

UNIVERSIDADE FEDERAL DO AMAZONAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Reidentificação de Pessoas em Rede de Câmeras

Ícaro Oliveira de Oliveira

MANAUS, AMAZONAS
Junho de 2009

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

UNIVERSIDADE FEDERAL DO AMAZONAS
DEPARTAMENTO DE CIÊNCIA DA COMPUTAÇÃO
CURSO DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

Reidentificação de Pessoas em Rede de Câmeras

Ícaro Oliveira de Oliveira

Dissertação apresentada ao Programa de Pós-Graduação em Ciência Computação do Instituto de Ciências Exatas da Universidade Federal do Amazonas como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Orientador: Prof. Dr. José Luiz de Souza Pio

MANAUS, AMAZONAS

Junho de 2009

Dedico esta dissertação às minhas queridas avós Edith (*in
memorian*) e Amira (*in memorian*).

Resumo

Esta dissertação aborda o problema da reidentificação da presença de pessoas em um ambiente de câmeras interconectadas por um sistema de comunicação. O problema da reidentificação ou reaquisição consiste no processo de realização dinâmica da correspondência entre imagens extraídas por câmeras diferentes. A metodologia aqui desenvolvida utiliza técnicas de detecção de pontos de interesse e cor dos alvos por meio do espaço *wavelet*. A aplicação deste trabalho é importante para os modernos sistemas de segurança nos quais a identificação da presença de alvos no ambiente monitorado amplia a capacidade de atuação dos agentes de segurança em tempo real e fornece a localização de cada pessoa de forma imediata. Resultados satisfatórios foram alcançados através dos experimentos reais por meio dos pontos de interesse aplicados em grandes bases públicas de vídeos e imagens sintéticas com ruído.

Palavras-chaves: reidentificação, wavelet, rede de câmeras.

Abstract

This dissertation presents an approach to the people reidentification problem in a camera network system. The reidentification or reacquisition problem consists essentially on the matching process of images acquired from different cameras. This work is applied in a monitored environment by cameras. This application is important to modern security systems, in which the targets presence identification in the environment expands the capacity of action by security agents in real time and provides important parameters like localization for each target. We used the target's color and target's interest points with features for reidentification. The satisfactory results were obtained from real experiments in public video datasets and synthetic images with noise.

Keywords: features, reidentification, color

Agradecimentos

Primeiramente à Deus, fonte de Sabedoria e Vida.

Aos meus queridos pais Edmilson e Vera Lúcia pelo apoio, compreensão e sacrifício realizados para esta conquista. E principalmente ao meu irmão Matheus pelo companheirismo e paciência durante os difíceis períodos enfrentados.

À minha namorada Itelvina pela paciência, ajuda e compreensão nos momentos difíceis que surgiram durante esta jornada.

Ao Prof. Dr. José Luiz de Souza Pio, orientador desta dissertação, por todo empenho, sabedoria, compreensão e, acima de tudo, exigência. Gostaria de ratificar a sua competência, participação com discussões, correções, sugestões que fizeram com que concluíssemos este trabalho.

Aos coordenadores do Programa de Pós-Graduação em Informática da UFAM, Prof. Dr. Raimundo da Silva Barreto e Prof. Dr. Edleno Silva de Moura, pela oportunidade de crescimento, aprendizado, realização profissional e pessoal e pela confiança em mim depositada.

Ao Prof. Dr. José Luiz de Souza Pio, Prof. Dr. José Pinheiro de Queiróz Neto e Prof. Dr. Horácio Antônio B. Fernandes de Oliveira por aceitarem participar da Banca de Defesa desta Dissertação, proporcionando discussões e sugestões que servirão para crescimento, aprendizado e incentivo à pesquisa.

Ao Prof. Dr Paulo Renan pelo auxílio com a língua inglesa.

À secretaria do Programa de Pós-Graduação em Informática da UFAM, ao Júnio e à Nazaré por tornarem factíveis a realização deste trabalho sendo

prestativos, dedicados e competentes.

Ao CNPq, cujo auxílio financeiro permitiu a dedicação necessária ao trabalho.

Aos amigos Fred, Geraldo e Isaac, pelo companheirismo, amizade e paciência durante todo o período em que estivemos juntos, pelas dicas e ajudas durante vários trabalhos realizados durante o curso, amizade e crescimento profissional e pessoal proporcionado durante todo o período de realização do trabalho, aos demais colegas do Grupo de Visão Computacional e do Grupo de Pós-Graduação em Informática da UFAM.

À todos MUITO OBRIGADO!

Sumário

Dedicatória	i
Resumo	ii
Abstract	iii
Agradecimentos	iv
1 Introdução	1
1.1 Motivação	3
1.2 Justificativa	6
1.3 Objetivos	7
1.3.1 Objetivo Geral	7
1.3.2 Objetivos Específicos	7
1.4 Organização do Trabalho	7
2 Fundamentação Teórica	9
2.1 Sistema de Rede de Câmeras	9
2.1.1 Topologia de Rede em Malha	12
2.2 Reconhecimento de Padrões	13
2.3 Características Visuais	15
2.4 Sistemas de Cor	16

2.4.1	Modelo de Cor RGB (<i>Red, Green, Blue</i>)	17
2.4.2	Modelo de Cor HSV (<i>Hue, Saturation, Value</i>)	18
2.4.3	A Cor como Característica Visual	19
2.5	Pontos de Interesse	22
2.5.1	Integral de Imagem	23
2.5.2	Detecção dos Pontos de Interesse	23
3	Trabalhos Relacionados	25
3.1	Reidentificação por Padrões Biométricos	26
3.2	Reidentificação pela Aparência do Objeto	27
3.2.1	Característica Visual: Cor	28
3.2.2	Característica Estrutural: Pontos de Interesse	34
4	Descrição Funcional da Reidentificação	40
4.1	O Componente Funcional de Controle	41
4.1.1	Aquisição da Imagem	42
4.2	Componente Funcional de Reidentificação	42
4.2.1	Conversão de Modelos de Cor	43
4.2.2	Integração da Imagem	44
4.2.3	Detecção dos Pontos de Interesse	45
4.2.4	Construção do Espaço Escalar	48
4.2.5	Extração das Características	52
4.2.6	Reconhecimento	57
5	Resultados Experimentais	59
5.1	Especificação da Configuração Utilizada	59
5.2	Fixação e Ajuste de Parâmetros para Implementação	60

5.2.1	Conversão entre os Modelos	60
5.2.2	Detecção dos Pontos de Interesse	61
5.2.3	Extração das Características	63
5.3	Grupo 1: Avaliação da Metodologia usando Imagens Sintéticas	64
5.3.1	Base de Imagens Sintéticas	69
5.3.2	Avaliação da Metodologia com Inserção do Ruído Gaussiano somente na Imagem de Consulta	70
5.3.3	Avaliação da Metodologia com Inserção do Ruído Gaussiano em cada objeto contido na imagem	79
5.3.4	Avaliação da Metodologia com Inserção do Ruído Gaussiano na Imagem de Consulta e em cada Objeto da Imagem	88
5.4	Grupo 2: Avaliação da Metodologia com Vídeos Reais	95
5.4.1	Descrição da Base Pública de Vídeos Reais	96
5.4.2	Descrição da Base de Consulta	98
5.5	Grupo 3: Avaliação da Complexidade dos Algoritmos	103
5.5.1	Integral de Imagens	103
5.5.2	Detecção dos Pontos de Interesse	103
5.5.3	Orientação	104
5.5.4	Descritor dos pontos de interesse	104
5.5.5	Reconhecimento do Método de Casamento do Sinal Laplaciano com o Vetor de Características formado pelos Pontos de Interesse	104
5.6	Grupo 4: Avaliação da Comunicação entre Nós	105
5.6.1	Parâmetros da rede	105

5.6.2	Avaliação de Multi-Saltos	107
6	Conclusão	109
6.1	Trabalhos Futuros	110
6.2	Considerações Finais	111

Lista de Figuras

1.1	Sistema de vigilância que monitora vários ambientes através de diversas câmeras distribuídas. Figura obtida em http://images.quebarato.com.br/photos/big/F/8/A1AF8_2.jpg .	2
1.2	Esse operador realiza o registro das atividades e a troca das fitas para manter a vigilância de uma loja. Figura obtida em http://1.bp.blogspot.com/_AM1xfKiynQs/SLPx6KlLiI/AAAAAAAAA1Y/m9e4eVmCqGQ/s400/B43AF_2.jpg .	4
1.3	Vários operadores realizam a tarefa árdua de visualizar cada monitor a fim de manter a vigilância e a segurança de um grande ambiente com várias câmeras distribuídas. Figura obtida em http://gcnny.us/surveillance/surveillance2.jpg .	5
2.1	Esquema do sistema de câmeras distribuídas obtida em (Pio, 2006).	11
2.2	Uma rede infraestruturada em malha composta por diversos pontos de acessos, conhecidos como nós da rede. Figura obtida em http://www.vivasemfio.com/blog_images/mesh.jpg .	13

-
- 2.3 Sistema RGB representado em um cubo RGB, onde os vértices mostram as combinações resultantes no espaço de cor normalístico. Figura obtida em <http://www6.ufrgs.br/engcart/PDASR/imagem7.JPG>. 18
- 2.4 Modelo de Representação do sistema HSV para algumas cores. Figura obtida em <http://www6.ufrgs.br/engcart/PDASR/imagem8.JPG>. 18
- 2.5 Histograma representando a distribuição de cores do canal RGB dividido em 3 gráficos para os canais *Red*, *Green* e *Blue*. 21
- 3.1 No canto esquerdo superior, representa a localização do objeto selecionado. No canto superior direito, apresenta a pessoa que foi detectada. E abaixo representa as pessoas detectadas e os seus respectivos tempos. Figura obtida em Wei et al. (2004). 27
- 3.2 Apresenta o rastreamento do objeto utilizando 2 câmeras com sobreposição de campo. Figura obtida em Morioka et al. (2006). 29
- 3.3 Um diagrama esquematizado do funcionamento do ViSE. A busca é feita por características primitivas que são dadas como entrada por um operador humano e apresenta o conjunto de pessoas visualizadas pelas câmeras em rede que possuem aquelas características. Figura obtida em Park et al. (2006). 30

-
- 3.4 A imagem dada como consulta é mostrada na coluna da esquerda, e as outras colunas representam o resultado dos principais conjuntos de imagens trazidos de cada consulta. Para cada consulta as imagens trazidas corretamente estão destacadas por uma caixa. A terceira linha mostra um exemplo onde a correta correspondência não está presente. Figura obtida em Gheissari et al. (2006). 32
- 3.5 Interface do sistema robusto desenvolvido por Pham et al. (2007). A parte A representa os 3 humanos detectados. A parte B a localização no mapa. A parte C e D a busca pelas aparências similares dos humanos selecionados em A. E a parte E representa o resultado obtido. Figura obtida em Pham et al. (2007). 33
- 3.6 A primeira coluna representa a imagem dada como consulta. Depois, nessa imagem são extraídas as características e comparadas com o banco de imagens. As outras colunas representam as imagens do banco similares a imagem dada como consulta. Figura obtida em Sivic and Zisserman (2003). 35
- 3.7 Dado uma imagem são extraídas suas características. É gerado um descritor e realizado o casamento de padrões com as características extraídas de uma rede de câmeras organizadas pela estrutura de árvore gerado pelo KNN, resultando nas reidentificações. Figura obtida em Arth et al. (2007). 36

-
- 3.8 Visão esquemática da reidentificação de uma consulta, onde no meio representa a Árvore KD contendo os pontos de interesse das pessoas. Figura obtida em Hamdoun et al. (2008). 38
- 4.1 Esquema geral da metodologia proposta. O quadro pontilhado representa os agentes externos do sistema (operador e sistemas de câmeras), a elipse o procedimento utilizado no componente funcional de controle e os retângulos os procedimentos utilizados no componente funcional de reidentificação. 41
- 4.2 Cálculo da área usando integral de imagens. Figura obtida em Evans (2009). 45
- 4.3 Aproximação Laplaciana de Gaussiano. Linha de cima: Os discretizados e descartados derivadas Gaussianas de segunda ordem no \mathbf{x} , y e direções xy . Referentes ao L_{xx} , L_{yy} e L_{xy} . Linha de baixo: Aproximações do filtro de caixa com peso no \mathbf{x} , y e direções xy . Referem-se estes como D_{xx} , D_{yy} , D_{xy} . Figura obtida em Evans (2009). 47
- 4.4 Amostragens em Diversas Escalas. Utilizada pelo SURF, à Direita, Diferente da abordagem tradicional à esquerda. Figura obtida em Evans (2009) e adaptada. 49
- 4.5 Os tamanhos dos Filtros posteriores devem diferenciar-se por um mínimo de seis para preservar a estrutura do filtro. Figura obtida em Evans (2009). 50

-
- 4.6 Supressão não-maximal. O pixel marcado "X" é selecionado como o máximo se for maior que os pixels que o rodeiam em seu intervalo e nos intervalos acima e abaixo. Figura obtida em Evans (2009) e adaptada. 51
- 4.7 *Wavelets de Haar* representados pelos dois filtros. Figura obtida em Evans (2009). 52
- 4.8 A determinação da orientação é obtida pela soma das produções dos vetores aqui mostrados em azul. O maior desses vetores determina a orientação dominante. Figura obtida em Evans (2009). 54
- 4.9 Janelas do descritor. O tamanho da janela é 20 vezes a escala do ponto detectado e é orientado ao longo da direção dominante apresentada em verde. Figura obtida em Evans (2009). 55
- 4.10 Componentes do descritor. Figura obtida em Evans (2009). 56
- 5.1 Na esquerda está a imagem descrita no modelo RGB, ao seu lado a mesma em tons de cinza e no final a mesma descrita no modelo HSV. 61
- 5.2 Esses são exemplos da detecção dos pontos de interesse. A primeira imagem representa a imagem sintética de consulta, a segunda imagem é a imagem de consulta da base1 e a terceira imagem é a imagem de um quadro do vídeo da base2. 63
- 5.3 Na esquerda está a imagem descrita no modelo RGB e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 65

-
- 5.4 Na esquerda está a imagem descrita no modelo HSV e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 66
- 5.5 Na esquerda está a imagem em tons de cinza e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 67
- 5.6 Na esquerda está a imagem em tons de cinza com escala $\frac{1}{1.5} \times$ e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 68
- 5.7 Na esquerda está a imagem em tons de cinza com escala $\frac{1}{3.0} \times$ e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 68
- 5.8 Na esquerda está a imagem em tons de cinza com escala $\frac{1}{4.5} \times$ e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 69
- 5.9 Essa figura apresenta as imagens utilizadas para representar as pessoas no lado esquerdo e algumas imagens de carro no lado direito para dificultar a reidentificação correta. 70

-
- 5.10 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 71
- 5.11 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. . . 72
- 5.12 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 73
- 5.13 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 74
- 5.14 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. . . 75
- 5.15 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 76
- 5.16 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 77
- 5.17 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. . . 78

-
- 5.18 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 79
- 5.19 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 80
- 5.20 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. . . 81
- 5.21 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 82
- 5.22 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 83
- 5.23 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. . . 84
- 5.24 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 85
- 5.25 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 86

-
- 5.26 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. 87
- 5.27 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 88
- 5.28 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 89
- 5.29 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. 90
- 5.30 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 91
- 5.31 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 92
- 5.32 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. 92
- 5.33 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 93

-
- 5.34 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse. 94
- 5.35 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor. . . 94
- 5.36 Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse. 95
- 5.37 Esses quadros são seqüências extraídas dos vídeos da base ACTIONS, os quais apresentam diferentes ações e diversas deformações. 97
- 5.38 Essas imagens são seqüências obtidas dos vídeos da base CAVIAR. Essas medidas apresentadas nas figuras foram descartadas neste trabalho. 98
- 5.39 Essas imagens de consulta representam as pessoas da base1. São imagens extraídas de visão frontal da pessoa com pouca deformação. 99
- 5.40 Essas imagens de consulta representam as pessoas da base2. São imagens extraídas de visão frontal e traseira da pessoa com pouca deformação. E com visão panorâmica. 99
- 5.41 Esse gráfico apresenta a precisão do vetor 1 sobre o reconhecimento utilizando os parâmetros de limiar e dimensões do descritor. 100

-
- 5.42 Imagem apresentando uma reidentificação da consulta, onde o ambiente apresenta uma grande quantidade de pessoas e ruído. 101
- 5.43 Imagem apresentando uma reidentificação da consulta onde aparece somente a pessoa que se deseja reidentificar. 101
- 5.44 Imagem apresentando uma reidentificação da consulta onde aparece um conjunto de pessoas próximo a pessoa que se deseja reidentificar. 102
- 5.45 Imagem apresentando uma reidentificação da consulta onde aparece um conjunto de pessoas próximo a pessoa que se deseja reidentificar e essa pessoa possui oclusão. 103
- 5.46 As Figuras 5.6.1, 5.6.1 e 5.6.1 apresentam o tempo consumido em um envio entre dois nós para cada imagem enviada ou recebida variando o tempo de espera entre 10 ms a 50 ms. . . . 106
- 5.47 Apresentação dos tempos de envios variando os saltos. 108

Lista de Tabelas

- 2.1 Tabela do comprimento de onda de um espectro de luz. Tabela obtida em <http://www.if.ufrj.br/teaching/luz/cor.html> 17
- 2.2 Tabela com algumas das muitas combinações de cores possíveis com o sistema RGB. Obtido em <http://www.efeitosespeciais.net/tabela.htm>. 17
- 2.3 Seleção de nível de cores. Tabela obtida em <http://pt.tech-faq.com/hsv.shtml>. 19

Capítulo 1

Introdução

O problema abordado nesta dissertação é a reidentificação de objetos de interesse em um ambiente monitorado por uma rede de câmeras. Neste trabalho, reidentificar significa encontrar padrões de pessoas que trafegam em um ambiente monitorado por câmeras convenientemente distribuídas. Esse problema pode ser sumarizado por meio da seguinte pergunta:

Como reidentificar padrões visuais de pessoas em outras câmeras da rede, quando previamente identificados em uma câmera?

Em muitas aplicações de vigilância por vídeo, torna-se desejável determinar se uma pessoa atualmente visível já foi observada em outro lugar monitorado pela rede de câmeras. Uma apresentação completa sobre esta área de pesquisa pode ser encontrada, por exemplo, no trabalho apresentado em Tu et al. (2007). Os algoritmos de reidentificação devem ser robustos, mesmo em situações que possuem diferentes orientações da câmera, diferentes condições de iluminação, pessoas em movimento e mudanças rápidas nas aparências.

Geralmente, grandes ambientes como aeroportos, hospitais, shoppings, vias públicas, centros comerciais, estádios, parque de estacionamento (Collins et al., 1999, 2000; Black et al., 2004; Micheloni et al., 2005), eventos esportivos (Saito et al., 1999; Bruno Müller and de Oliveira Anido, 2004) ou qualquer outro ambiente público ou privado são monitorados por um conjunto

de câmeras que garantem a maior cobertura de monitoração do ambiente e minimizam os efeitos da oclusão, conforme apresentado na Figura 1.1.



Figura 1.1: Sistema de vigilância que monitora vários ambientes através de diversas câmeras distribuídas. Figura obtida em http://images.quebarato.com.br/photos/big/F/8/A1AF8_2.jpg.

O problema da reidentificação de pessoas envolve-se com a área de visão computacional. Devido aos avanços tecnológicos da área de Visão Computacional e das Redes de Computadores tornou-se viável a utilização de câmeras convenientemente interconectadas em um sistema de comunicação, a fim de monitorar diversas regiões em tempo-real (Matsuyama, 1999; Nakazawa et al., 1998).

Em sua forma mais clássica, uma imagem é uma aplicação com domínio em uma região do R^2 e com valores em um espaço de cores que corresponde a um modelo matemático do agregado de energia luminosa que atinge um plano de células fotosensores. Algumas características visuais podem ser extraídas desse modelo matemático tais como forma, textura, localização e vetores de movimento dos objetos que compõem uma cena.

Essas imagens podem ser obtidas de diferentes maneiras, como através de uma única câmera, uma seqüência de vídeos ou, como tratado neste trabalho, por múltiplas câmeras interligadas por meio de um sistema de comunicação.

