Não	Manual	66,38%	72,59%
Sim	Histograma	ASM	67,87%	73,14%
Sim	Histograma	Manual	66,34%	73,88%
Sim	LogAbout	ASM	64,62%	65,03%
Sim	LogAbout	Manual	66,15%	67,77%
Sim	Log	ASM	69,95%	75,00%
Sim	Log	Manual	68,70%	75,92%

A Tabela 4 demonstra que a marcação manual e posteriormente uma normalização geométrica com os pontos marcados corretamente contribui para o aumento da taxa de acerto da classificação de gênero.

Ao analisar as tabelas anteriores, percebe-se que ao aumentar o número de imagens no conjunto de treinamento, o método piorou em relação ao treinamento feito com menos imagens.

Posteriormente, para aumentar o conjunto de imagens pré-rotuladas, um novo teste foi realizado onde cada imagem de entrada foi comparada com todas as outras. Os resultados obtidos podem ser analisados na Tabela 6.

**Tabela 6.** Taxas de acerto para o método de classificação de gênero um-para-um com normalização geométrica por marcação manual e com comparação entre todas as imagens.

<b>Detecção de Faces</b>	<b>Normalização Fotométrica</b>	<b>Normalização Geométrica</b>	<b>Acertos(%)</b>
Não	Não	Não	79,44%
Sim	Não	Não	85,74%
Sim	Histograma	Não	85,74%
Sim	Log	Não	85,92%
Sim	LogAbout	Não	77,59%
Sim	Não	Sim	86,48%
Sim	Histograma	Sim	85,00%
Sim	Log	Sim	86,66%
Sim	LogAbout	Sim	79,07%

Os resultados obtidos são comparáveis com os obtidos em trabalhos descritos na literatura, como Fahim (2007), que obteve uma taxa de acerto próxima a 90% utilizando PCA, ICA e SVM em um conjunto de 100 imagens. Ylioinas *et al.* (2011) que utilizaram variações do LBP somadas a informações de contraste obteve taxa de acerto de 91,53% no banco de dados FERET e Jain *et al.* (2005) que conseguiu alcançar uma taxa de acerto de 95,67% utilizando ICA e SVM para um conjunto de 500 imagens do banco FERET.

### 5.2.3 Reconhecimento de faces

Quanto ao reconhecimento de faces, os testes foram realizados sobre os bancos de faces Yale A (Georghiades *et al.*, 2001), Yale B(Georghiades *et al.*, 2001), GTAV (Tarrés e Rama) e VSOF-T-Faces (Marinho,2012).

Os testes foram realizados sobre conjuntos de indivíduos visando reconhecer apenas os que já são conhecidos previamente pelo classificador e utilizando os conceitos de validação cruzada e de verificação (um-para-um).

Para cada banco, após a etapa de detecção de faces, foram aplicadas técnicas de normalização geométrica e fotométrica com o intuito de analisar o impacto destas no

reconhecimento. Inicialmente, para encontrar a posição dos olhos requeridos pela normalização geométrica foi utilizado o algoritmo ASM.

Testes utilizando a abordagem um-para-um, que avalia o desempenho do método em relação à verificação, foram realizados. Variando o limiar escolhido previamente, pode-se gerar um gráfico que mostra as curvas que representam a Taxa de Falsa Rejeição (*FRR*), a Taxa de Falsa Aceitação (*FAR*) e o ponto de encontro destas curvas, Taxa de Erro Igual (*EER*). Os *EER*'s resultantes dos testes são apresentados na Tabela 7.

**Tabela 7.** *EER* para reconhecimento facial (1x1) com detecção utilizando pontos do ASM

<b>Detecção de Faces</b>	<b>Normalização Fométrica</b>	<b>Normalização Geométrica</b>	<b>GTAV</b>	<b>VSOFT</b>	<b>YALEA</b>	<b>YALEB</b>
Não	Não	Não	28%	17%	26%	8%
Não	Log	Não	30,9%	16,1%	30%	12%
Não	LogAbout	Não	27%	30%	35%	19%
Não	Histograma	Não	31%	32%	33,5%	9,0%
Sim	Não	Sim	32%	42,5%	36,5%	34%
Sim	Log	Sim	32%	41,3%	36,4%	33,5%
Sim	LogAbout	Sim	30%	39,3%	36,7%	32,4%
Sim	Histograma	Sim	34%	40%	37,5%	32,7%

Diante dos resultados apresentados na Tabela 7, percebe-se que o método proposto possui uma taxa elevada para o valor de *EER* e que quando utilizou-se a etapa de normalização geométrica o método piorou a taxa de acerto quando comparada com o não uso da etapa de normalização geométrica.

Após os testes um-para-um, foram realizados os testes utilizando o conceito de validação cruzada e a Tabela 8 mostra os resultados obtidos quando apenas uma imagem é utilizada para testes e o restante para treinamento. Para calcular a taxa de acerto do método de reconhecimento de faces foi utilizada a Equação 15.