A principal contribuição deste trabalho é a característica "inteligente" do processamento automático após um processo de reidentificação e de monitoramento. Com isso, pretende-se identificar os eventos e situações que requerem a atenção da segurança, aumentando a segurança no ambiente monitorado de modo eficiente e eficaz.

Diferente das abordagens atuais, a metodologia adotada neste trabalho utiliza a informação extraída a partir da decomposição do espaço de cores das pessoas e a detecção de **pontos de interesse** como características. Essa abordagem trabalha sobre o ambiente interno com câmeras com e sem sobreposição nos seus campos de visão, sem oclusão e não requer conhecimento topológico do ambiente ou treinamento dos padrões de pessoas.

1.1 Motivação

O crescimento na segurança e no controle de acesso de áreas protegidas têm favorecido um aumento considerável da informação visual nos sistemas de vigilância. A enorme quantidade de informação visual obtida na maioria desses sistemas têm provido grande impacto na segurança pública e privada. O monitoramento visual ainda é provido por meio de sistemas de circuito fechado de TV (CCTV), os quais não apresentam nenhuma característica "inteligente". Os operadores humanos desses sistemas ainda são centrais no

processo de decisão; permanecem com a responsabilidade de avaliar, planejar e executar todos os procedimentos, inclusive a tarefa árdua de reidentificar pessoas sobre uma enorme quantidade de informação visual obtida de diversas câmeras, conforme apresentado na Figura 1.2 (para um pequeno ambiente) e na Figura 1.3 (para um grande ambiente). Com isso, surge a necessidade de integrar a esses sistemas de CCTV uma característica inteligente capaz de auxiliar os operadores humanos sobre essa tarefa árdua de reidentificação de pessoas.



Figura 1.2: Esse operador realiza o registro das atividades e a troca das fitas para manter a vigilância de uma loja. Figura obtida em http://1.bp.blogspot.com/_AM1xfKiynQs/SLPx6KlflI/AAAAA-AAAA1Y/m9e4eVmCqGQ/s400/B43AF_2.jpg.



Figura 1.3: Vários operadores realizam a tarefa árdua de visualizar cada monitor a fim de manter a vigilância e a segurança de um grande ambiente com várias câmeras distribuídas. Figura obtida em <http://gcnyc.us/surveillance/surveillance2.jpg>.

As redes de vigilância possuem diversas câmeras que devem ser coordenadas a fim de identificar os alvos sobre grandes distâncias e períodos de tempo. A vigilância automatizada por câmeras captura em tempo-real os alvos em ambientes de difícil acesso e com ruídos. Isto significa que a solução do problema de reidentificação é muito importante.

Com isso, a motivação na implementação de soluções para o problema proposto está na possibilidade de agregar características inteligentes ao processo de monitoramento, viabilizando a utilização do operador humano para tarefas mais nobres e liberando-o de tarefas mais tediosas como a busca por

uma pessoa em uma grande coleção de vídeos de vigilância.

1.2 Justificativa

O aumento atual da violência urbana tem sido favorecida pelo desenvolvimento de sistemas capazes de identificar terroristas, traficantes de drogas, pessoas perdidas em grandes ambientes, vândalos, assaltantes, estupradores e assassinos são fatores marcantes e decisivos. Demandas dessa natureza justificam o estudo e a implantação de sistemas de reidentificação capazes de monitorar continuamente e otimizar a detecção e localização desses agentes.

A justificativa deste trabalho se concentra na necessidade de encontrar soluções que resolvam esses problemas presentes na sociedade. Dentro de um contexto científico e tecnológico, se propõe a estudar, investigar e desenvolver uma metodologia para a questão de reidentificação de pessoas em uma rede de câmeras distribuídas convenientemente sobre um ambiente monitorado. Essas pessoas podem ser identificadas em diferentes regiões desse ambiente, o que facilita e viabiliza o controle do fluxo das mesmas. O trabalho agrega a funcionalidade de reidentificação que é desprovida pela maioria dos sistemas atuais.

Muitas pesquisas têm sido desenvolvidas na área de visão computacional para sistemas de reidentificação de objetos de interesse (Sivic and Zisserman, 2003; Morioka et al., 2006) e de pessoas (Wei et al., 2004; Pham et al., 2007). Esses trabalhos têm focado a automatização desses sistemas tentando ampliar os sistemas restritos aos CCTV.

Na abordagem proposta são utilizadas técnicas simples como detecção, extração e correspondência de padrões de pessoas. Pretende-se, portanto, obter neste trabalho uma abordagem simples e eficiente que reidentifique em tempo real a presença de pessoas em um ambiente de múltiplas câmeras. O operador humano terá sempre atualizada a informação referente à posição atual de uma pessoa que foi visualizada pelas câmeras e não precisará mais reunir visualmente as informações de várias câmeras para a localização da

mesma de forma manual, difícil e tediosa.

1.3 Objetivos

Para melhor compreensão os objetivos desse trabalho foram divididos em geral e específico.

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é prover aos sistemas de segurança uma ferramenta para reidentificação da presença de uma determinada pessoa em um ambiente monitorado por uma rede de câmeras.

1.3.2 Objetivos Específicos

Como objetivos específicos, destacam-se:

- Implementação de uma metodologia de reidentificação de pessoas;
- Implementação de métodos para detecção e extração de pontos de interesse;
- Avaliação da precisão e do tempo de execução do método de correspondência de padrões por meio de experimentos reais aplicados em bases públicas de vídeo.
- Análise da robustez da metodologia sobre a escalabilidade de pessoas e variação no ruído sobre imagens sintéticas;
- Análise do tempo de comunicação entre dois nós de uma rede em malha;

1.4 Organização do Trabalho

O texto está organizado como segue. No Capítulo 2, é apresentada a fundamentação teórica sobre as abordagens e áreas de interesse relacionadas

que serão utilizadas no trabalho. No Capítulo 3, é apresentada uma revisão sobre as diferentes abordagens relacionadas ao problema de reidentificação da presença de pessoas em um ambiente de múltiplas câmeras. No Capítulo 4, é apresentada detalhadamente a metodologia aplicada. No Capítulo 5, são apresentados os resultados obtidos por meio de experimentos reais e simulações. No Capítulo 6, é apresentado a conclusão do trabalho.

Capítulo 2

Fundamentação Teórica

Neste capítulo serão abordados alguns temas e conceitos que permitirão entender melhor o desenvolvimento deste trabalho, como: técnicas de detecção de pontos de interesse, extração de características visuais e reconhecimento de padrões em um ambiente com câmeras distribuídas, os quais permitirão a compreensão das diferentes etapas do processo metodológico. Os fundamentos dessas técnicas estão apresentadas com detalhes em seguida.

2.1 Sistema de Rede de Câmeras

Atualmente existe uma grande variedade de aplicações nas quais um conjunto de câmeras interligadas é distribuída em um ambiente objetivando a visualização simultânea de diversas regiões de uma área de monitoramento. Pio (2006) mostra que o objetivo principal desses sistemas é adequar maior eficiência ao sistema de segurança por meio da combinação das técnicas de visão computacional com tecnologia de redes de computadores em aplicações de rastreamento de pessoas e veículos e como nesse trabalho aplicado em reidentificação de pessoas.

Em geral, os sistemas de rede de câmeras são projetados em circuito fechado de TV (CFTV) e interligados por uma rede com fio ou sem fio. Esses sistemas possuem resultados surpreendentes na identificação de suspeitos e

na detecção de atividades ilícitas. Porém, a grande maioria ainda deixa por conta do operador humano o processo de controle e tomada de decisão como apresentado no trabalho (Pio, 2006).

No trabalho de doutorado (Pio, 2006) descreve um sistema distribuído de câmeras para o rastreamento de características no espaço tridimensional conforme a Figura 2.1. Nesse sistema utiliza-se um procedimento de triangulação a partir de duas ou mais câmeras calibradas, onde essas câmeras possuem unidades de câmera de baixo custo que incluem uma pequena câmera CCD e dois mecanismos servo-controlados para rotações e posicionamentos fixos.

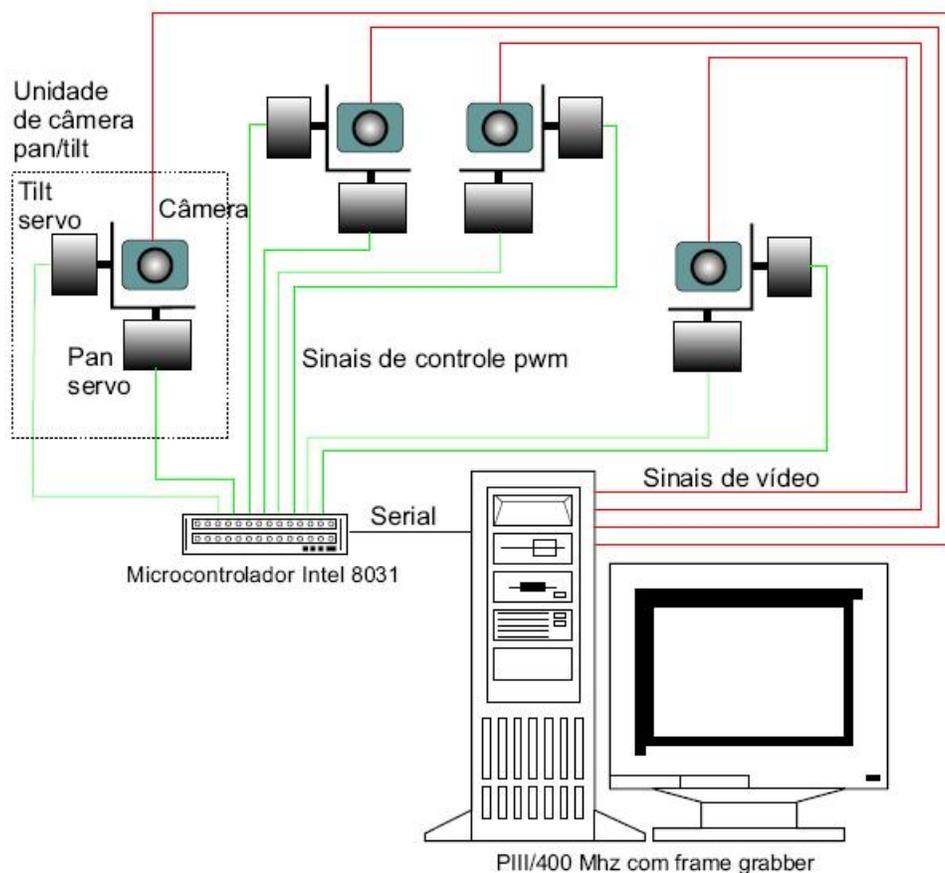


Figura 2.1: Esquema do sistema de câmeras distribuídas obtida em (Pio, 2006).

Uma outra forma de representar uma rede de câmeras também é encontrado em Pio (2006) como uma infraestrutura para áreas metropolitanas, uma vez que estas redes fornecem uma grande quantidade de informação visual, provida, na maioria dos casos, por um pequeno número de câmeras comuns.

O uso de mais de uma câmera para cobrir uma região muito grande, como um campo de futebol, aumenta consideravelmente a precisão de qualquer técnica de visão computacional, porém, leva a uma grande quantidade

de dados que devem ser processados e convenientemente associados. Essas câmeras são posicionadas objetivando maximizar a cobertura do campo de futebol e o rastreamento de objetos que se tornam pequenos no plano de projeção das câmeras, como apresentado em Pio (2006).

2.1.1 Topologia de Rede em Malha

A rede em Malha (Akyildiz et al., 2005) é uma alternativa de protocolo ao padrão 802.11 com uma rede de infraestrutura composta de Pontos de Acesso e clientes, os quais necessariamente devem utilizar o seu Ponto de Acesso para trafegarem conforme apresentado na Figura 2.2. Essa rede é composta de vários nós ou roteadores, que passam a se comportar como uma única e grande rede, possibilitando que o cliente se conecte em qualquer um desses nós. Os nós servem como repetidores, e cada nó está conectado a um ou mais dos outros nós. Desta maneira, é possível transmitir mensagens de um nó a outro por diferentes caminhos. A rede em Malha possui a vantagem de ser a rede de baixo custo, fácil implantação e bastante tolerante a falhas. Nessa rede, roteadores sem fio são tipicamente instalados no topo de edifícios e comunicam-se usando protocolos, como por exemplo, o *OLSR* em modo *ad hoc* através de múltiplos saltos de forma a encaminhar pacotes de dados aos seus destinos.

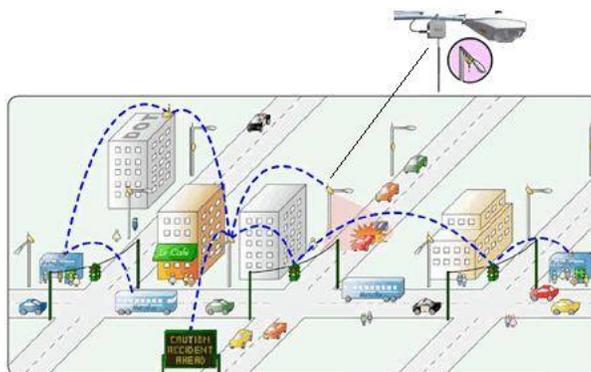


Figura 2.2: Uma rede infraestruturada em malha composta por diversos pontos de acessos, conhecidos como nós da rede. Figura obtida em http://www.vivasemfio.com/blog_images/mesh.jpg.

2.1.1.1 Cálculo do tempo de comunicação

Para calcular o tempo de envio de uma informação entre um nó a outro da rede utiliza-se a fórmula apresentado por Tanenbaum (2002) que consiste em

$$tempo = \text{tempo de espera} + \frac{\text{tamanho do envio}}{\text{tamanho da banda}}. \quad (2.1)$$

2.2 Reconhecimento de Padrões

O *Reconhecimento de Padrões* é uma disciplina científica que estuda e desenvolve técnicas com o objetivo de descrever, classificar e reconhecer regularidades significativas ou padrões, em meios ruidosos e complexos (Schalkoff, 1989).

O objetivo principal do reconhecimento de padrões é a *classificação*. O que se procura por meio da classificação é atribuir classes a objetos usando relações estabelecidas por meio de observações obtidas sobre estes últimos (Fu, 1982).

Em geral, os métodos de classificação diferem-se na forma de representação dos objetos, na maioria de como são definidas as regiões no espaço de

representação e na escolha de algum critério a ser minimizado baseado na amostra dada.

Historicamente, o reconhecimento de padrões inicialmente utilizou técnicas probabilísticas e da teoria da decisão estatística para sua fundamentação. Teve como fundamento a teoria das probabilidades, baseando-se na abordagem Bayesiana. Sob essa abordagem, um conjunto de medidas características (ou atributos) são extraídos dos padrões. Cada atributo extraído passa a ser visto como um ponto em um espaço n -dimensional. A idéia principal é dividir o espaço de características em regiões que agrupam objetos de mesma classe. A classificação é a busca pela região do espaço de características mais adequado para um dado padrão de entrada.

Como consequência, o problema de reconhecimento estrutural é tratado como um *problema de casamento de padrões*. Para ilustrar o *casamento de padrões*, considere, por exemplo, uma base de dados consistindo de cadeias de números, cada cadeia representando uma amostra de um padrão. A classificação de um padrão desconhecido é determinada pelo casamento entre o padrão dado como entrada e todos os protótipos contidos na base de dados. O objetivo final consiste em associar o padrão desconhecido a uma classe do protótipo mais similar da base de dados. Para determinar esta similaridade neste trabalho é utilizado o método de Soma das Diferenças Quadráticas (SDQ) para calcular a distância entre a cadeia do padrão desconhecido e a cadeia dos protótipos contidos na base de dados. Com isso,

$$\min_{x>0} SDQ = \sum_x^n (c_i(x) - c_q(x))^2 \quad (2.2)$$

onde c_i é a cadeia dos protótipos da base de dados, x é o índice de cada cadeia do protótipo, n é a quantidade de dimensões da cadeia e c_q é a cadeia do padrão desconhecido. Portanto, através do índice x do valor dessa formula será o protótipo mais similar do padrão desconhecido.

2.3 Características Visuais

A obtenção da informação visual é determinada por meio das características visuais de cada objeto contido na cena. Nesse contexto, as características visuais são definidas por Hutchinson et al. (1996) como medidas ou observações obtidas por uma câmera e são descritas como globais quando se obtém uma propriedade global da imagem, como o valor médio dos *pixels*, momentos ou o fluxo óptico. Por outro lado, são descritas como locais quando dizem respeito a subconjunto particular de *pixels*, como pontos, cantos ou segmentos de reta.

As características visuais são extraídas das imagens dos objetos tomados a partir de diferentes posições da câmera, de forma robusta e não ambígua. Um parâmetro de uma característica visual (Hutchinson et al., 1996; Simon et al., 1995) caracteriza-se como um valor real obtido a partir de uma ou mais características visuais. O comprimento de um segmento de reta, as coordenadas de um ponto, padrões de textura, cor, o valor dos semi-eixos de uma elipse e o raio da circunferência são exemplos de parâmetros de características visuais.

Dado um conjunto de n parâmetros de características visuais, defini-se um vetor $c = [f_1, \dots, f_n]^T$ como sendo o vetor de características visuais, onde $c \in \mathfrak{S} \subseteq R$ no qual \mathfrak{S} representa o espaço dos parâmetros de características visuais. Em geral, os parâmetros da característica visual são obtidos a partir de uma imagem de referência (Simon et al., 1995).

Dentre as abordagens para representar um objeto existe a abordagem que usa a informação sobre a luminosidade da imagem. A idéia é não impor o que tem de ser visto na imagem (pontos, linhas, etc) mas usar o que é realmente visto na imagem para caracterizar um objeto. Como a cor e pontos de interesse.

2.4 Sistemas de Cor

O sistema de cor consiste nos padrões que definem as propriedades de uma determinada cor. Por exemplo, o branco é composto por todas as outras cores, onde um espectro de luz visível, pode então assumir diversas cores (desde o violeta até o vermelho), em função do comprimento de onda observada na Tabela 2.1. Três características básicas de um espectro de luz são, a matiz (definida pelo comprimento da onda), brilho (intensidade da luz fonte) e saturação (predominância da matriz). Existem dois processos de formação de cores: aditivo e subtrativo. A formação do sistema de cores no processo aditivo é obtida a partir de duas fontes luminosas, onde, cores diferentes são projetadas em duas regiões; sendo assim, na interseção entre as cores há a formação de uma nova cor. Por outro lado, no processo subtrativo, é geralmente feito com a utilização de *slides*. Um exemplo tem-se quando emite-se uma luz branca (que possui todos os comprimentos de onda) sobre um filtro verde, o qual filtra todos os comprimentos de onda deixando passar somente o comprimento de onda relativa à cor verde, produzindo assim o verde. Dentro do sistema de cores, os padrões mais conhecidos são: *RGB* (*Red, Green, Blue*), *CMYK* (*Cyan, Magenta, Yellow, Black*), *HSV* (*Hue, Saturation, Value*), *RYB* (*Red, Yellow, Blue*), *Highcolor*, Imagem binária, *SRGB* (*standard RGB*), *Munsell*, entre outros. Neste trabalho limitaremos a explicação aos modelos *HSV* (*Hue, Saturation, Value*) e o *RGB* (*Red, Green, Blue*).

Tabela 2.1: Tabela do comprimento de onda de um espectro de luz. Tabela obtida em <http://www.if.ufrj.br/teaching/luz/cor.html>

Cor	Comprimento da Onda
Violeta	380-440 mm
Azul	440-490 mm
Verde	490-565 mm
Amarelo	565-590 mm
Vermelho	630-780 mm

2.4.1 Modelo de Cor RGB (*Red, Green, Blue*)

Formado pelas cores vermelha, verde e azul, o RGB é um padrão de cores utilizado em monitores/televisores, em tubos de raios catódicos, cristal líquido ou de plasma. Utiliza 8 bits de informação para representar cada uma de suas cores (vermelho, verde, azul), podendo variar de 0 a 255. É possível alcançar mais de 16 milhões de cores com suas combinações de tons. Algumas destas cores estão representadas na Tabela 2.2:

Tabela 2.2: Tabela com algumas das muitas combinações de cores possíveis com o sistema RGB. Obtido em <http://www.efeitosespeciais.net/tabela.htm>.

Cor	R	G	B
Preto	0	0	0
Azul	0	0	255
Verde	0	255	0
Ciano	0	255	255
Cinza	128	128	128

Na Figura 2.3, é possível observar o seu modelo mapeado em um cubo.

Onde os vértices extremos apresentam as cores principais do RGB.

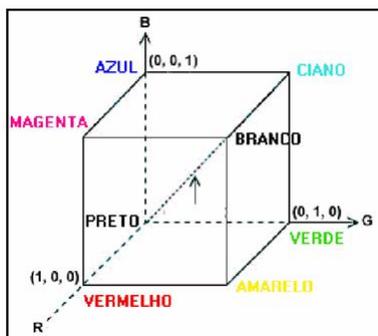


Figura 2.3: Sistema RGB representado em um cubo RGB, onde os vértices mostram as combinações resultantes no espaço de cor normalístico. Figura obtida em <http://www6.ufrgs.br/engcart/PDASR/imagem7.JPG>.

2.4.2 Modelo de Cor HSV (*Hue, Saturation, Value*)

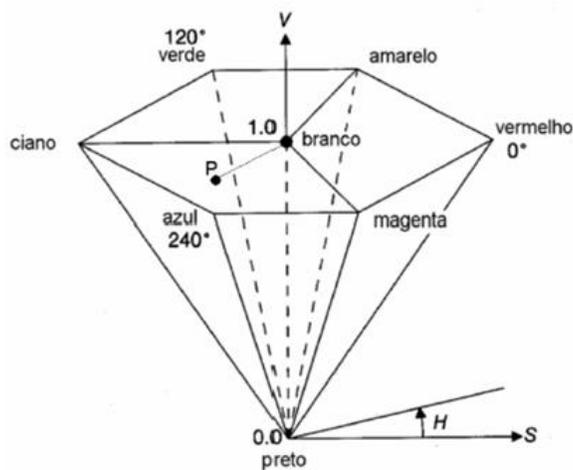


Figura 2.4: Modelo de Representação do sistema HSV para algumas cores. Figura obtida em <http://www6.ufrgs.br/engcart/PDASR/imagem8.JPG>.

Esse sistema de cores é o mais próximo à percepção da visão humana, por esse motivo, apresenta maiores vantagens em relação aos demais sistemas, em aplicações de visão computacional. Este sistema é composto pela Tonalidade, Saturação e Valor, onde a tonalidade é a cor definida pela faixa de valores de 0 a 360 ° (Tabela 2.3), representada por um círculo, a saturação varia pelo percentual da presença de tons de cinza de 0 a 100% e a valor é definido pelo brilho composto variando de 0 a 100%. Na Figura 2.4 apresenta o modelo HSV, a vantagem deste modelo de representação é a capacidade de representar as cores do espaço HSV em um único objeto. Devido à natureza bidimensional interfaces de computador, o modelo de cônico HSV é mais adequada para selecionar as cores para computação gráfica.

Tabela 2.3: Seleção de nível de cores. Tabela obtida em <http://pt.tech-faq.com/hsv.shtml>.

Angulo	Cor
0-60	Vermelho
60-120	Amarelo
120-180	Verde
180-240	Ciano
240-300	Azul
300-360	Magenta

2.4.3 A Cor como Característica Visual

Segundo a teoria do tri-estímulo da cor (Trucco and Verri, 1998; Gonzalez and Woods, 2001; Forsyth and Ponce, 2002), a visão colorida é resultante da excitação de três tipos de cones fotoreceptores da retina. O espaço de cor RGB baseia-se nessa teoria, relacionando a cor a um primeiro estágio na retina do processo de visão do ser humano. A codificação da percepção da

cor no espaço RGB não é eficiente. A descrição de uma cor é mais intuitiva quando se usam as características da matiz e da saturação do espaço HSV.

O espaço de cores conhecido como HSV (*Hue, Saturation, Value*) é um modelo de cor definido no espaço RGB de representação de cores. As cores são definidas pelos atributos de valor ou brilho, matiz e saturação. Esses atributos podem ser analisados e manipulados individualmente, ao contrário do sistema RGB, onde eles são interligados (Gonzalez and Woods, 2001).

O espaço HSV é representado graficamente como um cone e usa coordenadas cilíndricas polares para representar as cores. O vértice do cone do HSV representa o preto, enquanto o seu eixo coincide com o eixo acromático. O valor aumenta em sentido contrário ao do vértice. Uma seção circular do cone mostra a variação de matizes ao redor do perímetro. A saturação aumenta para fora do centro, passando de cinza para tons pastéis e destes para matizes espectrais puras (Gonzalez and Woods, 2001; Forsyth and Ponce, 2002).

A matiz informa a cor predominante em um determinado pixel da imagem. Pode ser compreendido como a medida do comprimento de onda médio da luz que um pixel reflete ou emite, definindo a cor do objeto. No modelo HSV, os valores da matiz estão representados por uma seqüência radial ao redor dos círculos de saturação e do eixo de valor, um ângulo entre 0° e 360° (Gonzalez and Woods, 2001; Forsyth and Ponce, 2002).

A saturação tem como função estimar a pureza da cor. Quanto mais saturada, isto é, quanto maior o valor de saturação mais pura será a cor. A pureza também está relacionada com a quantidade de cor branca presente na cor. O valor da saturação expressa o intervalo de comprimentos de onda ao redor do comprimento de onda médio no qual a energia é refletida ou transmitida (Gonzalez and Woods, 2001; Forsyth and Ponce, 2002). No modelo HSV a saturação é representada como distância radial do ponto até o eixo central do cone.

O valor fornece a noção geral de quão iluminada está a imagem naquele momento. É a medida da energia total envolvida em todos os comprimentos

de onda responsáveis pela sensação de brilho dessa energia incidente sobre o olho (Gonzalez and Woods, 2001; Forsyth and Ponce, 2002). Seu valor é dado no modelo HSV pela distância de um ponto até a origem ou ápice do cone.

O *histograma de cores* descreve a distribuição das cores no objeto, independente de sua forma e orientação, definindo, assim, um modelo de cor do objeto (Stricker and Swain, 1994; Gonzalez and Woods, 2001). O histograma visto na Figura 2.5 $\mathbf{h}(\mathbf{I}) \in x$, onde x é um vetor onde cada elemento representa o número de pixels da imagem \mathbf{I} que é descrito por uma determinada cor.

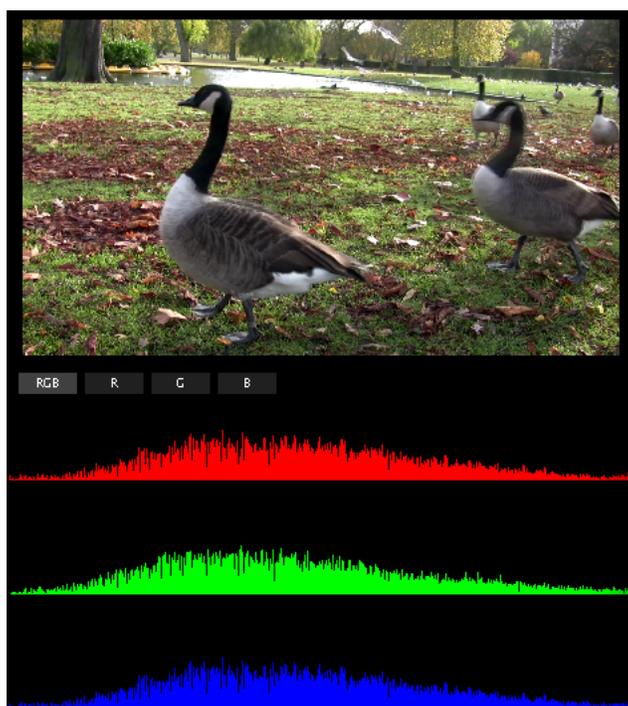


Figura 2.5: Histograma representando a distribuição de cores do canal RGB dividido em 3 gráficos para os canais *Red*, *Green* e *Blue*.

Para se quantificar a semelhança entre histogramas de cores, faz-se necessária a definição de uma métrica no espaço vetorial formado pela distri-

buição de cores representadas nos histogramas. Assim, para dois histogramas de cores \mathbf{h}_1 e \mathbf{h}_2 dados, as principais normas para medir semelhança entre elas são propostas por Stricker and Swain (1994); Gonzalez and Woods (2001).

$$d_{L1}(h_1, h_2) = \|h_1 - h_2\|_{L1} \quad (2.3)$$

$$= \sum_{i=1}^n |h_{1,i} - h_{2,i}|, \quad (2.4)$$

$$d_{L2}(h_1, h_2) = \|h_1 - h_2\|_{L2} \quad (2.5)$$

$$= \left(\sum_{i=1}^n |h_{1,i} - h_{2,i}|^2 \right)^{\frac{1}{2}}, \quad (2.6)$$

$$d_{L3}(h_1, h_2) = \frac{\sum_{i=1}^n h_{1,i} - h_{2,i}}{\sum_{i=1}^n h_{2,i}}. \quad (2.7)$$

Em alguns casos as medidas de similaridade do tipo d_{L1} por histogramas funciona bem quando os histogramas são esparsos. As medidas do tipo d_{L2} provocam o aparecimento de muitos falsos negativos, ou seja, não recupera todas as imagens cujos histogramas são perceptualmente similares. O uso da métrica d_{L3} , ocasiona o aparecimento de muitos falsos positivos (Stricker and Swain, 1994).

O uso da cor definida no espaço HSV e as métricas acima são combinadas de modo a fornecerem diferentes formas de identificação da cor dos objetos. Na abordagem, desenvolvida neste trabalho, um objeto tem suas características visuais representadas no espaço HSV e descritas por meio de um histograma de cor.

2.5 Pontos de Interesse

A abordagem descrita por Bay et al. (2006) usa cálculos de características de níveis de cinza local de pontos de interesse. Pontos de interesse são características locais com grandes quantidades de informações. As características locais usadas neste trabalho são baseadas sobre invariantes diferenciais de

níveis de cinza. Essas asseguram invariâncias sob um grupo de deslocamento dentro da imagem.

Para a determinação desses pontos de interesse nas imagens foi escolhido o algoritmo proposto por Bay et al. (2006). O SURF, além de ser robusto e com baixo custo computacional costuma ser invariante às diversas características presentes em imagens, como iluminação, rotação, escala, translação, dentre outras. Tornando assim, adequado a tarefa de identificação de um objeto em câmeras com diferentes campos de visões e ruídos de variações significativas entre cada quadro da câmera. Nas seções seguintes serão apresentados as etapas do algoritmo com mais detalhes. Na próxima seção será explicado um dos trabalhos que é fundamental na ganho de desempenho do algoritmo, o cálculo da integral da imagem.

2.5.1 Integral de Imagem

O conceito de integral de imagem foi desenvolvido por Viola and Jones (2001), tal tipo de imagem permite uma implementação extremamente de máscaras de convolução. Ou seja, tal representação, se utiliza do fato que características retangulares em imagens podem ser computadas de forma mais rápida utilizando a seguinte representação, dada uma imagem I e um ponto (x, y) a imagem integral I_{Σ} é calculada pela soma dos valores entre o ponto e o ponto de origem. Formalmente esta formula pode ser definida:

$$I_{\Sigma}(x, y) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(x, y) \quad (2.8)$$

2.5.2 Detecção dos Pontos de Interesse

Para detecção dos pontos de interesse de um objeto na cena é utilizado o método *Fast-Hessian* desenvolvido por Bay et al. (2006). O processo de implementação desse método pode ser visto como uma simplificação do cálculo da matriz Hessiana, com a melhora de desempenho proporcionada pela in-

tegral da imagem. Primeiramente será explicado como funciona a matriz Hessiana e como ela é aplicada em uma imagem I .

2.5.2.1 Matriz Hessiana

O detector SURF é baseado no determinante da matriz Hessiana. Foi considerado para esta matriz uma função contínua de 2 variáveis da qual o valor da função no ponto (x, y) é dado por $f(x, y)$. A matriz Hessiana, H , é a matriz das derivadas parciais da função f . Como apresentado na equação 2.9.

$$H(f(x, y)) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_2^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y_2^2} \end{bmatrix}. \quad (2.9)$$

O determinante desta matriz, conhecido como discriminante, é apresentado abaixo:

$$\det(H) = \frac{\partial^2 f}{\partial x_2^2} \frac{\partial^2 f}{\partial y_2^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2. \quad (2.10)$$

O valor da equação 2.10 é usado para obter o máximo e mínimo da função pelo teste derivado de segunda ordem. Visto que o determinante é o cálculo do produto dos autovalores da matriz Hessiana. Depois, são classificados os pontos através do sinal do resultado. Se o determinante é negativo então os autovalores tem sinais diferentes, portanto o ponto não é um extremo local. Caso, dê positivo então ambos os sinais são positivos ou negativos e neste caso o ponto é classificado como um extremo local.

Capítulo 3

Trabalhos Relacionados

Neste capítulo serão apresentados os trabalhos relacionados com o problema aqui proposto destacando as principais diferenças entre esses trabalhos e o aqui desenvolvido.

Pode-se dividir os trabalhos de reidentificação em duas categorias: a primeira categoria consiste em métodos que usam padrões biométricos como características para reidentificação de um objeto de interesse; a segunda categoria consiste em métodos que usam características visuais extraídas da aparência do objeto para reidentificação. Essas características não são únicas e podem ser formadas pela cor, forma, pontos de interesse, altura, entre outras informações visuais ou estruturais. Várias abordagens têm sido propostas. Recentemente alguns trabalhos têm proposto o uso de casamento de pontos de interesse para estabelecer a reidentificação de objetos (Sivic and Zisserman, 2003), de carros (Arth et al., 2007), e de pessoas (Gheissari et al., 2006; Hamdoun et al., 2008; Pham et al., 2007).

Para melhor compreensão foram sistematizadas as abordagens aqui apresentadas, os trabalhos relacionados estão subdivididos em Reidentificação por padrões Biométricos e Reidentificação pela aparência do objeto.

3.1 Reidentificação por Padrões Biométricos

O sistema proposto por Wei et al. (2004) apresenta a localização de pessoas aplicada em ambientes internos com múltiplas câmeras, conforme apresentado na Figura 3.1. O sistema pode utilizar um reconhecimento de face (biométrico) ou apenas características visuais extraídas da pessoa como altura, cor do cabelo e cor da roupa para identificação e localização de cada pessoa rastreada. O sistema opera um casamento temporal dos resultados obtidos por múltiplas câmeras e utiliza o conhecimento do domínio local (somente ambientes internos) para melhoria da sua eficiência. Entretanto, esse sistema utiliza algoritmos complexos para detecção e reconhecimento das características biométricas, tais como a face, que acarreta um alto custo computacional, tornando inviável a sua utilização em tempo real. Atualmente existem sistemas mais eficientes que utilizam características obtidas da aparência do objeto. Estes sistemas possuem resultados iguais ou melhores que o sistema desenvolvido por Wei et al. (2004) e trabalham com um menor custo computacional.

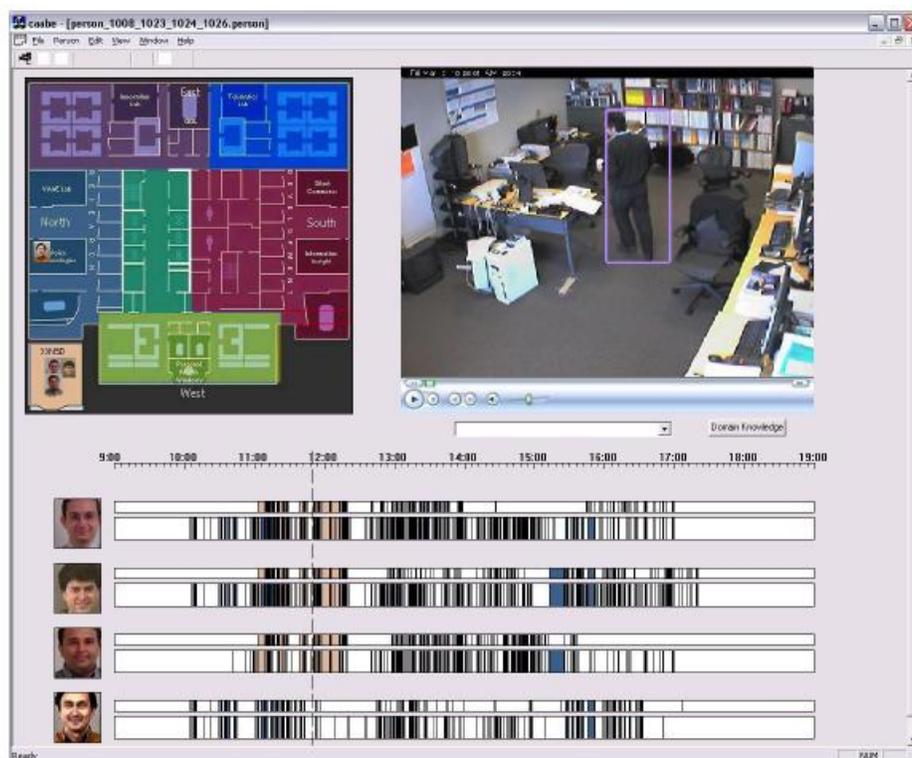


Figura 3.1: No canto esquerdo superior, representa a localização do objeto selecionado. No canto superior direito, apresenta a pessoa que foi detectada. E abaixo representa as pessoas detectadas e os seus respectivos tempos. Figura obtida em Wei et al. (2004).

3.2 Reidentificação pela Aparência do Objeto

Recentemente, os trabalhos de reidentificação pela aparência utilizam a cor como característica, que pode ser obtida através do rastreamento de pessoas ou da detecção dos pontos de interesse. Essas características podem ser aplicadas em qualquer objeto como pessoas e carros como apresentado nos próximos trabalhos relacionados. Elas são trabalhadas de forma mais eficiente e possuem os mesmos resultados que as técnicas biométricas, entretanto as

técnicas biométricas são mais precisas, pois através dos algoritmos complexos eles conseguem reconhecer uma pessoa com grandes deformações, como exemplo, no reconhecimento de faces utilizado no trabalho de Wei et al. (2004), o qual permite reconhecer a mesma pessoa, utilizando óculos, ou qualquer objeto que dificulte o seu reconhecimento. Enquanto que nas características da aparência utilizando cor existe a dificuldade de reidentificar pessoas diferentes que estejam trajando roupas com cores similares. Já os *wavelets* não trabalham com cor, pois trabalham principalmente com imagens em tons de cinza utilizando informações estruturais e intensidade, entre outras características, dependendo de cada abordagem. Primeiramente, serão apresentados os trabalhos relacionados que utilizam a cor como característica.

3.2.1 Característica Visual: Cor

Morioka et al. (2006) propõe um histograma de cor baseado na configuração do modelo do objeto conforme os resultados apresentados na Figura 3.2. Esse modelo denominado modelo global é eficiente para rastreamento de objetos, principalmente pessoas, e casamento de padrões. O alto espaço dos histogramas de cores gerados pela cor da aparência do objeto extraído em diferentes posições e orientações é usado para sua configuração. Entretanto, é difícil configurar um modelo apropriado baseado na antecipação dos movimentos do objeto, assim como, o casamento de padrões de características de cor não é consistente quando a disparidade é alta na localização e orientação do mesmo. Quando a pessoa se movimenta dentro e fora da região observada pelas câmeras, ou movimenta-se entre as regiões sobrepostas das câmeras, isso pode gerar alvos desconectados. Os resultados dos experimentos apresentados na Figura 3.2 utilizaram testes simples, onde apresenta somente uma pessoa andando em uma única sala com várias câmeras. Não houve experimentos com uma quantidade maior de pessoas e nem com pessoas com cores similares e de diferentes deformações, tanto na altura quanto na largura. Esse sistema trabalha sobre algoritmos complexos para criar o modelo de cada pessoa rastreada possuindo um alto custo computacional e limitando-se aos ambientes

internos.



Figura 3.2: Apresenta o rastreamento do objeto utilizando 2 câmeras com sobreposição de campo. Figura obtida em Morioka et al. (2006).