**Tabela 8.** Taxa de acerto do reconhecimento facial com apenas uma imagem de teste por indivíduo

<b>Detecção de Faces</b>	<b>Normalização Fométrica</b>	<b>Normalização Geométrica</b>	<b>GTAV</b>	<b>VSOFT</b>	<b>YALEA</b>	<b>YALEB</b>
Não	Não	Não	83,1%	100%	96,7%	99,6%
Não	Log	Não	90,7%	94,9%	81,1%	100%
Não	LogAbout	Não	89,5%	93,6%	76,6%	100%

Não	Histograma	Não	89,5%	100%	76,6%	100%
Sim	Não	Sim	66,8%	70,5%	81,1%	83,3%
Sim	Log	Sim	66,8%	70,5%	82,5%	84,6%
Sim	LogAbout	Sim	78,5%	75,6%	77,9%	85,4%
Sim	Histograma	Sim	65,1%	69,2%	79,9%	83,7%

Analisando a Tabela 8 percebe-se que o método proposto obteve, em alguns casos, taxas de acerto elevadas e que na maioria dos testes a normalização geométrica piorou a taxa de acerto alcançada.

Diante da preocupação de que métodos atuais devam ser eficientes tendo poucas imagens no banco de dados de treinamento, foram realizados testes *leave-one-out*, onde para cada rodada apenas uma imagem de cada indivíduo é utilizada no treinamento e o restante das imagens é utilizado nos testes. Os resultados podem ser vistos na Tabela 9.

**Tabela 9.** Taxa de acerto do reconhecimento facial com apenas uma imagem de treinamento por indivíduo

Detecção de Faces	Normalização Fotométrica	Normalização Geométrica	GTAV	VSOFT	YALEA	YALEB
Não	Não	Não	53,3%	91%	70%	88,5%
Não	Log	Não	58,3%	76,4%	55,6%	88,9%
Não	LogAbout	Não	65,5%	75,1%	52,7%	83,3%
Não	Histograma	Não	60,2%	71,3%	53,4%	89,1%
Sim	Não	Sim	32,9%	37,9%	42,8%	40,2%
Sim	Log	Sim	34,1%	39,5%	43,8%	40,6%
Sim	LogAbout	Sim	46,9%	44,6%	35,8%	45,3%
Sim	Histograma	Sim	33,3%	42,8%	43,3%	40%

Na Tabela 9 pode-se perceber que ao utilizar apenas uma imagem para o treinamento o método não conseguiu reproduzir as taxas alcançadas com um conjunto de treinamento maior, como observa-se na Tabela 8.

Ao detectar a face, a imagem é recortada retirando todo *background*. Um fato que pode ser observado através das tabelas é que ao perder a informação do *background*, o classificador proposto sofre uma queda significativa na taxa de acerto. Portanto, para efetuar o recorte é necessário que uma boa normalização geométrica seja realizada para que as faces fiquem uniformes. Porém, ao observar as imagens

normalizadas geometricamente, algumas apresentaram uma normalização com anormalidades como demonstra a Figura 32. Essas anormalidades nas faces normalizadas podem ter afetado o reconhecimento piorando a taxa de acerto como observado nas tabelas anteriores.



**Figura 32.** Imagens normalizadas geometricamente pelos pontos extraídos do ASM

Para avaliar de forma mais precisa o método de reconhecimento proposto e retirar os erros ocasionados pela falha na detecção de olhos e posterior normalização geométrica, uma marcação de pontos nos olhos foi feita manualmente no banco YALE A. Após a marcação destes pontos, a etapa de normalização geométrica foi refeita com os pontos dos olhos nas posições corretas em todas as faces. A Figura 33 mostra as mesmas imagens da Figura 32 com a nova normalização.



**Figura 33.** Imagens normalizadas geometricamente pelos pontos extraídos manualmente

A Tabela 10 mostra os resultados obtidos após a nova normalização e comparados com os resultados anteriores. Os testes realizados foram utilizando uma imagem para testes e o restante para treinamento.

**Tabela 10.** Reconhecimento facial com etapa de normalização geométrica baseada em marcação manual dos olhos

<b>Normalização Fotométrica</b>	<b>Normalização Geométrica</b>	<b>YALEA</b>
Não	ASM	81,1%
Não	Manual	96,1%
Log	ASM	82,5%
Log	Manual	98,0%
LogAbout	ASM	77,9%
LogAbout	Manual	92,2%
Histograma	ASM	79,9%
Histograma	Manual	96,7%

Pela Tabela 10, percebe-se que ao efetuar a normalização geométrica através da marcação manual dos olhos, o método proposto alcançou taxas de acerto superiores às taxas alcançadas quando a normalização geométrica foi realizada através do pontos dos olhos encontrados pelo ASM. Testes com marcação manual serão iniciados nos outros bancos utilizados para o trabalho proposto.

Por fim, foi feito um teste onde as imagens do banco de dados VSOFTE-Faces (Marinho, 2012) foram utilizadas no treinamento e foram extraídas faces do vídeo interno gravado nas instalações da empresa para serem usadas no conjunto de testes. No conjunto de faces frontais para testes, foram excluídas todas as faces que possuíam tamanho inferior a 100 x 100 *pixels*, restando assim 681 faces para realizar o teste.