A abordagem ViSE desenvolvida por Park et al. (2006) propõe um componente semi-automático de um sistema de vigilância com câmeras interconectadas em uma rede conforme apresentado na Figura 3.3, que representa uma máquina de busca visual. O ViSE visa ajudar na operação de monitoramento de enormes quantidades de vídeos capturados, rastreando e procurando por pessoas em cada vídeo com base em suas características primitivas, como a cor das roupas e altura, com a interação de um operador humano. Nessa abordagem é utilizada a detecção, rastreamento e o reconhecimento de pessoas. O reconhecimento é o casamento entre o modelo das características primitivas de cada pessoa com as informações dadas pelo operador. O componente trabalha com módulos complexos como remoção de fundo e sombra, rastreamento e a criação do modelo utilizando a cor da aparência do objeto. Todos eles possuem um alto custo computacional e baixa eficiência, não podendo portanto, trabalhar em tempo real. O componente como visto em Park et al. (2006) possui a limitação de trabalhar em ambientes internos e somente com uma pessoa. Este trabalho possui uma idéia interessante, pois consiste em reidentificar as pessoas pela cor da roupa. Recentemente, não existe um algoritmo de subtração de fundo que apresente ótimos resultados quanto a eficiência e extração sem ruído dos objetos em movimento. Os seus experimentos apenas utilizaram uma base de dez pessoas com três câmeras

distintas, necessitando de mais experimentos que comprovem a sua aplicação.

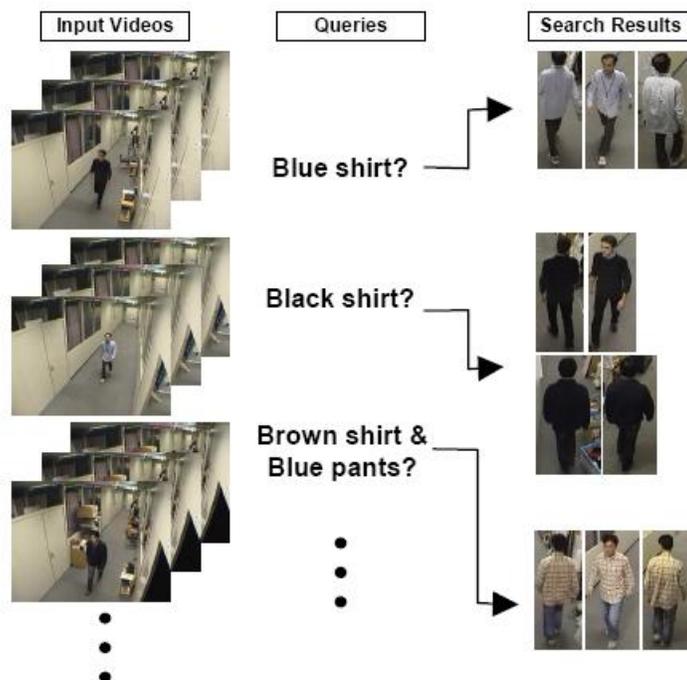


Figura 3.3: Um diagrama esquematizado do funcionamento do ViSE. A busca é feita por características primitivas que são dadas como entrada por um operador humano e apresenta o conjunto de pessoas visualizadas pelas câmeras em rede que possuem aquelas características. Figura obtida em Park et al. (2006).

Entretanto, caso um operador deseje reidentificar uma pessoa, ele não conseguirá de maneira precisa com o trabalho desenvolvido em Park et al. (2006), pois a pessoa possui uma natureza dinâmica. E esse problema é tratado no trabalho desenvolvido em Gheissari et al. (2006), onde ele apresenta duas abordagens para reidentificação de pessoas usando operadores de interesse e casamento de modelos, respectivamente, para estabelecer as correspondências espaciais entre as mesmas. Nesse trabalho foi desenvolvido um novo segmentador espaço temporal que utiliza o espaço e o tempo extraídos

das assinaturas invariantes para a cor da roupa. O tempo de extração dos valores de vários operadores de pontos de interesse é limitado, pois a aparência das pessoas é sempre dinâmica. Por isso, esse trabalho escolhe um operador que gera um grande número de valores na região com alto conteúdo de informação, aumentando a probabilidade de estabelecer correspondentes verdadeiros entre imagens da mesma pessoa. Utiliza o operador invariante de *Hessian*, o mesmo operador utilizado neste trabalho. Um casamento de assinaturas é utilizado para estabelecer a correspondência entre dois conjuntos de pontos de interesse. Uma pontuação de casamento é computada baseada na cardinalidade dos conjuntos finais de correspondências, focando no modelo de forma articulado de cada pessoa. Esse modelo é composto por partes diferentes do corpo como braços, pernas, cabeça e tronco. Para casamento dos modelos gerados é usado um algoritmo de programação dinâmica, onde as assinaturas invariantes são formadas pela combinação de cor e informação de estrutura extraída de diferentes regiões, de tal forma que a informação de cor é formada pela matiz e saturação. Ele aplicou de forma interessante o uso das características de cor buscando resolver o problema de mudança dinâmica da cor do objeto, entretanto não solucionou o problema de reidentificar pessoas diferentes com cores similares, conforme pode ser visto em seus resultados apresentados na Figura 3.4.



Figura 3.4: A imagem dada como consulta é mostrada na coluna da esquerda, e as outras colunas representam o resultado dos principais conjuntos de imagens trazidos de cada consulta. Para cada consulta as imagens trazidas corretamente estão destacadas por uma caixa. A terceira linha mostra um exemplo onde a correta correspondência não está presente. Figura obtida em Gheissari et al. (2006).

O sistema desenvolvido por Pham et al. (2007) demonstra a sua robustez por meio do pré-processamento dos vídeos de vigilância e uma interface conforme apresentada na Figura 3.5 que possui uma busca interativa. O sistema é formado por diferentes componentes, como detecção do movimento, classificação do objeto, modelagem e casamento de padrões extraídos do objeto. Entretanto, no módulo de casamento de padrões do objeto, os experimentos dos autores mostraram uma limitação do sistema na qualidade do casamento que consiste na dependência pela máscara do objeto. Sem a máscara do objeto a precisão é muito fraca na distinção entre o objeto e o fundo da cena. No módulo de classificação do objeto utiliza a informação do SIFT, uma abordagem que apresenta dificuldades na detecção dos pontos de interesse com variação de luz, segundo testes desenvolvidos por Mikolajczyk and Schmid (2005). O conjunto de treinamento utilizado não é totalmente

representativo para os dados usados como testes conforme apresentado na Figura 3.5 onde observar-se o problema de identificar falsos positivos com cores similares.



Figura 3.5: Interface do sistema robusto desenvolvido por Pham et al. (2007). A parte A representa os 3 humanos detectados. A parte B a localização no mapa. A parte C e D a busca pelas aparências similares dos humanos selecionados em A. E a parte E representa o resultado obtido. Figura obtida em Pham et al. (2007).

Portanto, a cor não é uma característica precisa para identificar uma pessoa. A abordagem desenvolvida por Park et al. (2006) trabalha somente com a reidentificação de cores, entretanto atualmente um operador possui interesse em reidentificar uma pessoa em vez de reidentificar cores. Para isso, existem abordagens que utilizam os *wavelets* obtidos de pontos de interesse. Esses *wavelets* não utilizam a cor e possuem aspectos invariantes.

3.2.2 Característica Estrutural: Pontos de Interesse

Um dos primeiros sistemas para recuperação de objeto genérico é o sistema de busca de vídeo do Google desenvolvido por Sivic and Zisserman (2003) que obteve um resultado impressionante para um limitado número de vídeos. O sistema opera com pontos de interesse, empregando a abordagem de extração dos pontos e casamento de padrões desenvolvido por Lowe (2003). Em vez de impor restrições geométricas fechadas, os autores definem restrições que contam o número de correspondências em um casamento de padrões por vizinhança, resultando em um alto custo computacional. Os seus resultados são apresentados na Figura 3.6. Entretanto, a abordagem SIFT desenvolvida por Lowe (2003) apresentou dificuldades na detecção dos pontos de interesse com variação de luzes, segundo testes desenvolvidos por Mikolajczyk and Schmid (2005), portanto, o trabalho desenvolvido por Sivic and Zisserman (2003) terá dificuldades em obter os pontos em imagens com grande variação de luz.

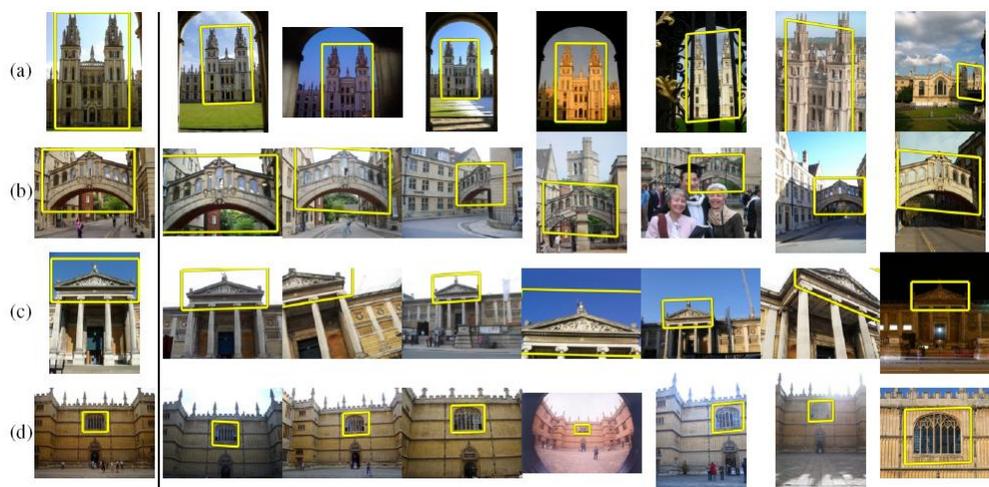


Figura 3.6: A primeira coluna representa a imagem dada como consulta. Depois, nessa imagem são extraídas as características e comparadas com o banco de imagens. As outras colunas representam as imagens do banco similares a imagem dada como consulta. Figura obtida em Sivic and Zisserman (2003).

Atualmente, a técnica *PCA-SIFT* diminui as dimensões do *SIFT* e gera uma assinatura compacta com a mesma qualidade de informação, porém apresenta um alto custo computacional para isso. Esta técnica foi utilizada no trabalho desenvolvido por Arth et al. (2007), onde apresenta um novo arcabouço para reaquisição de objetos em uma rede de câmeras inteligentes. As suas maiores contribuições se devem ao fato dos objetos usarem características locais, comunicação entre os nós adjacentes e o algoritmo proposto de casamento de padrões conforme apresentado na Figura 3.7. Primeiro, os objetos são identificados e descritos pelo *PCA-SIFT*. O *PCA-SIFT* gera assinaturas robustas locais através das características extraídas dos quadros. Essa assinatura pode mais adiante ser comunicada de forma eficiente para os nós vizinhos da câmera. No módulo de casamento de padrões são comparadas as assinaturas do objeto, permitindo uma reaquisição eficiente e rastreamento de objetos sobre a rede inteira de câmeras inteligentes. O arcabouço funciona

em câmeras descalibradas, já que a calibração de multi-câmeras é uma tarefa difícil. Entretanto, esta abordagem necessita que o domínio de treinamento do objeto utilize exemplos de imagens que garantam um bom desempenho. É necessário que se faça sempre uma remoção dos nós desnecessários para não existir sobrecarga, resultando em um alto custo de manutenção. O arcabouço possui dificuldades em distinguir dois carros de formas similares com pinturas diferentes ou defeitos em suas superfícies. O *SIFT* possui problemas na detecção dos pontos de interesse em mudanças de luz, segundo testes desenvolvidos por Mikolajczyk and Schmid (2005). Atualmente, as câmeras não possuem meio de armazenamento como apresentados nesse trabalho para armazenar o modelo, onde o armazenamento é feito em memória RAM, apresentando desta forma uma limitação.

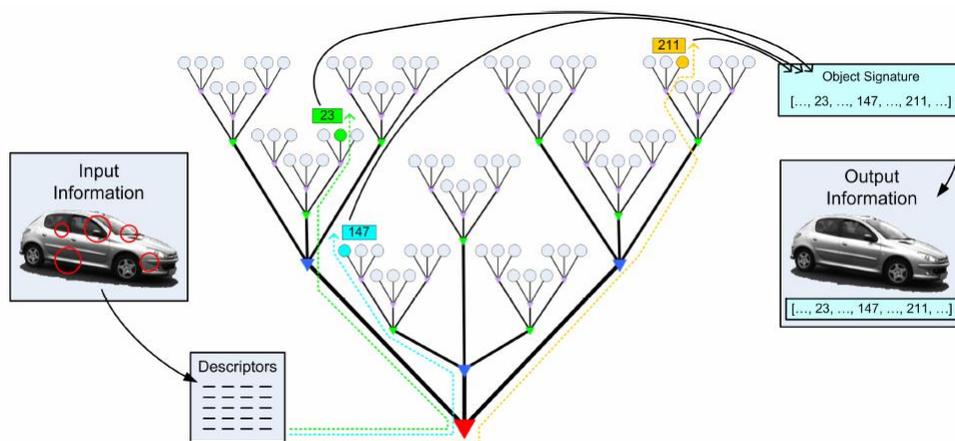


Figura 3.7: Dado uma imagem são extraídas suas características. É gerado um descritor e realizado o casamento de padrões com as características extraídas de uma rede de câmeras organizadas pela estrutura de árvore gerado pelo KNN, resultando nas reidentificações. Figura obtida em Arth et al. (2007).

Recentemente, o método SURF desenvolvido por Bay et al. (2006), utiliza tecnologias mais precisas e eficientes do que o SIFT e o PCA-SIFT

e não possui dificuldades em trabalhar sobre variação de luz. No trabalho desenvolvido por Hamdoun et al. (2008) apresenta uma abordagem que usa o casamento de assinaturas baseados no descritor de pontos de interesses obtidos do método SURF e coletados de uma pequena seqüência de quadros do vídeo. A novidade consiste no acúmulo de pontos de interesse durante o rastreamento de pessoas sobre cada câmera, para capturar variabilidades na aparência. A abordagem é formada pela construção dos modelos de cada pessoa, construção da consulta, comparação dos descritores, o casamento e a identificação. A construção do modelo consiste na geração da assinatura de cada pessoa detectada ou rastreada e para não existir redundância na informação, ela é extraída a cada meio segundo. A assinatura da pessoa é obtida pelo acúmulo de descritores dos pontos de interesse obtidos nestas imagens. A construção da consulta é feita da mesma forma que a construção do modelo porém utiliza uma quantidade menor de imagens. A comparação do descritor é calculado pela Soma das Diferenças Absolutas, onde é obtido a similaridade dos descritores dos pontos de interesse entre os modelos e a consulta. Conforme visto na Figura 3.8, o casamento entre os descritores é feito por uma busca através do *BBF* em uma *Árvore KD* que contém todos os modelos das pessoas. Finalmente, a identificação consiste em uma abordagem de votação onde cada ponto de interesse extraído da consulta é comparado com todos os pontos dos modelos armazenados na *Árvore KD*, e um voto é adicionado para cada modelo contendo um descritor casado. O modelo mais votado representa a reidentificação da consulta. Essa abordagem necessita do conhecimento de vários ângulos de visões da mesma pessoa.

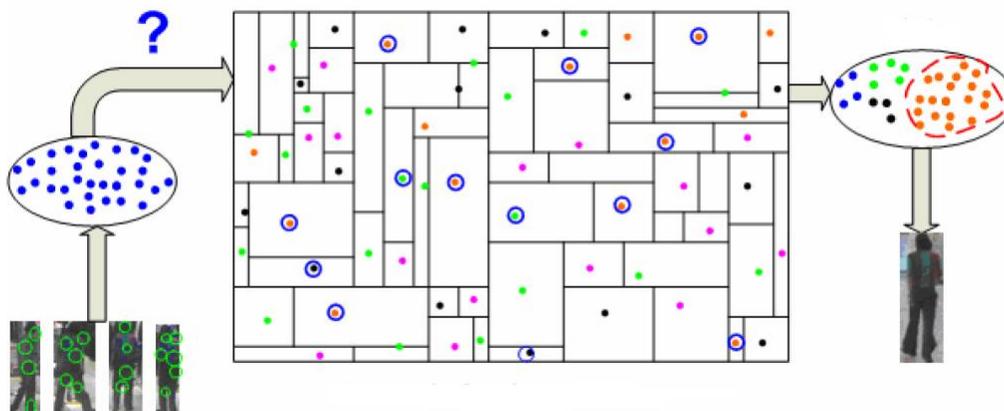


Figura 3.8: Visão esquemática da reidentificação de uma consulta, onde no meio representa a Árvore KD contendo os pontos de interesse das pessoas. Figura obtida em Hamdoun et al. (2008).

Por outro lado, a abordagem apresentada neste trabalho não utiliza métodos de rastreamento como em Wei et al. (2004); Hamdoun et al. (2008); Morioka et al. (2006); Park et al. (2006), detecção de pessoas como usado em Pham et al. (2007) entre outros métodos como apresentado nos trabalhos relacionados para obter a região onde se encontra o objeto em movimento. Este trabalho utiliza somente o detector SURF de pontos de interesse como utilizado em Sivic and Zisserman (2003); Arth et al. (2007). A novidade apresentada neste trabalho está no casamento dos pontos de interesse e suas respectivas cores, diferente do trabalho desenvolvido em Gheissari et al. (2006) que utiliza casamento entre imagens, ou Wei et al. (2004) que utiliza características biométricas e em Park et al. (2006); Morioka et al. (2006); Pham et al. (2007) que utilizam casamento de cores. O casamento utiliza apenas o sinal de *laplace* obtido pelo método SURF apresentado em Bay et al. (2006) e o cálculo de similaridade realizado pela Soma das Diferenças Quadráticas. Desta forma, possui um menor custo computacional e maior eficiência comparada as abordagens que utilizam modelos ou estruturas para casamento como em Hamdoun et al. (2008); Park et al. (2006); Arth et al. (2007). Es-

As abordagens possuem uma falha em relação a quantidade, pois a medida que existe o movimento de pessoas ou carros é inserido na estrutura como se fosse um novo objeto, isso também ocorre em Hamdoun et al. (2008) que não trata o dinamismo das pessoas como é tratado em Gheissari et al. (2006). A estrutura usada em Arth et al. (2007) para carros possui um método de manutenção para remover nós desnecessários, entretanto acaba tornando-se um gargalo no sistema. Portanto, a abordagem apresentada neste trabalho é mais eficiente e possui menor custo computacional em relação aos trabalhos relacionados, podendo ser aplicado tanto em ambiente externo quanto interno, com campos de visões das câmeras sobrepostos e não-sobrepostos e principalmente com câmeras descalibradas e calibradas. Em ambientes com câmeras calibradas informações mais precisas de localização da pessoa podem ser obtidas mais facilmente. Câmeras sem sobreposição dificilmente obtêm informações de todos os ângulos de visão da pessoa em movimento, mostrando que a abordagem utilizada nesse trabalho não necessita da visão de todos os ângulos da pessoa para realizar o casamento de forma precisa e rápida apenas de um ângulo. O detector do SURF não detecta sombras e não necessita de um removedor de sombras como utilizado em Park et al. (2006).

Capítulo 4

Descrição Funcional da Reidentificação

Neste capítulo será descrito a metodologia do trabalho. Para melhor compreensão dessa metodologia é necessário que se considere que um determinado ambiente é monitorado por um conjunto de n câmeras $C_1, C_2, C_3, \dots, C_n$ com capacidade local de processamento e interconectadas por um sistema de comunicação em malha. Considera-se ainda que o sistema monitora um ambiente interno, com controle de luminosidade, e que os campos de visão das câmeras não necessariamente se sobrepõem.

A metodologia do trabalho é dividida em dois principais componentes funcionais: controle e reidentificação. A Figura 4.1 apresenta uma visão esquemática da metodologia, onde o quadro pontilhado representa as câmeras e as elipses representam os componentes funcionais do sistema.

O componente funcional de controle possui o conjunto de procedimentos necessários para viabilizar a comunicação automatizada entre o operador e as câmeras a partir da consulta. A consulta é processada a partir de dados fornecidos por uma imagem do alvo dado como entrada pelo operador. A consulta obtida pelo componente de controle é repassada para o componente de reidentificação.

O componente funcional de reidentificação caracteriza-se como um pro-

cesso de reconhecimento de padrões, onde consiste em detectar, extrair as características visuais e reidentificá-las em cada câmera do sistema. Finalmente, esse componente funcional irá apresentar as reidentificações ao centro de vigilância. Esses dois componentes funcionais estão descritos nas seções seguintes.

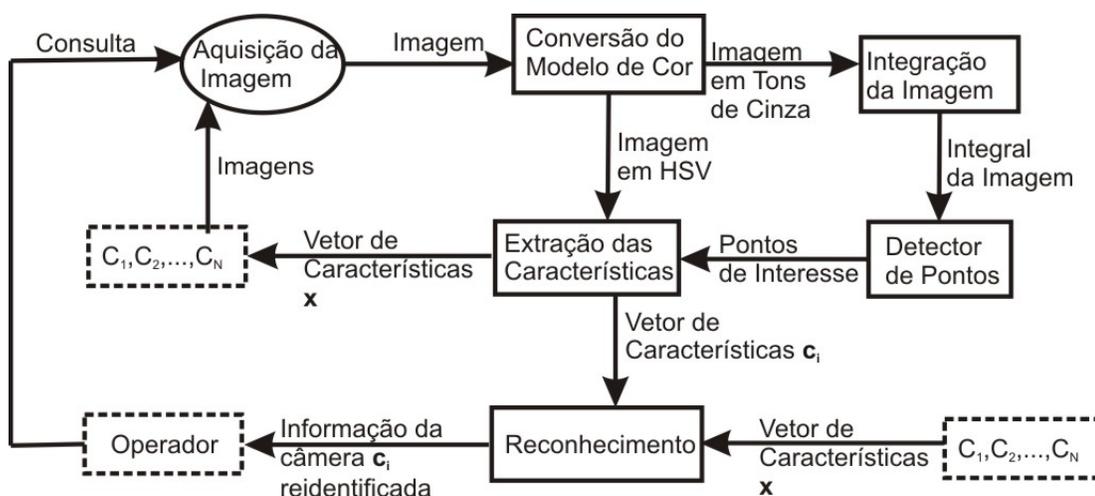


Figura 4.1: Esquema geral da metodologia proposta. O quadro pontilhado representa os agentes externos do sistema (operador e sistemas de câmeras), a elipse o procedimento utilizado no componente funcional de controle e os retângulos os procedimentos utilizados no componente funcional de reidentificação.

4.1 O Componente Funcional de Controle

O controle do sistema provê os procedimentos necessários para a comunicação das informações visuais entre o operador e as câmeras do sistema. Fundamentalmente, esse componente trabalha com a aquisição de imagens. Esse processo recebe como entrada a consulta do operador e um conjunto de imagens referentes a amostragem de cada câmera do sistema. Esse processo distribui as imagens adquiridas pela entrada para o componente funcional de

reidentificação. O controle recebe as reidentificações da consulta dada como entrada pelo operador através do componente de reidentificação e apresenta-as ao centro de vigilância conforme apresentado na Figura 4.1.

4.1.1 Aquisição da Imagem

Para melhor compreensão do processo de aquisição ele foi dividido em 2 modos de entrada: *offline* e *online*. Onde o *offline* possui dois diferentes tipos de entrada do sistema e o *online* é uma entrada realizada por meio de cada câmera. Abaixo estão descritos da seguinte maneira:

1. *Offline*:
 - (a) O operador fornece dados por meio de uma imagem do alvo de interesse que ele deseja reidentificar.
 - (b) O operador fornece dados por meio da seleção da câmera onde o alvo de interesse que ele deseja reidentificar se encontra.
2. *Online*:
 - (a) A amostragem da sequência de imagens obtidas do conjunto de câmeras $C_1, C_2, C_3, \dots, C_n$ onde $n \geq 0$ é obtida pelos quadros da mesma. A quantidade dessa amostragem é obtida pelo fps (*frames per second*) de cada câmera, por padrão esse fps é 25.

A partir do processo de aquisição, a imagem é enviada para a conversão entre os modelos de cor do componente funcional de reidentificação.

4.2 Componente Funcional de Reidentificação

O componente funcional de reidentificação é formado pelos principais processos para um reconhecimento de padrão. Esses processos são a conversão do modelo de cor, a integração da imagem, detector de pontos, extração das características e o reconhecimento.

Seja \mathbf{x}_j o vetor de pontos de interesse obtidos a partir das características da consulta. A reidentificação realiza o reconhecimento de padrões entre os m vetores de características obtidos de cada ponto de interesse do vetor \mathbf{x}_j , $j = 1, 2, 3, \dots, m$, obtido da consulta adquirida pelo operador e os z vetores de características obtidos de cada ponto de interesse do vetor \mathbf{c}_i , $i = 1, 2, 3, \dots, z$, de cada câmera do sistema. Esse componente possui o processo de conversão de modelos de cor, detecção de pontos, extração de características e reconhecimento, conforme visto na Figura 4.1. A seguir serão apresentados em detalhes os procedimentos internos do reconhecimento.

4.2.1 Conversão de Modelos de Cor

A primeira etapa do componente de reidentificação é a conversão do modelo de cor. Onde esse processo converte uma imagem descrita no modelo RGB em uma imagem em tons de cinza e transforma uma imagem colorida descrita em RGB para o HSV. Nas seções seguintes serão apresentados os detalhes da conversão.

4.2.1.1 Conversão de uma imagem descrita no modelo RGB em uma imagem em tons de cinza

Seja $I(x, y)$ a imagem descrita no modelo RGB e cada canal R, G e B está descrito em uma imagem de 8 bits, onde cada canal é representado entre os valores 0 a 255.

A conversão entre uma imagem $I(x, y)$ para uma imagem em tons de cinza $GRAY(x, y)$ consiste em para cada pixel i da imagem $I(x, y)$ é representado como o pixel g na imagem $GRAY(x, y)$ como $g = 0.299 * R + 0.587 * G + 0.114 * B$.

Com isso a imagem $GRAY(x, y)$ é enviada para o processo de integração da imagem. onde é utilizada no processo de integral da imagem como especificado na seção 4.2.2.

4.2.1.2 Conversão de uma imagem colorida descrita no modelo RGB para HSV

Seja $I(x, y)$ a imagem descrita no modelo RGB e cada canal R, G e B está descrito em uma imagem de 8 bits, onde cada canal é representado entre os valores 0 a 255.

A conversão entre uma imagem $I(x, y)$ descrita em RGB para uma imagem descrita no modelo HSV $HSV(x, y)$ consiste em decompor a imagem $I(x, y)$ nos canais R, G e B e a imagem $HSV(x, y)$ nos canais H, S e V, onde $V = \max(R, G, B)$

$$S = (V - \min(R, G, B))/V \text{ se } V > 0, 0 \quad (4.1)$$

$$(G - B) * 60/S, \text{ if } V = R \quad (4.2)$$

$$H = 180 + (B - R) * 60/S, \text{ if } V = G \quad (4.3)$$

$$240 + (R - G) * 60/S, \text{ if } V = B \quad (4.4)$$

se $H < 0$ então $H = H + 360$.

Na saída $0 \leq V \leq 255$, $0 \leq S \leq 1$ e $0 \leq H \leq 360$.

Os valores são então convertidos para $V = V * 255$, $S = S * 255$ e $H = H / 2$. Com isso, a imagem $HSV(x, y)$ é enviada para o processo de extração das características.

4.2.2 Integração da Imagem

A integração recebe como entrada a imagem em tons de cinza. Essa integração é implementada eficientemente considerando apenas a soma dos três valores anteriores, à esquerda, acima e na diagonal superior esquerda. Uma vez tendo calculada a integral da imagem em tons de cinza usando a fórmula apresentada na Seção 2.5.1, somente é necessário quatro operações de forma

a determinar a soma de qualquer região quadrada na imagem, independente do tamanho da mesma. A Figura 4.2 ilustra o processo de uso da integral da imagem, a área $S = A + D - (B + C)$. Isto mostra que no algoritmo o tamanho do retângulo não influencia no desempenho. Muito utilizado portanto em grandes áreas.

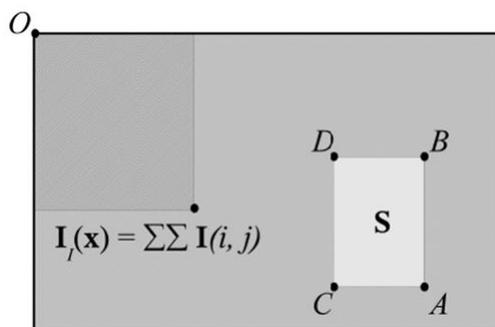


Figura 4.2: Cálculo da área usando integral de imagens. Figura obtida em Evans (2009).

Depois esta imagem é dada como entrada para a próxima etapa, esta etapa usa o *Fast-Hessian* para detectar os pontos de interesse.

4.2.3 Detecção dos Pontos de Interesse

Aplicando a teoria de Hessian sobre imagens em vez de uma função contínua $f(x, y)$ como apresentado na Seção 2.5.2, tem-se que inicialmente troca-se os valores da função $f(x, y)$ pelas intensidades do pixel $\mathbf{I}(x, y)$. Depois, é utilizado um método para calcular as derivadas parciais de segunda ordem da imagem. Estas derivadas podem ser calculadas por convolução com um núcleo apropriado.

O Gaussiano normalizado escalar de segunda ordem é o filtro escolhido uma vez que permite a análise sobre as escalas, bem como o espaço (a teoria de espaço escalar é apresentado na Seção 4.2.4).

Podemos construir núcleos pela derivada Gaussiana em \mathbf{x} , y e combi-

nado com a direção xy como o cálculo das quatro entradas da matriz Hessiana apresentada na fundamentação teórica. Com o uso da Gaussiana é permitido variar a quantidade de nitidez durante a etapa de convolução, com isso o determinante é calculado em diferentes escalas. Além disso, a Gaussiana é uma convolução com o núcleo invariante à rotação. Desta forma, agora calcula-se a matriz Hessiana, \mathbf{H} , em função de ambos espaço $\mathbf{x} = (x, y)$ e da escala σ .

$$\mathbf{H}(\mathbf{x}, \sigma) = \begin{bmatrix} L_{xx}(\mathbf{x}, \sigma) & L_{xy}(\mathbf{x}, \sigma) \\ L_{xy}(\mathbf{x}, \sigma) & L_{yy}(\mathbf{x}, \sigma) \end{bmatrix}. \quad (4.5)$$

Onde L_{xx} consiste na convolução da derivada gaussiana de segunda ordem $\frac{\partial^2 g(\sigma)}{\partial x^2}$ com a imagem no ponto $\mathbf{x}=(x, y)$ e similarmente para L_{yy} e L_{xy} . Estas derivadas são conhecidas como Laplacianas da Gaussiana.

Depois, calcula-se o determinante da matriz Hessiana, $\det \mathbf{H}$, para cada pixel na imagem e através do resultado é obtido apenas os pontos de interesse. Esta variação do detector Hessiano é similar ao proposto por Beaudet (1978). Lowe (2003) encontrou um aumento de desempenho em aproximar o Laplaciano de Gaussianos por uma diferença de Gaussianos. De uma maneira similar, em Bay et al. (2006) é proposto uma aproximação para o Laplaciano de Gaussianos pelo uso de representações do filtro de caixa dos núcleos respectivos, essa aproximação é utilizada neste trabalho.

Na Figura 4.3 é ilustrada uma similaridade entre os núcleos discretizados e descartados e em contrapartida os filtros de caixas dos mesmos. Considerável aumento de desempenho é encontrado quando esses filtros são usados em conjunção com a integral da imagem descrita na Seção 2.5.1. Para quantificar a diferença considera-se o número de acessos e operações no vetor requeridas na etapa de convolução. Um filtro de 9×9 requer 81 acessos e operações no vetor, enquanto que para a representação do filtro de caixa requer somente oito acessos e operações. Como o filtro é aumentado em tamanho, o custo computacional aumenta de forma significativa para o Laplaciano original enquanto que o custo para os filtros de caixa é independente do tamanho.

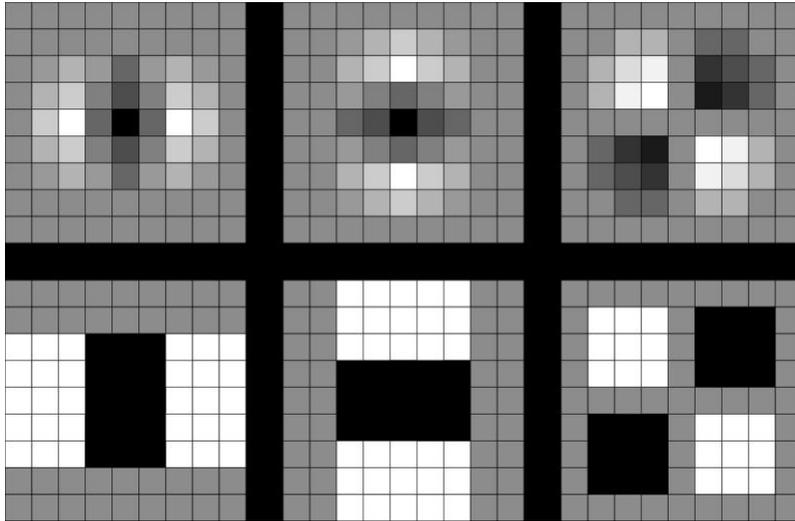


Figura 4.3: Aproximação Laplaciana de Gaussiano. Linha de cima: Os discretizados e descartados derivadas Gaussianas de segunda ordem no \mathbf{x} , y e direções xy . Referentes ao L_{xx} , L_{yy} e L_{xy} . Linha de baixo: Aproximações do filtro de caixa com peso no \mathbf{x} , y e direções xy . Referem-se estes como D_{xx} , D_{yy} , D_{xy} . Figura obtida em Evans (2009).

Na Figura 4.3 apresenta de forma simples os pesos aplicados para cada seção de filtro. Para o filtro D_{xy} , as regiões pretas são ponderadas com o valor 1, as regiões brancas com o valor -1 e o resto das áreas não são ponderadas. Os filtros D_{xx} e D_{yy} são ponderadas de forma similar mas as regiões brancas tem o valor 1 e os pretos o valor 2. Desta forma, a simples ponderação permite um rápido cálculo das áreas. Em Bay et al. (2006) é apresentado a fórmula 4.6 com uma aproximação precisa para o determinante Hessiano usando os Gaussianos aproximados, onde

$$\det(H_{approx}) = D_{xx}D_{yy} - (0.9D_{xy})^2. \quad (4.6)$$

Esse determinante é reduzido em relação à resposta do *blob* na localização $\mathbf{x} = (x, y, \sigma)$. A busca pelo máximo local desta função sobre ambos espaço e escala produzem os pontos de interesse para uma imagem. Depois, é

enviado os pontos de interesse para o processo de extração das características que é explicado na próxima seção.

4.2.4 Construção do Espaço Escalar

Antes de detectar os pontos de interesse usando o determinante Hessiano, é necessário introduzir uma noção sobre espaço escalar. Um espaço escalar, na matemática, é uma função contínua da qual pode ser utilizada para encontrar o extremo cruzamento de todas as escalas possíveis (Witkin, 1983).

Na área de visão computacional o espaço escalar é tipicamente modelado como uma pirâmide onde a imagem de entrada é interativamente envolvida por um núcleo Gaussiano e repetidamente sub-amostrado (reduzido em tamanho).

Este mesmo método é utilizado para grandes efeitos na abordagem SIFT desenvolvida por Lowe (2003) onde cada camada baseia-se na camada anterior. Desta forma, as imagens precisam ser redimensionadas, essa tarefa não é eficiente computacionalmente.

Como o tempo de processamento dos núcleos usados no SURF é invariante no tamanho, o seu espaço escalar pode ser criado para aplicar núcleos de tamanho aumentado para a imagem original. Com isso, permite que para múltiplas camadas da pirâmide de espaço escalar sejam processados simultaneamente e exclui a necessidade de subamostrar a imagem, logo prover um aumento no desempenho. Na Figura 4.4 é ilustrada a diferença entre a estrutura do espaço escalar tradicional e em contra partida a estrutura do espaço escalar do SURF.

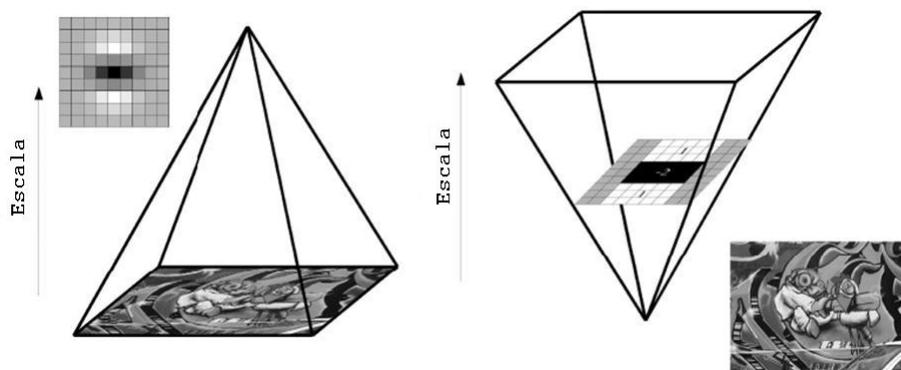


Figura 4.4: Amostragens em Diversas Escalas. Utilizada pelo SURF, à Direita, Diferente da abordagem tradicional à esquerda. Figura obtida em Evans (2009) e adaptada.

O espaço escalar é dividido dentro de um número de *oitavos*, onde cada oitavo representa uma série de respostas, mapas que cobrem o dobro da escala. Em Bay et al. (2006) o nível menor do espaço escalar é obtido da saída dos filtros 9×9 . Camadas posteriores são obtidas pelo aumento de escala do filtro, enquanto que mantêm a mesma proporção do esboço do filtro. A medida que o tamanho do filtro aumenta o valor da escala Gaussiana associado é aumentado, e em vez de existir proporções constantes do esboço, pode-se calcular esta escala pela fórmula:

$$\sigma_{approx} = \frac{1.2}{9} x \frac{\text{Escala do filtro de base}}{\text{Tamanho do Filtro de Base}} \quad (4.7)$$

Quando grandes filtros são construídos, há um número de fatores que devem ser considerados. O aumento no tamanho é limitado pelo tamanho dos pontos positivos e negativos da base das derivadas Gaussianas de segunda ordem. Na Figura 4.5 é ilustrada a estrutura dos filtros a medida que eles aumentam de tamanho.

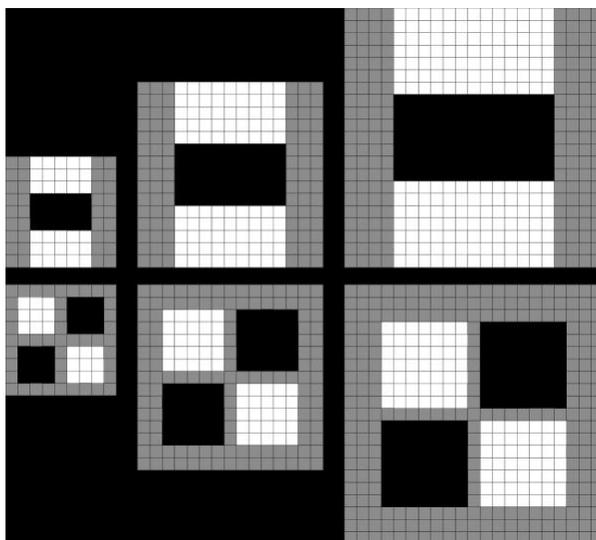


Figura 4.5: Os tamanhos dos Filtros posteriores devem diferenciar-se por um mínimo de seis para preservar a estrutura do filtro. Figura obtida em Evans (2009).

A tarefa de localizar a escala e rotação invariante dos pontos de interesse na imagem pode ser dividido em três etapas. Primeiro as respostas são limiarizadas do mesmo modo que todos os valores abaixo do limiar predeterminado são removidos. Aumentar o limiar diminui o número de pontos de interesse detectados, deixando apenas os mais fortes, enquanto que diminuir o limiar permite aumentar a quantidade de detectados. Assim, o limite pode ser adaptado para adequar a detecção para a aplicação.