Para o reconhecimento em vídeo, os resultados obtidos podem ser vistos na Tabela 11.

**Tabela 11.** Taxa de acerto par reconhecimento facial em vídeo

<b>Normalização Fotométrica</b>	<b>Normalização Geométrica</b>	<b>Taxa de Acerto (%)</b>
Não	ASM	24,54546%
Log	ASM	26,72956%
Não	Manual	61,82085%
Log	Manual	59,32452%

A Tabela 11 apresenta uma taxa de acerto inferior às taxas encontradas no reconhecimento de faces realizado nos bancos de imagens de faces utilizados na

literatura, isto pode ser explicado pelo fato das imagens de faces retiradas dos vídeos possuírem uma maior variação de pose e *background* e também efeito de *blur* devido ao movimento dos indivíduos. Também observa-se que o ASM utilizado para encontrar o ponto dos olhos não conseguiu obtê-los de maneira correta, prejudicando a normalização geométrica e ocasionando uma taxa de acerto inferior a taxa alcançada quando os olhos foram marcados manualmente.

#### 5.2.4 Rastreamento de Faces

Para analisar o efeito do método proposto para rastreamento de faces, os testes foram realizados sobre o banco de vídeo gravado internamente, assim como no banco de dados *Incremental Learning for Visual Tracking* (Ross *et al.*, 2007). Apenas as faces frontais são buscadas nos vídeos.

Inicialmente, um teste para mensurar a importância do rastreamento de faces foi realizado. Para cada vídeo, a detecção de faces foi aplicada em cada quadro, contabilizando a quantidade de faces que são encontradas. Posteriormente, o rastreamento de faces é aplicado no vídeo e através da reidentificação de faces, é possível identificar se a face presente no quadro que está sendo analisado pertence a um novo indivíduo ou se ela pertence a indivíduo que esteve presente em quadros anteriores. A Tabela 12 demonstra a importância do rastreamento para reunir informações que são perdidas caso a análise seja realizada separadamente para cada quadro.

**Tabela 12.** Rastreamento de faces x faces isoladas

<b>Vídeo</b>	<b>Quantidade de faces Encontradas</b>	<b>Quantidade de Indivíduos Marcados Manualmente na Cena</b>	<b>Quantidade de Indivíduos Distintos Encontrados na Cena</b>
Vídeo Interno 01	164	14	13
Vídeo Interno 02	1572	13	21
<i>Outdoor sequence</i> (Ross <i>et al.</i> 2007)	315	1	1

<i>Indoor sequence</i> (Ross et al. 2007)	407	1	3
---	-----	---	---

Uma característica dos vídeos utilizados é a existência de indivíduos utilizando óculos e estes aparecem na cena em momentos distintos com e sem a presença dos óculos. O método proposto considerou, em algumas ocasiões, estes indivíduos como sendo distintos devido à presença dos óculos, e por este motivo aumentou a quantidade de indivíduos encontrados na cena quando comparado com o número real de indivíduos presentes.

Outro teste foi realizado para analisar o método proposto. A cada face encontrada, esta é comparada com as faces já conhecidas até o momento em que ela foi encontrada e caso ela seja considerada uma nova face, de acordo com o método proposto, que utiliza informação espacial e o LBPH, é atribuído a ela um novo identificador. Considerando que uma marcação prévia em cada face foi definida manualmente, para cada atribuição, foi considerado acerto caso o método tenha atribuído conforme o identificador, e erro, caso contrário. A Tabela 13 apresenta a taxa de acerto alcançada nos teste de reidentificação de faces.

**Tabela 13.** Taxa de acertos para reidentificação de faces no rastreamento.

<b>Vídeo</b>	<b>Faces Encontradas</b>	<b>Indivíduos Distintos na Cena</b>	<b>Taxa de Acertos(%)</b>
Vídeo Interno 01	164	14	90,96%
Vídeo Interno 02	1572	13	61,00%
<i>Outdoor sequence</i> (Ross et al. 2007)	315	1	88,88%
<i>Indoor sequence</i> (Ross et al. 2007)	407	1	59,00%

Ao analisar os resultados, percebe-se que grande parte dos erros deve-se à lógica de comparação do método, pois como, inicialmente, a comparação é feita utilizando-se

apenas a posição espacial das faces do quadro anterior com as posições da face do quadro atual, caso a face seja considerada uma nova face por esta etapa e o reconhecedor de faces considere que esta face não pertence ao banco de faces conhecidas, um novo identificador é atribuído a face atual. Porém, ao atribuir este novo identificador, todas as faces deste indivíduo que sejam reconhecidas posteriormente apenas pela posição espacial, terão o identificador distinto do que lhe foi atribuído anteriormente. Este erro pode ser ilustrado na Figura 34.



**Figura 34.** Ilustração da falha no método de rastreamento  
Fonte: Faces retiradas do banco de Learning for Visual Tracking Dataset (Ross *et al.*, 2007).