Depois da limiarização, um algoritmo de supressão não-maximal é executado para encontrar um conjunto de pontos candidatos. Para fazer isso cada pixel no espaço escalar é comparado com seus 26 vizinhos, composto de oito pontos na escala nativa e nove em cada uma das escalas acima e abaixo. Na Figura 4.6 é ilustrada um passo de supressão não-maximal. Neste estágio, tem-se um conjunto de pontos de interesse com a força mínima determinada pelo valor do limiar e da qual são, além disso, mínimo e máximo local no espaço escalar.

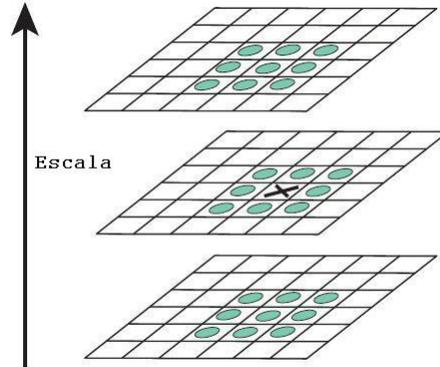


Figura 4.6: Supressão não-maximal. O pixel marcado "X" é selecionado como o máximo se for maior que os pixels que o rodeiam em seu intervalo e nos intervalos acima e abaixo. Figura obtida em Evans (2009) e adaptada.

O passo final em localizar os pontos envolvidos, consiste em interpolar os dados próximos para encontrar a localização em ambos escala e espaço para precisão do sub-pixel. Isto é feito para preencher um quadrático 3D como proposto por Brown and Lowe (2002). A fim de fazer isto é expressado o determinante da função Hessiana, $H(x, y, \sigma)$, como uma expansão por Taylor para centralizados termos quadráticos até a localização detectada. Isto é expressado como:

$$H(x) = H + \frac{\partial H^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 H}{\partial x^2} x. \quad (4.8)$$

A localização interpolada do extremo, $\hat{x} = (x, y, \sigma)$, é encontrado através das derivadas desta função e atribuído zero como abaixo:

$$\hat{x} = - \frac{\partial^2 H^{-1} \partial H}{\partial x^2 \partial x}. \quad (4.9)$$

As derivadas aqui são aproximadas por diferenças finitas da vizinhança dos pixels. Se \hat{x} é maior do que 0.5 no x, y ou direções σ , ajustando a

localização e executando a interpolação novamente. Este procedimento é repetido até \hat{x} seja menor que 0.5 em todas as direções ou o número de passos da interpolação predeterminada for excedido. Aqueles pontos da qual não convergem são descartados do conjunto de pontos de interesse deixando apenas os mais estáveis e repetidos.

4.2.5 Extração das Características

As características são formadas pelo descritor obtido a partir dos pontos de interesse determinados pelo método descrito na seção anterior e pela cor. A seguir são apresentados com detalhes os métodos de extração utilizados para formação do vetor de características.

4.2.5.1 Descritor do Ponto de Interesse

O descritor descreve como as intensidades do pixel são distribuídas dentro de uma escala que depende da vizinhança de cada ponto de interesse detectado pelo *Fast-Hessian*. Esta abordagem é a mesma usada no método SIFT desenvolvido por Lowe (2003), entretanto imagens integrais trabalham em conjunção com filtros conhecidos como *wavelets de Haar*, eles são usados para aumentar a robustez e diminuir o tempo de computação. *Wavelets de Haar* são filtros simples da qual podem ser usados para encontrar gradientes nas direções x e y .



Figura 4.7: *Wavelets de Haar* representados pelos dois filtros. Figura obtida em Evans (2009).

Como pode ser visto na Figura 4.7 o filtro à esquerda computa a resposta na direção x e o filtro à direita computa a resposta na direção y . Pesos

são 1 para regiões pretas e -1 para regiões brancas. Quando usado com imagens integrais cada *wavelet* requer somente 6 operações para computar.

A formação do descritor pode ser dividida em duas tarefas distintas: determinação da orientação e componentes do descritor. Primeiro, uma escala que depende da janela é construída e para cada ponto de interesse é determinado uma orientação reprodutível na qual é extraído um vetor de 64 ou 128 dimensões. É importante que todos os cálculos para o descritor estejam baseados nas medidas relativas à escala detectada afim de alcançar os resultados da escala invariante. O procedimento para determinação da orientação é explicado em seguida.

Determinação da Orientação

A fim de alcançar invariância para rotação da imagem cada ponto de interesse detectado é atribuído uma orientação reprodutível. A extração dos componentes do descritor é executada em relação à direção obtida pela orientação, por isso, é necessário que essa direção seja repetitiva em diferentes condições. Para determinar a orientação, as respostas do *wavelet de Haar* de tamanho 4σ são calculados para um conjunto de *pixels* dentro de um raio de 6σ do ponto detectado, onde σ refere-se a escala de cada ponto que já foi detectado. Um conjunto específico de *pixels* é determinado por amostragem daqueles *pixels* que se encontram dentro do círculo de tamanho σ .

As respostas são ponderadas com um Gaussiano centralizado no ponto de interesse. Já o resto dos pontos Gaussianos dependem da escala do ponto e do desvio padrão 2.5σ . Uma vez ponderadas as respostas são representadas como pontos no espaço vetorial, com as respostas do x ao longo da abscissa e as respostas do y ao longo da ordenada. A orientação dominante é selecionada por rotação de um segmento circular abrangendo um ângulo em radianos de $\frac{\pi}{3}$ em torno da origem. Em cada posição, as respostas do x e do y dentro do segmento são somadas e utilizadas para formar um novo vetor. O maior vetor proporciona sua orientação do ponto de interesse. Este processo é ilustrado na Figura 4.8.

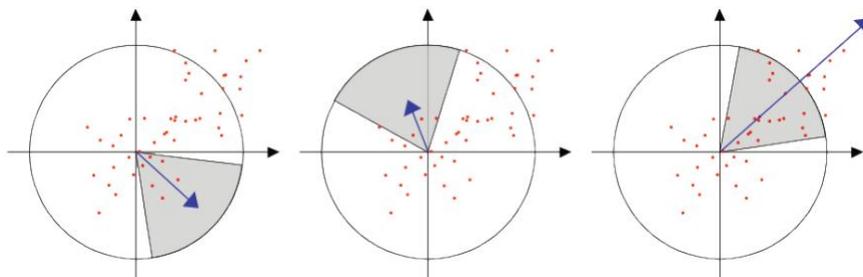


Figura 4.8: A determinação da orientação é obtida pela soma das produções dos vetores aqui mostrados em azul. O maior desses vetores determina a orientação dominante. Figura obtida em Evans (2009).

A seguir são apresentados os passos para obter os componentes do descritor.

Componentes do Descritor

O primeiro passo para extrair os descritores é construir uma janela quadrada em torno do ponto de interesse. Essa janela contém os *pixels* que formam as entradas no vetor descritor de tamanho 20σ , novamente σ se refere à escala detectada. Além disso, a janela é orientada ao longo da direção encontrada, tal que todos os cálculos posteriores são relativos a essa direção, como pode ser visto na Figura 4.9.

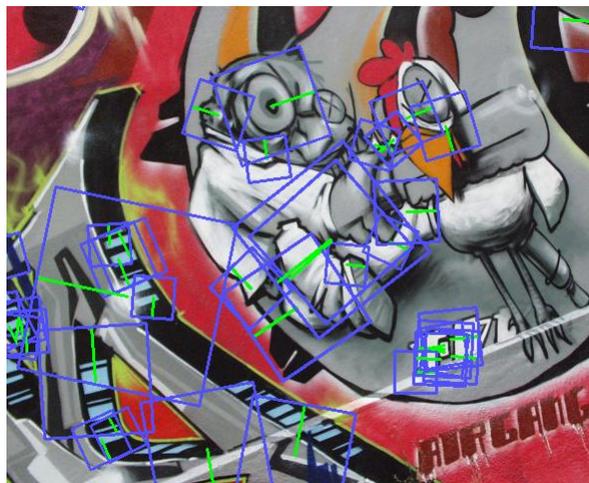


Figura 4.9: Janelas do descritor. O tamanho da janela é 20 vezes a escala do ponto detectado e é orientado ao longo da direção dominante apresentada em verde. Figura obtida em Evans (2009).

A janela do descritor é dividida dentro de subregiões regulares 4×4 . Dentro de cada uma dessas subregiões, os *wavelets de Haar* de tamanho 2σ são calculados para 25 pontos amostrados e distribuídos regularmente. As respostas do *wavelet* x e y são referenciados por d_x e d_y respectivamente, então para cada 25 pontos amostrados (cada subregião) é colecionado,

$$v_{subregion} = [\Sigma d_x, \Sigma d_y, \Sigma |d_x|, \Sigma |d_y|]. \quad (4.10)$$

Portanto, cada subregião apresenta 4 valores para o vetor de descritor conduzindo a um vetor de tamanho global de $4 \times 4 \times 4 = 64$. O descritor resultante é invariante na rotação, escala, brilho e contraste.

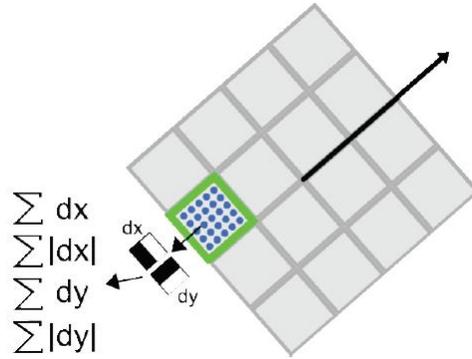


Figura 4.10: Componentes do descritor. Figura obtida em Evans (2009).

Como pode ser visto na Figura 4.10, o quadrado verde limita uma das 16 sub-regiões e os círculos azuis representam a amostragem dos pontos na qual são calculadas as respostas do *wavelet*, onde as respostas do x e y são calculados em relação à orientação dominante. Os componentes do descritor formam também o vetor de 128 posições que apresenta-se mais distinto e ainda assim eficiente para computar, entretanto é lento no casamento sobre altas dimensões, onde a somas do d_x e $|d_x|$ são computados separadamente para $d_y < 0$ e $d_y \geq 0$. Similarmente a soma do d_y e $|d_y|$ são divididos de acordo com o sinal de d_x , conseguindo desta forma o dobro da quantidade de características. Esse vetor também trabalha com 16 subregiões.

4.2.5.2 Cor

Dada uma imagem $I(x, y)$ convertida para o modelo HSV é obtida a imagem $HSV(x, y)$, a partir dos valores x e y da posição de cada ponto de interesse obtida pelo método descrito em 4.2.3. Uma imagem digital colorida é representada por três matrizes. A partir da decomposição da imagem $HSV(x, y)$ em três matrizes, obtêm as matrizes $\mathbf{H}_{m,n}$, $\mathbf{S}_{m,n}$ e $\mathbf{V}_{m,n}$. Com isso, na matriz $\mathbf{H}_{m,n}$ onde as posições m e n são respectivamente as posições x e y de cada ponto de interesse, é obtido o canal H do mesmo.

4.2.5.3 Descritor das Características

Os vetores de pontos de interesses \mathbf{x}_j e os \mathbf{c}_i possuem o seu descritor para cada ponto descrito de duas formas:

1. (descritor dos pontos de interesse): esse descritor consiste nos quatro valores obtidos nos *wavelets de Haar* das 16 subregiões, onde obtém um vetor com 64 dimensões. Somando-se o d_x e $|d_x|$ separadamente para $d_y < 0$ e $d_y \geq 0$ e similarmente a soma do d_y e $|d_y|$ divididos de acordo com o sinal de d_x serão obtidos 128 dimensões.
2. (descritor dos pontos de interesse e o canal H): consiste na união entre as características do descritor e do canal H de cada ponto de interesse detectado. Com isso o descritor irá possuir 65 ou 129 dimensões.

Se o descritor gerado é referente ao vetor \mathbf{x}_j , ele é enviado para todas as câmeras para ser salvo na memória RAM da mesma, se for o descritor referente ao vetor \mathbf{c}_i gerado, ele será enviado para o processo de reconhecimento.

4.2.6 Reconhecimento

O reconhecimento pode ser realizado de duas maneiras. A primeira utiliza apenas a comparação do sinal *Laplaciano* e a outra utiliza o sinal *Laplaciano* e a cor.

Esse processo realiza o casamento dos pontos de interesse do vetor \mathbf{x}_j , obtido pela câmera contendo os pontos representativos do alvo que deseja reidentificar, com os pontos do vetor \mathbf{c}_i extraído atualmente do quadro de uma câmera da rede. Abaixo serão apresentados as duas formas de casamento dos vetores de características.

Casamento com o Sinal Laplaciano

Para cada ponto contido no vetor \mathbf{x}_j e no vetor \mathbf{c}_i são obtidos os seus respectivos sinais Laplacianos. De acordo com o trabalho apresentado por Bay

et al. (2006), esse sinal representa o sinal positivo do ponto máximo positivo ou o sinal negativo do ponto máximo negativo obtido pela matriz Hessiana. Essa informação é útil para o casamento entre os pontos.

Para cada ponto a contido no vetor \mathbf{c}_i é casado com cada ponto b do vetor \mathbf{x}_j . Onde, é verificado se o sinal Laplaciano do ponto b é igual ao sinal Laplaciano do ponto a .

Se o sinal Laplaciano for igual é então calculada a distância entre os descritores do ponto b com os do ponto a pelo método da Soma das Diferenças Quadráticas (SDQ) como apresentado na Seção 2.2. Com isso, as similaridades entre os pontos do vetor \mathbf{c}_i com os pontos do vetor \mathbf{x}_j são obtidas através da distância entre os descritores.

Casamento com o Sinal Laplaciano e a Cor

Para cada ponto contido no vetor \mathbf{x} e no vetor \mathbf{c}_i são obtidos os seus respectivos sinais Laplacianos e as cores. Para cada ponto b contido no vetor \mathbf{c}_i é casado com cada ponto a do vetor \mathbf{x} . Desta forma, para cada ponto a é verificado se o sinal Laplaciano e a cor de b é igual ao sinal Laplaciano e a cor de algum ponto a .

Para os pontos a que forem de sinais Laplacianos e cores iguais ao do ponto b é calculada a distância entre os descritores do ponto q com os do ponto b pelo método SDQ como apresentado na Seção 2.2. Dessa forma, as similaridades entre os pontos de \mathbf{c}_i com \mathbf{x} são obtidas.

Depois do casamento, é verificado se a similaridade do vetor \mathbf{c}_i é muito próximo do vetor \mathbf{x} , utilizando o método de proximidade da vizinhança. Inicialmente a similaridade é ordenada em ordem decrescente. Se o valor da primeira similaridade de \mathbf{c}_i for menor que a segunda similaridade de \mathbf{c}_i multiplicada por um limiar l , então isso significa que o quadro atual da câmera i é a reidentificação da pessoa contida na imagem dada pelo operador.

Finalmente, o i e o quadro da câmera são enviados ao componente funcional de controle, onde o conteúdo será apresentado ao centro de vigilância.

Capítulo 5

Resultados Experimentais

Nesse trabalho as avaliações foram realizadas em quatro grupos de experimentos. O primeiro grupo utiliza uma base de imagens sintéticas para avaliar a robustez da combinação dos métodos de reconhecimentos com os vetores de características utilizando a inserção de ruídos e a escalabilidade de pessoas. O segundo grupo utiliza bases públicas com o objetivo de formar uma base de vídeos reais com informações heterogêneas como ambientes, ângulos de visão da câmera, quantidade de pessoas, campos de visão da câmera sobrepostos e não sobrepostos, tipos de câmeras e tipos de movimentos diferentes. Esse grupo avaliou na metodologia a precisão do reconhecimento a partir das características visuais obtidas desses vídeos reais. O terceiro grupo apresenta a complexidade dos métodos utilizados no segundo grupo. O quarto grupo apresenta a avaliação do tempo de execução da comunicação entre dois nós. Para melhor compreensão dos experimentos primeiramente serão especificadas as tecnologias e a configuração utilizada. Posteriormente os grupos de avaliações serão detalhados.

5.1 Especificação da Configuração Utilizada

As configurações utilizadas nos experimentos são especificadas abaixo:

1. Configuração 1:

- Processador: Pentium 4 Core Duo 3.2GHZ
- Memória: 2GB de RAM
- HD: 250 GB
- Linguagem de Programação: C++
- Biblioteca: OPENCV
- Sistema Operacional: OPENSUSE 11.1

2. Configuração 2:

- Máquina: notebook amazon pc tablet l97.
- Linguagem de Programação: C++
- Biblioteca: OPENCV
- Sistema Operacional: OPENSUSE 11.1

5.2 Fixação e Ajuste de Parâmetros para Implementação

Os experimentos foram executados com alguns parâmetros inerentes a implementação previamente ajustados, onde esses parâmetros serão discutidos nas seções seguintes.

5.2.1 Conversão entre os Modelos

Como especificado na Seção 4.2.1, a imagem descrita no modelo RGB é convertida pela função `cvCvtColor`. Abaixo são especificados os parâmetros dessa função. Na Figura 5.1 é apresentado um exemplo de conversão entre os modelos.

- a imagem RGB

- a imagem de destino
- parâmetro de conversão:
 - CV_BGR2HSV, para converter a imagem descrita no modelo RGB para o modelo HSV.
 - CV_BGR2GRAY, para converter a imagem descrita no modelo RGB para tons de cinza.

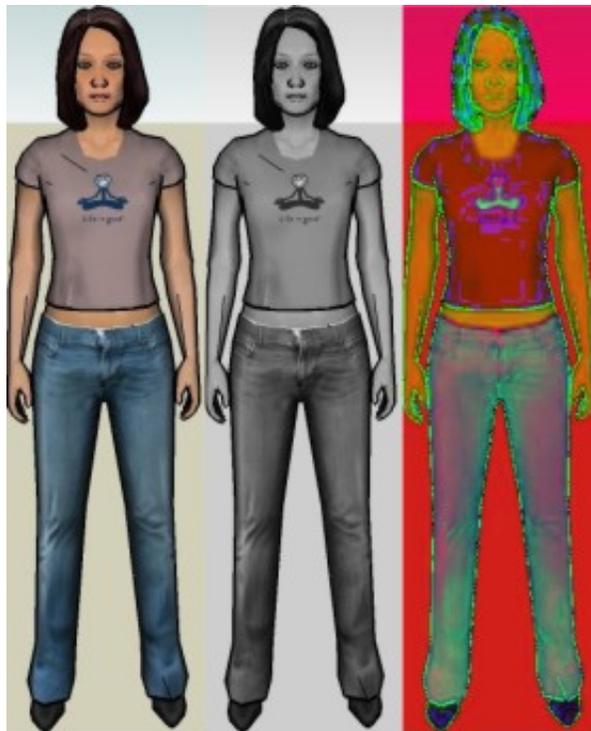


Figura 5.1: Na esquerda está a imagem descrita no modelo RGB, ao seu lado a mesma em tons de cinza e no final a mesma descrita no modelo HSV.

5.2.2 Detecção dos Pontos de Interesse

Como apresentado na Seção 4.2.3, esse detector captura os pontos de interesse a partir de cada imagem dada como entrada. Neste trabalho, foi utilizado a

função `cvExtractSurf` obtida na ferramenta proposta por Bradski (2000). Os parâmetros de entrada dessa função são:

1. a imagem de entrada, obrigatoriamente em tons de cinza;
2. a máscara que representa uma matriz com as mesmas dimensões da imagem de entrada em tons de cinza, onde os valores 1 significam a região que se deseja detectar os pontos de interesse;
3. os parâmetros do SURF:
 - o *extendido*:
 - 0: descritor básico de 64 dimensões;
 - 1: descritor *extendido* de 128 dimensões;
 - *threshold*:500;
 - número de oitavos: 3;
 - número de camadas dos oitavos: 4.

Os parâmetros de saída dessa função são:

1. lista contendo os pontos de interesse detectados:
 - posições do ponto;
 - o laplaciano;
 - a orientação;
2. lista com o descritor de cada ponto.