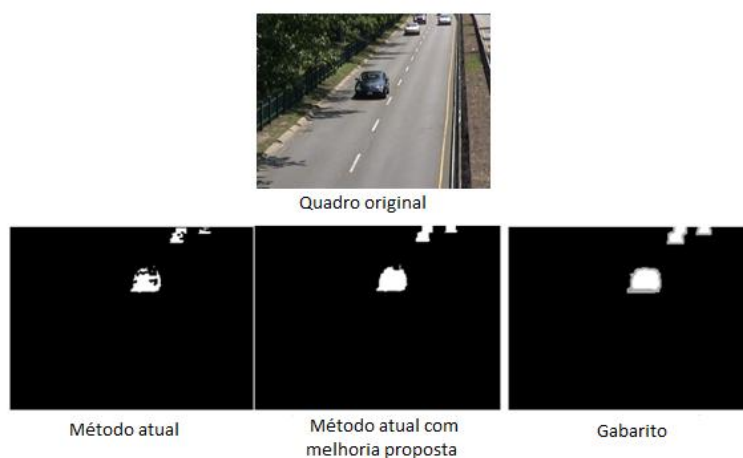
Para tornar o método robusto ao problema encontrado, pode-se realizar uma fusão da informação da posição espacial com o resultado do classificador para todas as faces.

## 6. Considerações Finais e Trabalhos futuros

Tecnologias de análise de cenas para interpretar conteúdos e extrair características de maneira automática estão se tornando cada vez mais relevantes para o cotidiano. Métodos computacionais que analisem e interpretem movimentações em cenas e que consigam extrair informações da presença de humanos ou até mesmo reconhecer características humanas são particularmente úteis em vários cenários.

O presente trabalho teve como objetivo geral pesquisar, desenvolver e avaliar métodos computacionais que analisem dados relevantes de cenas, vídeos ou imagens, extraíam características das informações visuais com o intuito de auxiliar procedimentos na Ciência Forense ou em levantamento de dados. Os métodos propostos foram desenvolvidos e contribuíram para que ao longo do trabalho uma sumarização pudesse ser realizada nas cenas.

Durante os testes, algumas considerações quanto ao trabalho puderam ser feitas. Quanto à detecção de movimentação, o método proposto mostrou-se eficiente e comparável com métodos existentes na literatura. Um artifício que pode melhorar as taxas das métricas avaliadas neste trabalho é a aplicação de um pós-processamento para auxiliar na segmentação dos objetos, como exemplo, após a detecção de movimentação do método. Estudos já estão sendo realizados aplicando-se métodos de dilatação e erosão nas imagens para que os objetos encontrados em um determinado quadro possam ser melhor definidos. A Figura 35 demonstra como esta melhoria futura poderá ajudar a segmentação dos objetos no banco de referência para detecção de mudanças em quadros (Goyette *et al.*, 2012).

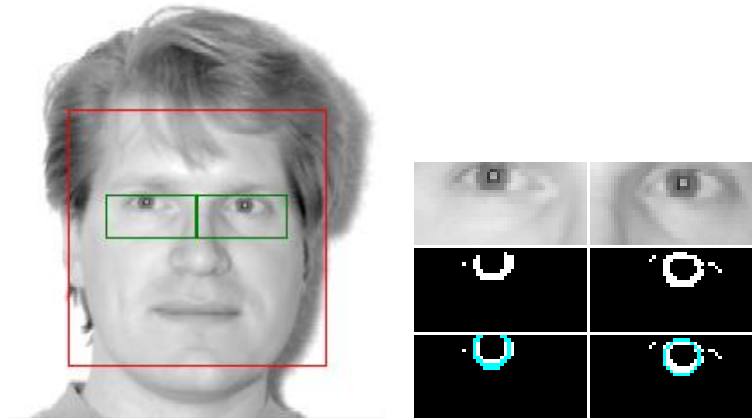


**Figura 35.** Pós-processamento para segmentação de objetos em quadros de movimentação

Analisando o reconhecedor de cor predominante no vestuário de um indivíduo, percebe-se que o método possui limitação quanto às camisas listradas ou com estampas, e também sofre influência da luminosidade do ambiente. Trabalhos futuros podem ser iniciados para estudar a correção destas falhas, detectando mais de uma cor predominante ou até mesmo com descrições mais precisas, como camisa com listras verticais, assim como uma melhor segmentação para separar o vestuário, podendo assim analisar uma região maior para definir a cor predominante.

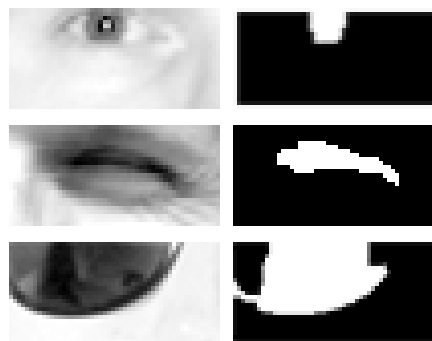
Para o reconhecimento de gênero e de faces, como trabalho futuro propõe-se testar outras variações do LBP que se enquadrem no reconhecimento de faces, testar a variação de seus parâmetros (P,R) e testar novos métodos de comparação dos histogramas do LBPH para comparar com o método de correlação que está sendo utilizado.

Algumas dificuldades foram encontradas e podem ser mencionadas como o problema na normalização geométrica das faces. A normalização geométrica depende de pontos que possam fornecer a rotação da face e sua escala, para isto, o trabalho propôs utilizar os pontos dos olhos. Entretanto, após testes com pontos marcados manualmente, percebeu-se que o método utilizado para a detecção desses pontos, *Active Shape Models* (ASM), mostrou falhas na detecção dos pontos característicos da face que influenciaram na normalização, ocasionando erros. Para corrigir estes problemas, um método para detecção de olhos já está sendo estudado e implementado. O método utiliza alguns conceitos bastante utilizados na literatura (Lin e Yang, 2004) (Ito *et al.*, 2012) (Khairoufa e Nor'aini, 2009) e consiste, basicamente, em dois passos: Inicialmente, a região dos olhos é encontrada a partir da face utilizando conceitos de proporção. Ao encontrar a região dos olhos, um filtro de bordas é aplicado e, posteriormente, aplica-se a "Transformada de Hough" para procurar por círculos na região dos olhos (Figura 36).



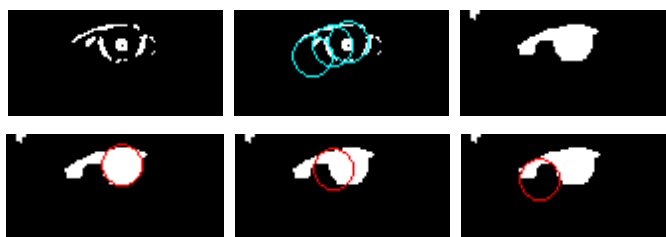
**Figura 36.** Transformada de Hough sobre região dos olhos

Após encontrar estes círculos, alguns filtros são aplicados para validar o círculo que representa a íris. Aplica-se uma transformação logarítmica na imagem para normalizar a iluminação, em seguida uma expansão de histograma é aplicada com o intuito de realçar a íris dos olhos, posteriormente o filtro negativo é aplicado seguido de uma limiarização. Após estes processamentos, a imagem da região do olho seguirá alguns padrões como pode ser visto na Figura 37.



**Figura 37.** Padrões encontrados após aplicação de filtros de pré-processamento

Após o processamento da imagem, caso a imagem se enquadre no padrão que foi encontrado para olhos abertos, cada círculo encontrado é analisado e então é atribuído como íris aquele que esteja sobre uma região predominantemente branca, caso ela exista(Figura 38). Caso a região do círculo não seja predominantemente branca, o detector proposto considera que a íris não está visível.



**Figura 38.** Exemplo onde o método encontra 3 círculos e verifica qual é mais predominantemente branco

Algumas melhorias ainda são necessárias para tratar imagens com algumas características como óculos de grau e olhos parcialmente fechados. Apesar de não estar concluído, alguns testes já foram aplicados para saber a evolução deste novo método em relação a encontrar a posição dos olhos. Como alguns bancos já haviam sido marcados manualmente com a posição dos olhos para cada face, foi feita uma comparação entre o desvio padrão da posição dos olhos encontrada pelo ASM e a posição dos olhos encontrada pelo método em desenvolvimento. Os resultados podem ser analisados na tabela abaixo.

**Tabela 14.** Tabela comparativa entre o ASM e o método proposto em desenvolvimento

<b>Banco de Dados</b>	<b>Deteção de Olhos</b>	Desvio Padrão (em <i>pixels</i> )	Erro Médio ( em <i>pixels</i> )
Yale A	ASM	5,11	7,26
Yale A	Método em desenvolvimento	4,29	5,00
Vsoft	ASM	2,92	3,60
Vsoft	Método em desenvolvimento	2,07	1,86

Uma melhoria futura para o reconhecedor de gênero é variar o atual método de comparação, o *KNN* (*k-Nearest Neighbors*), para aceitar mais de um vizinho, ou seja, não observar apenas um vizinho mais próximo e sim vários. O método de reconhecimento utilizado não gera um modelo para cada gênero, e sim compara o LBPH da imagem de entrada com todos os outros LBPH já conhecidos, e atribui o gênero do LBPH mais parecido ao gênero da imagem de face de entrada. Portanto, ao usar o *KNN* a vantagem será que para atribuição de gênero serão avaliados os *k*-vizinhos mais próximos, podendo ser atribuído a face de entrada o gênero mais comum entre os *k* vizinhos.

Diante dos resultados expostos na seção anterior, pode-se observar que as técnicas de normalização fotométrica utilizadas, em muitos casos contribuíram para o melhoramento das taxas de acerto. Apesar disto, através dos testes realizados não foi possível determinar qual método de normalização pode ser considerado superior para a abordagem do problema proposto.

A partir do estudo desenvolvido neste trabalho, outro trabalho futuro que pode ser desenvolvido é utilizar os métodos propostos em conjunto para que novas aplicações possam surgir. Um exemplo de aplicação através da combinação dos métodos propostos é armazenar a trajetória de movimentação de um indivíduo em cenas com várias câmeras de vigilância, ou utilizar informação de gênero ou cor da camisa para melhorar taxas de acerto na reidentificação de pessoas ou reconhecimento de faces.