No caso do método que utiliza somente o canal H, o descritor não é computado. Na Figura 5.2, são apresentados os exemplos de detecção dos pontos de interesse nas imagens.

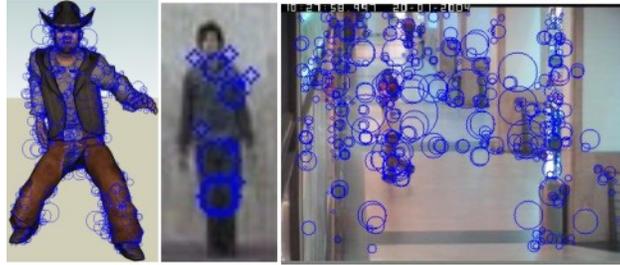


Figura 5.2: Esses são exemplos da detecção dos pontos de interesse. A primeira imagem representa a imagem sintética de consulta, a segunda imagem é a imagem de consulta da base1 e a terceira imagem é a imagem de um quadro do vídeo da base2.

5.2.3 Extração das Características

Como apresentado na Seção 4.2.5, a partir dos pontos detectados são obtidos o canal H e o descritor para formação do vetor de características de duas formas diferentes. As duas formas são:

1. usando somente o descritor;
2. usando somente o canal H e o descritor.

O descritor foi obtido no parâmetro *descriptor* da função *cvExtractSurf* proposta por Bradski (2000) utilizada no detector de pontos de interesse. O canal H foi obtido pela função *CV_IMAGE_ELEM* onde seus parâmetros são a imagem convertida para *HSV*, o tipo do resultado de retorno *uchar*, as dimensões do ponto de interesse *y* e *x*, onde a posição *x* é multiplicada por três. O valor do *x* é multiplicado por três porque a imagem está em três canais (*HSV*). Então, para obter o canal H multiplica-se *x* por três; para obter o canal S multiplica-se o *x* por três e soma-se um; e para obter o canal V multiplica-se o *x* por três e soma-se dois. Essa informação é obtida tanto na imagem dada pelo operador quanto na imagem obtida do quadro do vídeo. Os métodos de formação são apresentados nas próximas subseções.

5.3 Grupo 1: Avaliação da Metodologia usando Imagens Sintéticas

Esse grupo analisa a robustez da metodologia sobre a escalabilidade de até 38 objetos de interesse. Esse grupo desconsiderou um sistema de câmeras, pois o objetivo dessa avaliação é verificar a robustez do reconhecimento das características de uma consulta sobre um quadro com vários objetos de interesse. Para cada objeto foi verificado a reidentificação com a inserção de ruído Gaussiano somente na imagem de consulta, apenas em cada pessoa de um quadro da câmera, e considera-se também a inserção do ruído sobre a imagem de consulta e em cada pessoa.

Nesse grupo de avaliação foram analisados os dois vetores de características sobre os dois métodos de casamento com o limiar variando entre 0.7 e 0.8 e os pontos de interesses descritos em 128 dimensões, considere que o tamanho do ruído variou entre 91 a 121. Considere também a escalabilidade formada por 26 modelos de pessoas e mais 12 modelos de carros para avaliar a robutez das combinações entre o reconhecimento e o vetor de características. Considere também a inserção de escala $\frac{1}{1.5} \times$, $\frac{1}{3.0} \times$ e $\frac{1}{4.5} \times$ nas imagens de consulta. Nas próximas subseções serão apresentadas a base de imagens sintéticas e os detalhes da avaliação.



Figura 5.3: Na esquerda está a imagem descrita no modelo RGB e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 .

Na Figura 5.3, é apresentada a imagem sem ruído e com ruído 91×91 , 105×105 e 121×121 . Isto mostra que a medida que aumenta o ruído a imagem torna-se difícil de reidentificar.



Figura 5.4: Na esquerda está a imagem descrita no modelo HSV e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 .

Na Figura 5.4, é apresentada a imagem descrita no modelo HSV, utilizada para obter o canal H como característica. Ela é apresentada sem ruído e com ruído 91×91 , 105×105 e 121×121 . Pode-se observar que a medida que aumenta o ruído a imagem torna-se difícil de reidentificar a olho.

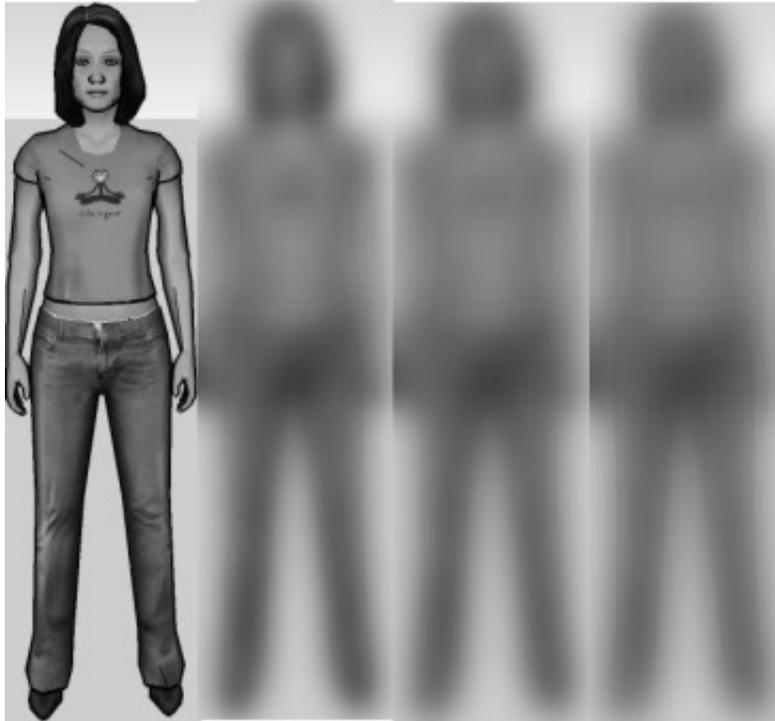


Figura 5.5: Na esquerda está a imagem em tons de cinza e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 .

Na Figura 5.5, é apresentada a imagem em tons de cinza, utilizada na integração da imagem. Ela é apresentada sem ruído e com ruído 91×91 , 105×105 e 121×121 . Pode-se observar que a medida que aumenta o ruído a imagem torna-se irreconhecível.



Figura 5.6: Na esquerda está a imagem em tons de cinza com escala $\frac{1}{1.5} \times$ e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 .

Na Figura 5.6, é apresentada a imagem em tons de cinza com uma escala $\frac{1}{1.5} \times$, utilizada na integração da imagem. Essa imagem é a imagem de consulta. Ela é apresentada sem ruído e com ruído 91×91 , 105×105 e 121×121 . Pode-se observar que a medida que aumenta o ruído dificulta o reconhecimento do que a imagem sem escala.



Figura 5.7: Na esquerda está a imagem em tons de cinza com escala $\frac{1}{3.0} \times$ e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 .

Na Figura 5.7, é apresentada a imagem em tons de cinza com uma escala $\frac{1}{3.0} \times$, utilizada na integração da imagem. Essa imagem é a imagem de consulta. Ela é apresentada sem ruído e com ruído 91×91 , 105×105 e 121×121 . Pode-se observar que a medida que aumenta o ruído dificulta o reconhecimento do que a imagem sem escala, principalmente porque essa imagem está muito menor que a sem escala.



Figura 5.8: Na esquerda está a imagem em tons de cinza com escala $\frac{1}{4.5} \times$ e sem ruído, ao seu lado a mesma com ruído de tamanho 91×91 , depois a mesma com ruído de tamanho 105×105 e no final a mesma com ruído de tamanho 121×121 .

Na Figura 5.8, é apresentada a imagem em tons de cinza com uma escala $\frac{1}{4.5} \times$, utilizada na integração da imagem. Essa imagem é a imagem de consulta. Ela é apresentada sem ruído e com ruído 91×91 , 105×105 e 121×121 . Pode-se observar que a medida que aumenta o ruído dificulta o reconhecimento do que a imagem sem escala, principalmente porque essa imagem consiste na metade do tamanho da imagem, e levando em consideração que a medida que aumenta a escala, a imagem sem ruído adicionado já apresenta um pouco de ruído.

Depois do processo de escala da imagem é realizado um processo de equalização do histograma a fim de aumentar a nitidez da imagem e em seguida é adicionado o ruído.

5.3.1 Base de Imagens Sintéticas

Essa base é formada por imagens de modelos 3D de pessoas e carros contendo um total de 38 modelos diferentes extraídos do *software Google Sketchup*.

Onde 26 são de pessoas e 12 de carros. Diferentemente da base do grupo de experimento 1, essas imagens além de serem precisas no objeto elas possuem resoluções altas onde a menor imagem possui 193 pontos por 567 pontos e a imagem maior possui 955 por 577 pontos. Na Figura 5.9, são apresentadas as imagens.



Figura 5.9: Essa figura apresenta as imagens utilizadas para representar as pessoas no lado esquerdo e algumas imagens de carro no lado direito para dificultar a reidentificação correta.

5.3.2 Avaliação da Metodologia com Inserção do Ruído Gaussiano somente na Imagem de Consulta

Nessa avaliação foi inserido o ruído com o tamanho de 91 a 121 nos modelos de pessoas utilizados como consulta, onde verificou-se a robustez do reconhecimento. A seguir serão apresentados os detalhes dos resultados separados por escala, método de casamento e vetor de característica.

5.3.2.1 Escala 1.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

Na Figura 5.10 observa-se que mesmo iniciando com uma alta taxa de ruído (91) o casamento conseguiu reidentificar mais que a metade da base. Percebe-se também que usando um limiar de 0.8 obteve resultados na maioria das vezes melhores que o limiar 0.7, mas com a taxa de ruído 99 e 107 mostrou que o limiar 0.7 obteve uma precisão acima do limiar 0.8, isto acontece pois apenas nestes casos o limiar 0.8 aumentou a taxa de erros levando à reidentificar falsos positivos. E no pior caso, com a taxa de ruído em 121 o resultado foi acima de 20% do total da base reidentificada corretamente. Este resultado é bom sobre esta alta taxa de ruído que dificilmente se encontra em uma câmera atualmente.

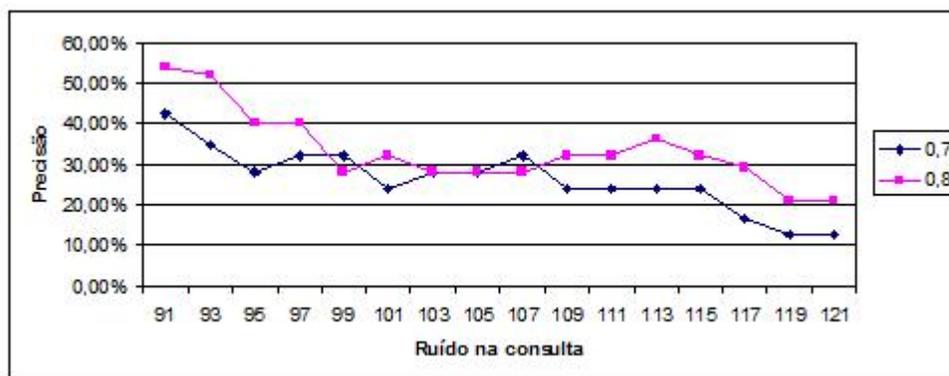


Figura 5.10: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.2.2 Escala 1.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

Na Figura 5.11 observa-se que comparado aos resultados da seção anterior o uso do canal H aumentou a taxa de erros, ou seja, o algoritmo obteve uma alta quantidade de falsos positivos. Por isso, sobre altas taxas de ruído o canal H não pode ser utilizado.

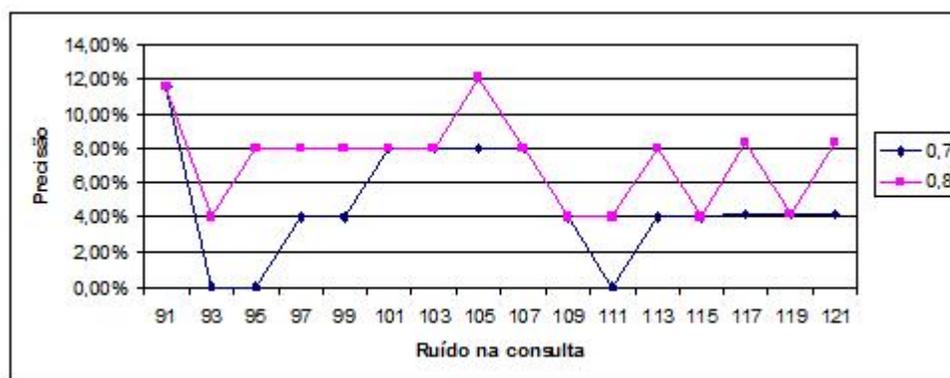


Figura 5.11: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.2.3 Escala 1.5; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

Na Figura 5.12 observa-se que comparados aos resultados da seção anterior com a taxa em 91 obteve resultados melhores que o da seção anterior entretanto com o aumento da taxa de ruído aumentou o erro, levando à reidentificar nenhum modelo de pessoas corretamente. Mesmo assim, entre as taxas de ruído 97 a 115 o casamento obteve um resultado constante apresentando ser mais robusto que os resultados anteriores.

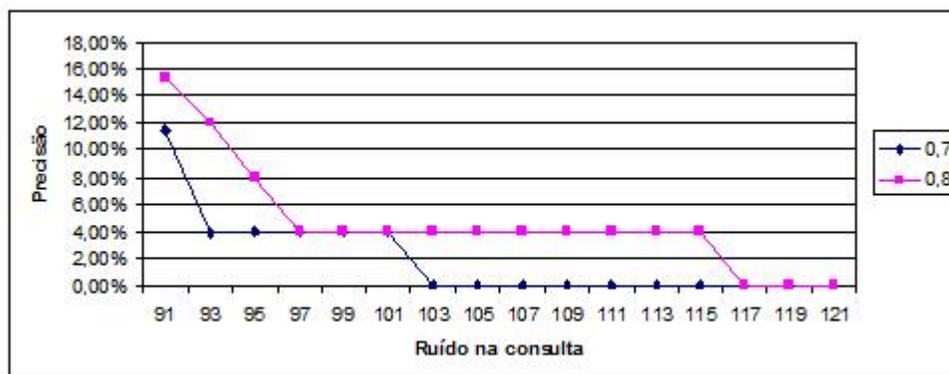


Figura 5.12: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

5.3.2.4 Escala 3.0; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

O aumento da escala para 3.0 permitiu a dificuldade na reidentificação de todos os elementos na maioria dos tamanhos de ruídos inseridos como pode ser visto na Figura 5.13. Entretanto no ruído com tamanho 99 observa-se 4% do total da base reidentificada corretamente.



Figura 5.13: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.2.5 Escala 3.0; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

O aumento da escala para 3.0 permitiu a dificuldade na reidentificação de todos os elementos em todos os tamanhos de ruídos inseridos como pode ser visto na Figura 5.14. Diferente do experimento anterior.

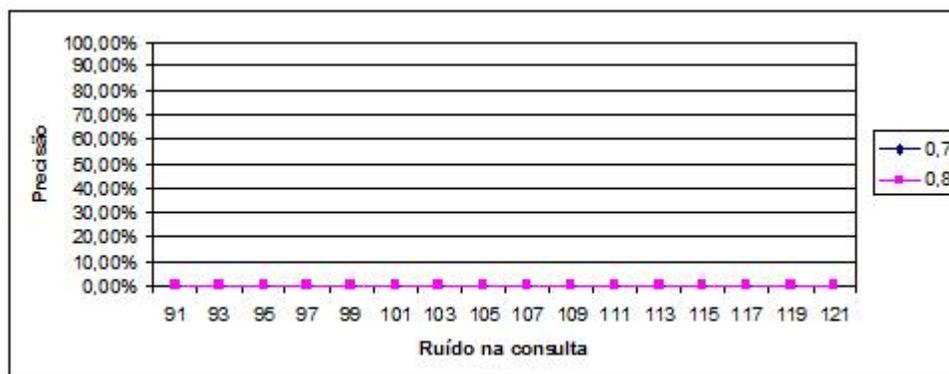


Figura 5.14: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.2.6 Escala 3.0; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

O aumento da escala para 3.0 permitiu a dificuldade na reidentificação de todos os elementos na maioria dos tamanhos de ruídos inseridos como pode ser visto na Figura 5.15. Entretanto no ruído com tamanho 99 observa-se 4% do total da base reidentificada corretamente. Ou seja, usando a cor no processo de casamento permitiu obter melhores resultados do que utilizando-o no cálculo da distância SDQ.



Figura 5.15: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

5.3.2.7 Escala 4.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

O aumento da escala para 4.5 permitiu a dificuldade na reidentificação de todos os elementos em todos os tamanhos de ruídos inseridos como pode ser visto na Figura 5.16.

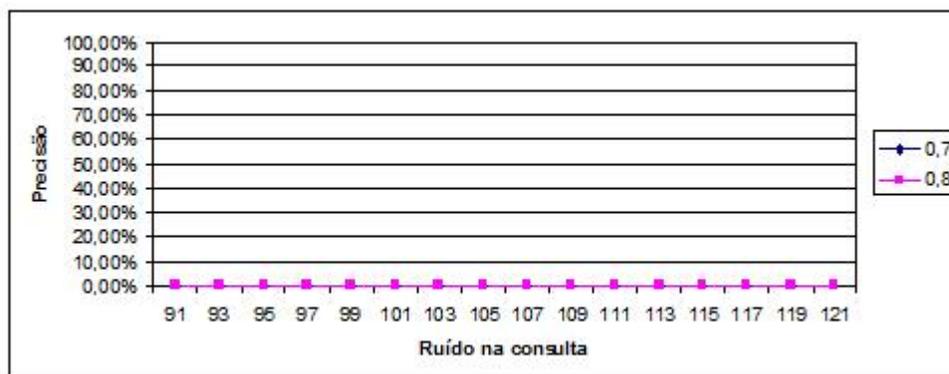


Figura 5.16: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.2.8 Escala 4.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

O aumento da escala para 4.5 permitiu a dificuldade na reidentificação de todos os elementos em todos os tamanhos de ruídos inseridos como pode ser visto na Figura 5.17.



Figura 5.17: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.2.9 Escala 4.5; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

O aumento da escala para 4.5 permitiu a dificuldade na reidentificação de todos os elementos em todos os tamanhos de ruídos inseridos como pode ser visto na Figura 5.18.



Figura 5.18: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

Portanto, com o aumento da escala para 3.0 e 4.5 os resultados não foram satisfatórios. Por isso, o uso de ruídos altos em imagens de pessoas para consulta com escalas altas torna-se difícil de reidentificá-los. Mas, dentre todos os resultados o limiar 0.8 com o método de casamento usando sinal Laplaciano e o vetor de característica formado por pontos de interesse apresenta-se os melhores resultados.

5.3.3 Avaliação da Metodologia com Inserção do Ruído Gaussiano em cada objeto contido na imagem

Nessa avaliação foi inserido o ruído com o tamanho de 91 a 121 nos modelos de pessoas utilizados na imagem que deseja reidentificar, onde verificou-se a robustez do reconhecimento. A seguir serão apresentados os detalhes dos resultados separados por escala, método de casamento e vetor de característica.

5.3.3.1 Escala 1.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

Na Figura 5.19 observa-se que no melhor caso no ruído com tamanho 103 obteve o melhor resultado com 38% de toda a base reidentificada corretamente e no pior caso 27% de toda a base reidentificada corretamente.