Quanto à ferramenta desenvolvida, testes iniciais para medir o tempo de gerar os gráficos de movimentação, assim como vídeos sumarizados, foram realizados obtendo em média menos de 40% do tempo do vídeo de entrada. A otimização dos métodos é fundamental para que testes mais robustos possam ser realizados, pois ainda é especialmente complexo extrair simultaneamente muitas características de um vídeo devido à grande quantidade de processamento envolvido. Portanto, utilização de algoritmos multi-tarefa, processamento em GPU, agrupamentos, entre outros recursos computacionais disponíveis, podem ser necessários para que a ferramenta se torne viável para o uso comercial. Ainda há desafios significativos na sumarização de vídeos com condições de captura não controladas, em grande parte devido a variações de iluminação, bem como oclusões e fundos confusos.

A interpretação automática destas cenas está se tornando cada vez mais necessária, pois o grande volume de dados que é gerado torna o processo de análise muito lento e muitas vezes humanamente impossível. Os métodos e a sumarização proposta neste trabalho podem auxiliar a análise de vídeos ou cenas concentrando o examinador apenas nas áreas de interesse facilitando assim sua interpretação e argumentação, o que pode ser útil para que os resultados esperados sejam obtidos mais rapidamente. A participação do examinador nas análises é fundamental, pois torna a investigação um processo de análise híbrido, combinando as habilidades e o conhecimento técnico do examinador com as capacidades do sistema automatizado.

## Referências

- Alves, F. F. (2007) Desenvolvimento de Aplicação Biométrica Para Reconhecimento de Impressão Digital Através de um Dispositivo Móvel. Trabalho de Conclusão de Curso, Universidade Federal da Bahia.
- Adam A., Rivlin E., Shimshoni I. e Reinitz D. (2008) "Robust real-time unusual event detection using multiple fixed-location monitors", *IEEE Trans. Patt Anal. Mach Intell.*.
- Alvy, R. Smith (1978) "Color gamut transform pairs", *Special Interest Group on Graphics and Interactive Techniques*, New York.
- Ahmad, T., Jameel, A. e Ahmad, B. (2011) "Pattern recognition using statistical and neural techniques", *International Conference on Computer Networks and Information Technology (ICCNIT)*, Abbottabad.
- Andaló, F. A. (2012) "Métodos de Visão Computacional aplicáveis à Ciência Forense", Tese de Doutorado, UNICAMP, Campinas - SP.
- Ahonen, T., Hadid, A. e Pietikäinen M. (2004). "Face Recognition with Local Binary Patterns", 8th *European Conference on Computer Vision (ECCV 2004)*, .
- Ardakany, A. R . e Joula, A. M. (2012) "Gender Recognition Based on Edge Histogram", *International Journal of Computer Theory and Engineering*,
- Black, M. J. e Jepson, A. D. (1996) EigenTracking: Robust matching and tracking of articulated objects using a view-based representation, *European Conference on Computer Vision (ECCV)*.
- Bradski, G. R. e Kaehler, A. (2008), "A Learning OpenCV: Computer Vision with the OpenCV Library", O'REILLY, 1th edition.
- Cascia, M. La, Sclaroff, S. e Athitsos, V. (2000) "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Chacon-Rodriguez, A., Julian, P., Castro, L., Alvarado, P. e Hernandez, N. (2011) "Evaluation of Gunshot Detection Algorithms", *IEEE Transactions on Circuits and Systems I: Regular Papers*.

- Chakraborty, M. e Fuentes, O. (2009) "Real-Time Image-Based Motion Detection Using Color and Structure", *Proceedings of the 6th International Conference on Image Analysis and Recognition (ICIAR '09)*.
- Chakravarthy, A. S. N. e Kumar S. T. V. (2012) "Survey on Computer Crime Scene Investigation Forensic Tools", *International Journal of Computer Trends and Technology*.
- Chang, Chih-Chung e Lin, Chih-Jen (2001) "LIBSVM : a library for support vector machines", *Software* disponível em <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- Chen, Chao-Ho, Chen, Tsong-Yi, Lin, Yu-Chun e Hu, Wu-Chih. (2014) "Moving-Object Intrusion Detection Based on Retinex-Enhanced Method", *Tenth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Kitakyushu.
- Chen, Cheng-Yao, Abdallah, A. e Wolf, W. (2006) "Audiovisual Gunshot Event Recognition", *IEEE International Conference on Systems, Man and Cybernetics*, Taipei.
- Chu, W. T. , Cheng, W. H. e Wu, J. L. (2005) "Generative and discriminative modeling towards semantic context detection in audio tracks", *Proceedings of IEEE International Conference on Multimedia Modeling*.
- Clavel, C., Ehrette, T. e Richard, G. (2005) "Events detection for an audiobased surveillance system", *Proceedings of IEEE International Conference on Multimedia and Expo*.
- Comaniciu, D, Ramesh, V. e Meer, P. (2003) "Kernel-based object tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Cootes, T.F., Taylor, C.J., Cooper, D.H. e Graham, J. (1995) "*Computer Vision and Image Understanding*".
- Costa, A. Fonseca (2013) "Classificação de pedestres em imagens degradadas", Dissertação de mestrado, Universidade Federal de Pernambuco.
- Dalal, N. e Triggs, B. (2005) "Histograms of oriented gradients for human detection", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, San Diego, CA, USA.

- Delp, E., Memon, N., e Wu, M. (2009) "Digital Forensics [From the Guest Editors]", *IEEE Signal Processing Magazine*.
- Fahim Mannan (2007) "Classification of Face Images Based on Gender using Dimensionality Reduction Techniques and SVM", *School of Computer Science, McGill University*. link acessado em 21 de outubro de 2013: [www.cim.mcgill.ca/~fmannan/comp652/comp652\\_report\\_FM.pdf](http://www.cim.mcgill.ca/~fmannan/comp652/comp652_report_FM.pdf).
- Franke, K. e Shihari, S. (2008) "Computational forensics: An Overview", *Computational Forensics*.
- Franke, K. e Shihari, S. (2007) "Computational Forensics: Towards Hybrid-Intelligent Crime Investigation", *Third International Symposium on Information Assurance and Security*, Manchester.
- Felzenszwalb P. F., Girshick R. B., McAllester D. A. e Ramanan D., (2010) "Object detection with discriminatively trained part-based models", *IEEE Trans. Patt Anal. Mach Intell.*
- Gonzalez, Rafael C. e Woods, Richards E. (1992) "Digital Image Processing". *Prentice Hall*, New Jersey.
- Goyette, N., Jodoin, P. -M., Porikli, F., Konrad, J., e Ishwar, P. (2012) "changedetection.net: A new change detection benchmark dataset", *Proc. IEEE Workshop on Change Detection (CDW-2012)*, Providence, RI.
- Haupt, Alexandre Gasparly (2004) "Detecção de Movimento, Acompanhamento e Extração de Informações de Objetos Móveis", Dissertação de mestrado, Universidade Federal do Rio Grande do Sul.
- Huwer, S. e Niemann, H. (2000) "Adaptive change detection for real-time surveillance applications", *In Third IEEE International Workshop on Visual Surveillance*, Dublin, Ireland.
- Isard, M. and A. Blake. (1996) "Contour tracking by stochastic propagation of conditional density", *European Conference on Computer Vision*.
- Ito, Yasutaka; Wataru Ohyama, Tetsushi Wakabayashi, Fumitaka Kimura (2012) "Detection of Eyes by Circular Hough Transform and Histogram of Gradient", *21st International Conference on Pattern Recognition (ICPR 2012)*, Tsukuba, Japan.

- Jain, A., Huang, J. e Fang, S., (2005) "Gender Identification Using Frontal Facial Images", *Proc. IEEE Int'l Conf. Multimedia and Expo*.
- Jerian M., Paolino S., Cervelli F., Carrato, S., Mattei A. e Garofano L. (2007) "A forensic image processing environment for investigation of surveillance video", *Forensic Science International*.
- Katkar, V., Kulkarni, S. e Bhatia, D. (2015) "Traffic Video Classification using edge detection techniques", *International Conference on Nascent Technologies in the Engineering Field (ICNTE-2015)*, Navi Mumbai.
- Kazienko, J. F. (2003) "Assinatura Digital de Documentos Eletrônicos Através da Impressão Digital", Dissertação de mestrado, Universidade Federal de Santa Catarina.
- Khairoufaizal, W. e Nor'aini, A. (2009) "Eye detection in facial images using circular Hough transform," *Proc. Colloquium of Signal Processing and its Application (CSPA)*.
- Kim, M., Kumar, S., Pavlovic, V. e Rowley, H. (2008) "Face tracking and recognition with visual constraints in real-world videos", *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognition*.
- Koschan, A. e Abidi, M (2008) "Digital Color Image Processing", John Wiley & Sons.
- Li, Y., Kittler, J., e Matas, J. (2000) "On matching scores of LDA-based face verification", *Proceedings of the British Machine Vision Conference*. British Machine Vision Association.
- Li, J., Li, R., Mohedano R., del Blanco, C. R., Cuevas, C., Gallego, G., e Garcia, N. (2015) "Real-time Surveillance Application by Multiple Detectors and Compressive Trackers", *IEEE International Conference on Consumer Electronics (ICCE)*, Las Vegas.
- Lin, D. e Yang, C. (2004) "Real-time eye detection using face-circle fitting and darkpixel filtering", *Proceedings of IEEE International Conference on Multimedia and Expo*.
- Liu, H., Miao, J., Li, J. *et al.* (2002) "A Novel Method to Compensate Variety of Illumination in Face Detection", *6<sup>th</sup> Joint Conference on Information Sciences (ICCVPRIP)*.

- Lu, N., Wang, J., Wu, Q.H. e Yang, L. (2008) "An Improved Motion Detection Method for Real-Time Surveillance", *Proc. of International Journal of Computer Science*, IAENG.
- Maltoni, D., Maio, D., Jain, A. e Prabhakar S. (2009). "Handbook of Fingerprint Recognition", *Editora Springer*.
- Matos, F. M. S. (2008). "Reconhecimento de Faces Utilizando Seleção de Coeficientes da Transformada Cosseno Discreta", Dissertação de mestrado, Universidade Federal da Paraíba, João Pessoa.
- Marinho, A. S. (2012). "Uma nova versão de um sistema de detecção e reconhecimento de face utilizando a transformada cosseno discreta", Dissertação de mestrado, Universidade Federal da Paraíba, João Pessoa.
- Marques Filho, Ogê e Vieira Neto, Hugo. (1999) "Processamento Digital de Imagens", Rio de Janeiro: Brasport.
- Mehran R., Oyama A. e Shah M., (2009) "Abnormal crowd behavior detection using social force model", *Proc. Computer Vision and Pattern Recognition*.
- Miner, M. e Park, D.C. (2004). "A lifespan database of adult facial stimuli", *Behavior Research Methods, Instruments, & Computers*.
- Ogale, N.A. (2006) "A survey of techniques for human detection from video", *Department of Computer Science*, University Of Maryland, College Park, USA.
- Ojala, T., Pietikäinen, M. and Harwood, D. (1996), "A Comparative Study of Texture Measures with Classification Based on Feature Distributions.", *Pattern Recognition*.
- Ojala, T., Pietikäinen, M. e Maenpää, T. (2002) "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns." *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Otoom, A. F., Gunes, H. e Piccardi, M. (2008) "Feature extraction techniques for abandoned object classification in video surveillance", *15th IEEE International Conference on Image Processing (ICIP 2008)*, San Diego.
- Pietikäinen, M., Hadid, A., Zhao, G. e Ahonen, T. (2011) "Computer Vision Using Local Binary Patterns", *Computer Imaging and Vision*, London.

- Phillips, P. J., Moon, H., Rizvi, S. A. e Rauss, P. J. (2000) "The FERET evaluation methodology for face recognition algorithm", *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Ren, Y. e Chua, Ching-Seng (2001) "Motion detection with non-stationary background", *Proceedings of the 11th International Conference on Image Analysis and Processing*, Palermo, Italy.
- Rodriguez, Yann (2006) "Face Detection and Verification using Local Binary Patterns", Tese de doutorado nº 3681, *Faculté Des Sciences Et Techniques De L'ingénieur*, Lausanne, EPFL.
- Ross, D., Lim, J., Lin, R.-S. e Yang, M.-H. (2007) "Incremental Learning for Robust Visual Tracking", *International Journal of Computer Vision*, Special Issue: Learning for Vision.
- Sodemann, A. A., Ross, M. P. e Borghetti, B. J. (2012) "A Review of Anomaly Detection in Automated Surveillance", *IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews*.
- Stauffer, C. e Grimson, W.E.L. (1999) "Adaptive background mixture models for real time tracking", *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*.
- Strauss, O., Sidibé, D. e Puech, W. (2012) "Quasi-continuous histogram based motion detection", *Technical Report*, LE2I.
- Tarrés, F. e Rama, A. "GTAV Face Database", avaliado em <http://gps-tsc.upc.es/GTAV/ResearchAreas/UPCFaceDatabase/GTAVFaceDatabase.htm>, Outubro 2013.
- Takayanagi, Y. e Katto, J. (2010) "Human Body Detection using HOG with Additional Color Features", *International Workshop on Advanced Image Technology*.
- Thornton, J., Baran-Gale, J., Butler, D., Chan, M. e Zwahlen, H. (2011) "Person Attribute Search For Large-Area Video Surveillance", *IEEE International Conference on Technologies for Homeland Security (HST)*, Waltham, MA.
- Turk, M. A. e Pentland, A. P. (1991) "Face Recognition Using Eigenfaces", *Proc. of the IEEE Computer Society Conference*.

- Vacca, John R. (2005) "Computer Forensics: Computer Crime Scene Investigation", *Second Edition*, Charles River Media, Boston - Massachusetts.
- Viola, P. e Jones, M. (2004) "Robust Real-time Face Detection", *International Journal of Computer Vision*.
- Xiao, T., Zhang, C. e Zha, H. (2015) "Learning to Detect Anomalies in Surveillance Video", *IEEE Signal Processing Letters*.
- Yao, Jian e Jean-Marc Odobez (2007) "Multi-Layer Background Subtraction Based on Color and Texture", *Visual Surveillance Workshop (CVPR-VS)*, Minneapolis.
- Yilmaz, A., Javed, O., e Shah, M. (2006). "Object Tracking: A Survey", *ACM Computing Surveys*.
- Ylioinas, J., Hadid, A. e Pietikainen, M. (2011) "Combining Contrast Information and Local Binary Patterns for Gender Classification", *Image Analysis*.
- Yoshinaga, S., Shimada, A., Nagahara, H. e Taniguchi, R. (2013) "Background Model Based on Intensity Change Similarity Among Pixels", *19th Japan-Korea Joint Workshop on Frontiers of Computer Vision*.