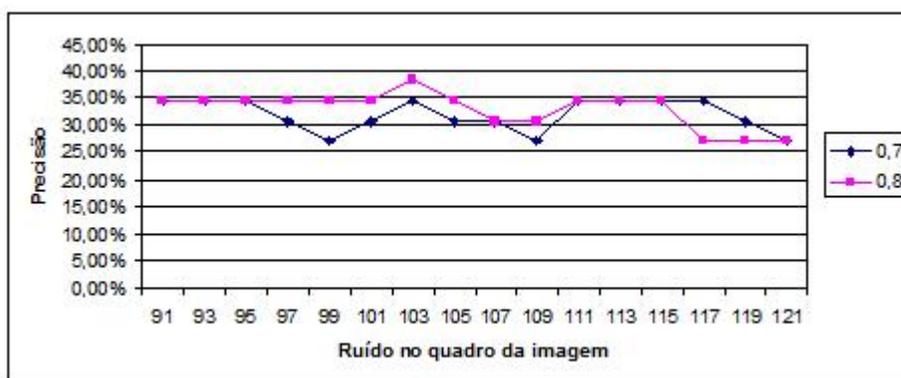


Figura 5.19: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.3.2 Escala 1.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

Na Figura 5.20 observa-se que no melhor caso no ruído com tamanho 91 obteve o melhor resultado com 38% de toda a base reidentificada corretamente o mesmo resultado apresentado no experimento anterior e no pior caso 19% de toda a base reidentificada corretamente, ou seja, menor que o pior caso do experimento anterior com o limiar 0.8. No limiar 0.7, o pior caso foi 11,5%.

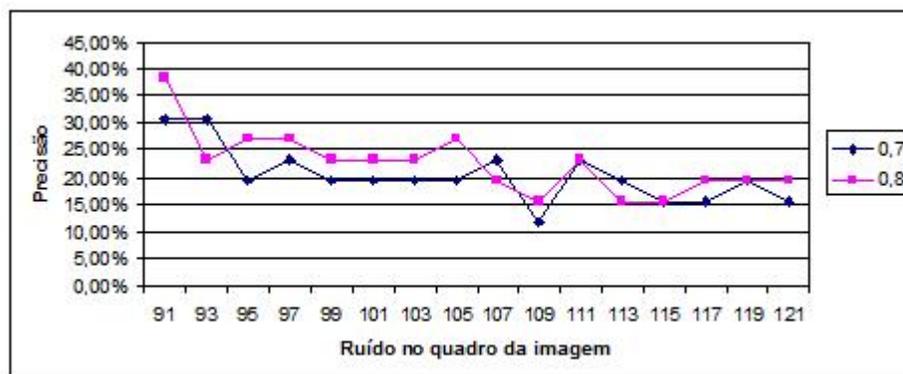


Figura 5.20: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.3.3 Escala 1.5; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

Na Figura 5.21 observa-se que o uso da cor no casamento permitiu no melhor caso 38% de toda a base reidentificada corretamente com o limiar 0.8 nos ruídos com tamanho 101 e 103 e no limiar 0.7 nos ruídos com tamanho 113 e 115. No pior caso, no limiar 0.7 obteve 15,3% de toda a base reidentificada corretamente e no limiar 0.8 obteve um resultado um pouco melhor de 19,2% de toda a base reidentificada corretamente.

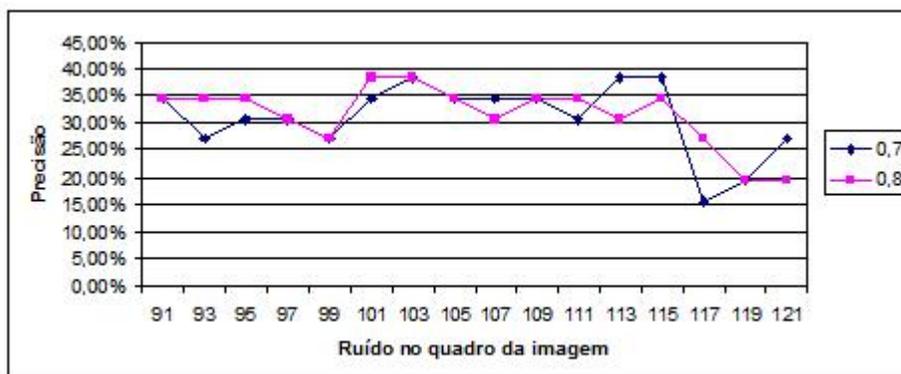


Figura 5.21: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

5.3.3.4 Escala 3.0; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

Na Figura 5.22 observa-se que no limiar 0.8 obteve o melhor caso em 38% de toda a base reidentificada corretamente e no pior caso 23% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 35% de toda a base reidentificada corretamente e no pior caso 23% de toda a base reidentificada corretamente. Ou seja, com o aumento da escala para 3.0 não afetou na reidentificação.

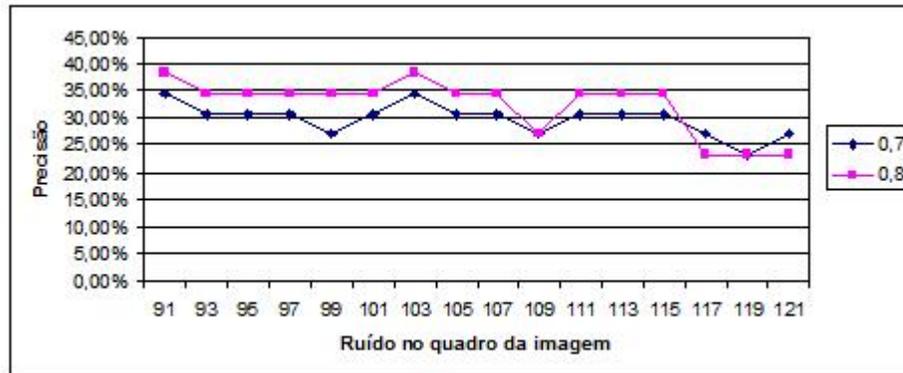


Figura 5.22: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.3.5 Escala 3.0; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

Na Figura 5.23 observa-se que no limiar 0.8 obteve o melhor caso em 35% de toda a base reidentificada corretamente e no pior caso 19% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 27% de toda a base reidentificada corretamente e no pior caso 15% de toda a base reidentificada corretamente. Ou seja, com o aumento da escala para 3.0 não afetou na reidentificação.

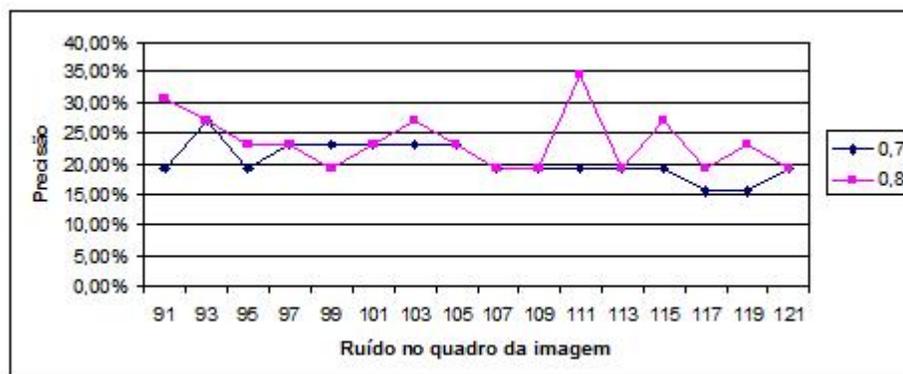


Figura 5.23: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.3.6 Escala 3.0; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

Na Figura 5.24 observa-se que no limiar 0.8 obteve o melhor caso em 38% de toda a base reidentificada corretamente e no pior caso 23% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 38% de toda a base reidentificada corretamente e no pior caso 30% de toda a base reidentificada corretamente. Ou seja, com o aumento da escala para 3.0 aumentou a quantidade de pessoas reidentificadas corretamente no pior caso.

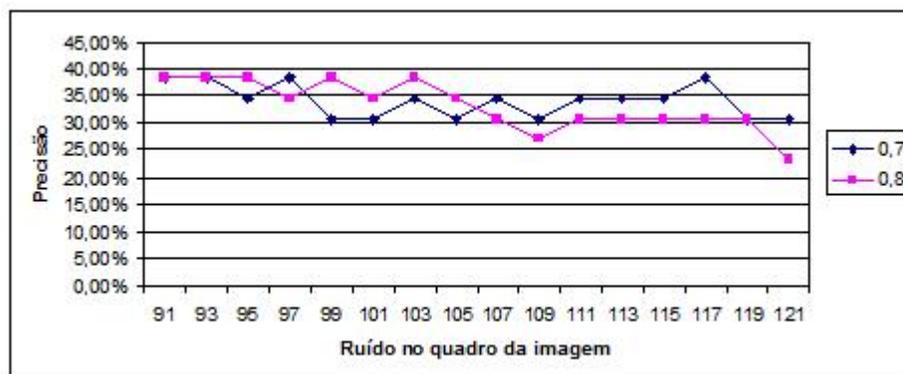


Figura 5.24: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

5.3.3.7 Escala 4.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

Na Figura 5.25 observa-se que no limiar 0.8 obteve o melhor caso em 47% de toda a base reidentificada corretamente e no pior caso 23% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 42% de toda a base reidentificada corretamente e no pior caso 23% de toda a base reidentificada corretamente. Ou seja, com o aumento da escala para 4.5 aumentou a quantidade de pessoas reidentificadas no melhor caso.

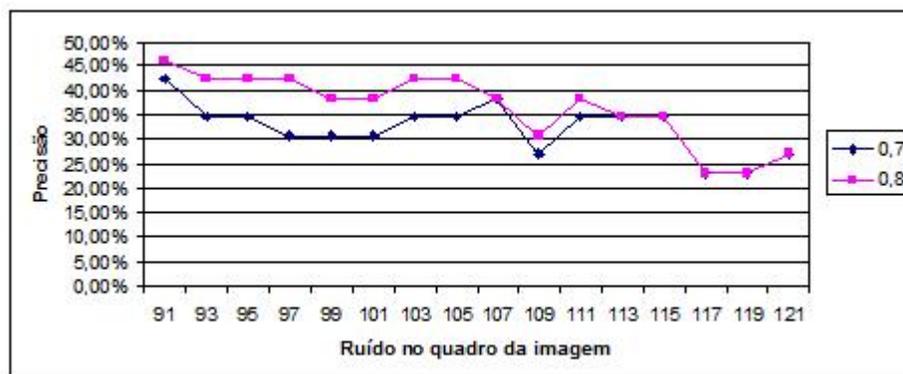


Figura 5.25: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.3.8 Escala 4.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

Na Figura 5.26 observa-se que no limiar 0.8 obteve o melhor caso em 35% de toda a base reidentificada corretamente e no pior caso 19% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 35% de toda a base reidentificada corretamente e no pior caso 15% de toda a base reidentificada corretamente. Ou seja, com o aumento da escala para 4.5 melhorou os resultados para o limiar 0.7 que obteve um acréscimo de 8% de toda a base reidentificada corretamente.

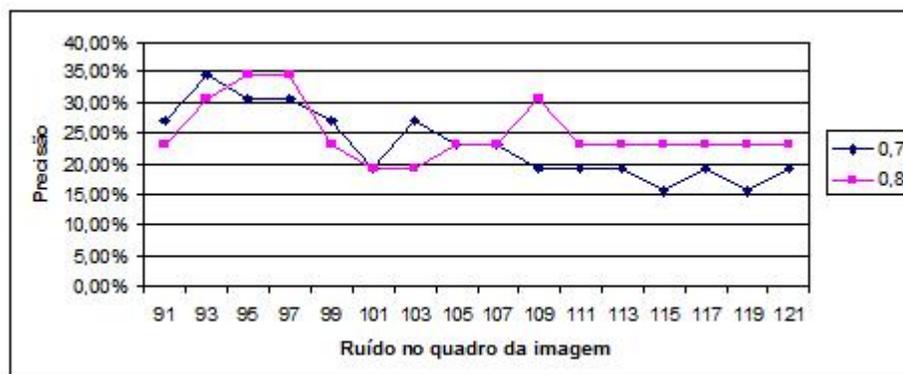


Figura 5.26: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.3.9 Escala 4.5; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

Na Figura 5.27 observa-se que no limiar 0.8 obteve o melhor caso em 42% de toda a base reidentificada corretamente e no pior caso 26% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 42% de toda a base reidentificada corretamente e no pior caso 23% de toda a base reidentificada corretamente. Ou seja, com o aumento da escala para 4.5 melhorou os resultados para o limiar 0.7 e 0.8 que obteve um acréscimo de 4% de toda a base reidentificada corretamente e no pior caso para o limiar 0.8 um acréscimo de 3% de toda a base reidentificada corretamente e para o limiar 0.7 um decréscimo de 7% de toda a base reidentificada corretamente.

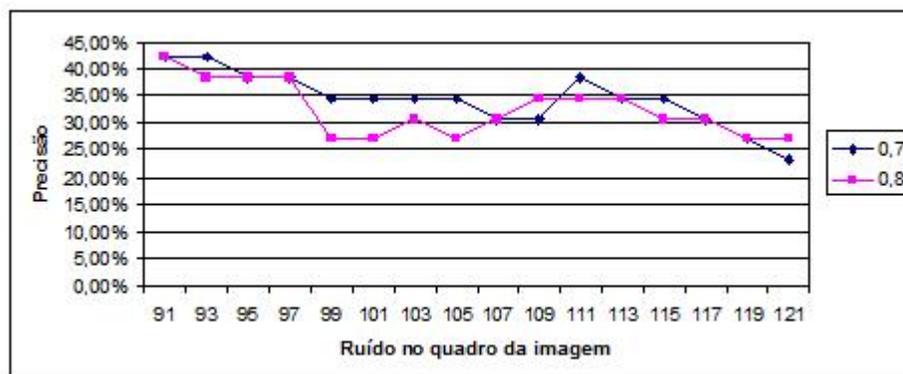


Figura 5.27: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

Portanto, com o aumento da escala nos objetos contidos na image aumentou a qualidade na reidentificação permitindo que dentre todos os casos o limiar 0.8 com o método de casamento usando sinal Laplaciano e o vetor de característica formado por pontos de interesse apresenta-se os melhores resultados.

5.3.4 Avaliação da Metodologia com Inserção do Ruído Gaussiano na Imagem de Consulta e em cada Objeto da Imagem

Nessa avaliação foi inserido o ruído com o tamanho de 91 a 121 nos modelos de pessoas utilizados como consulta e nos modelos usados como objetos na imagem que deseja-se reidentificar, onde verificou-se a robustez do reconhecimento. A seguir serão apresentados os detalhes dos resultados separados por escala, método de casamento e vetor de característica.

5.3.4.1 Escala 1.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

Na Figura 5.28 observa-se que no limiar 0.8 obteve o melhor caso em 53% de toda a base reidentificada corretamente e no pior caso 21% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 42% de toda a base reidentificada corretamente e no pior caso 12% de toda a base reidentificada corretamente.

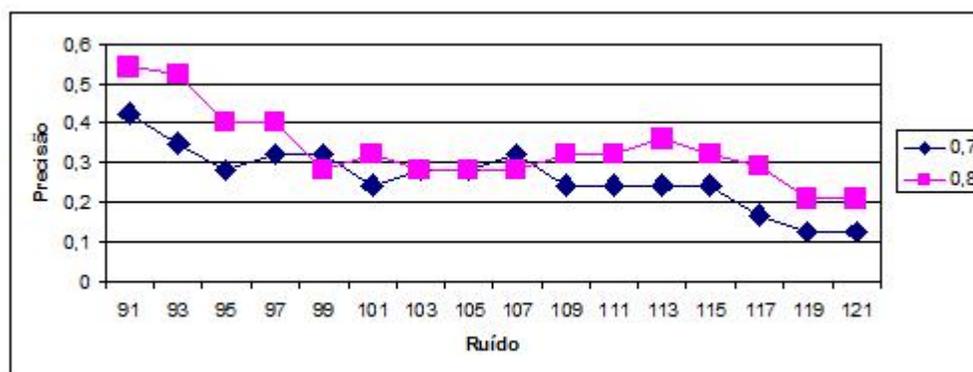


Figura 5.28: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.4.2 Escala 1.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

Na Figura 5.29 observa-se que no limiar 0.8 obteve o melhor caso em 12% de toda a base reidentificada corretamente e no pior caso 4% de toda a base reidentificada corretamente. E no limiar 0.7 obteve o melhor caso em 12% de toda a base reidentificada corretamente e no pior caso nenhuma pessoa reidentificada.

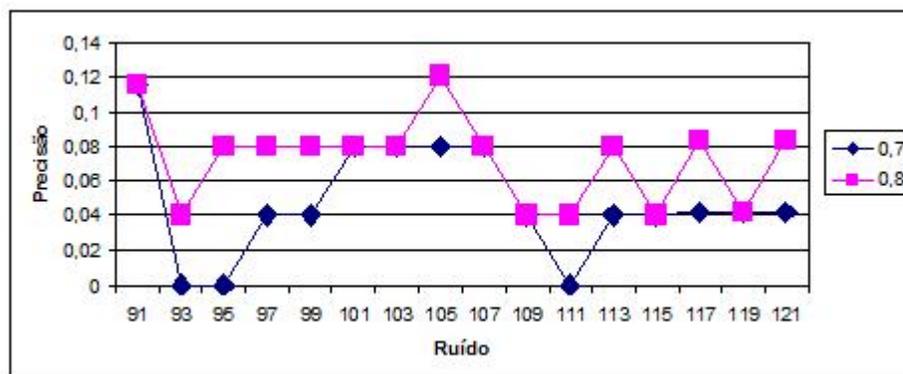


Figura 5.29: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.4.3 Escala 1.5; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

Na Figura 5.30 observa-se que no limiar 0.8 obteve o melhor caso em 15% de toda a base reidentificada corretamente e no pior caso nenhuma pessoa reidentificada. E no limiar 0.7 obteve o melhor caso em 12% de toda a base reidentificada corretamente e no pior caso nenhuma pessoa reidentificada.

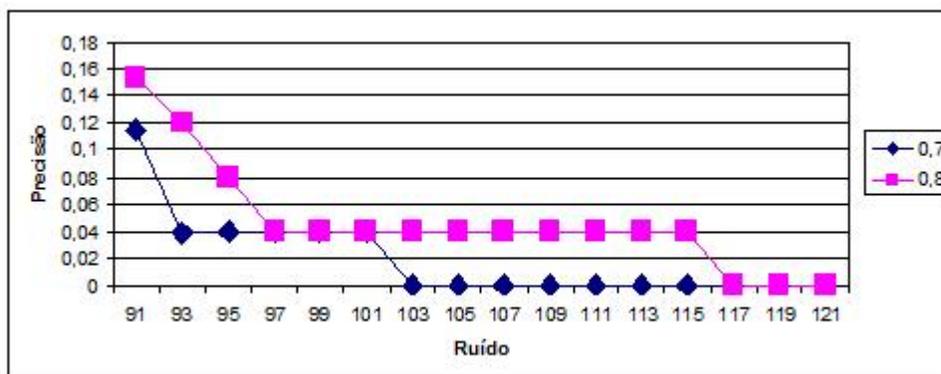


Figura 5.30: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 1.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

5.3.4.4 Escala 3.0; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

Na Figura 5.31 observa-se que no limiar 0.8 obteve o melhor caso em 4% de toda a base reidentificada corretamente e no pior caso nenhuma pessoa reidentificada. E no limiar 0.7 não obteve alguma pessoa reidentificada.

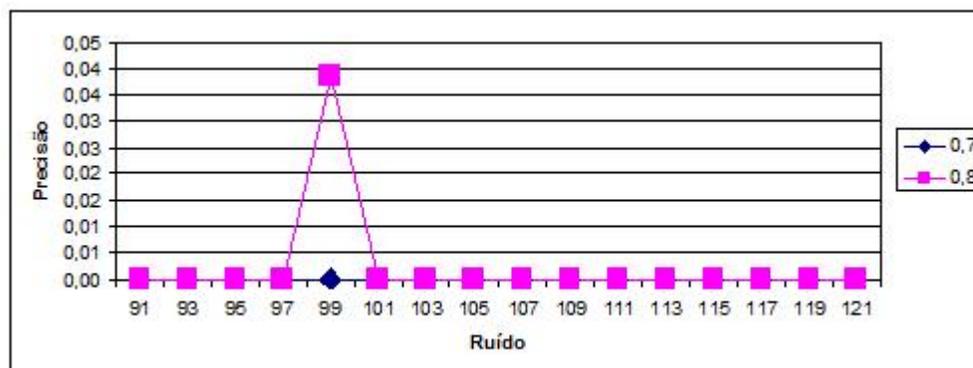


Figura 5.31: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.4.5 Escala 3.0; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

Na Figura 5.32 observa-se que no limiar 0.8 e no limiar 0.7 nenhuma pessoa foi reidentificada corretamente.

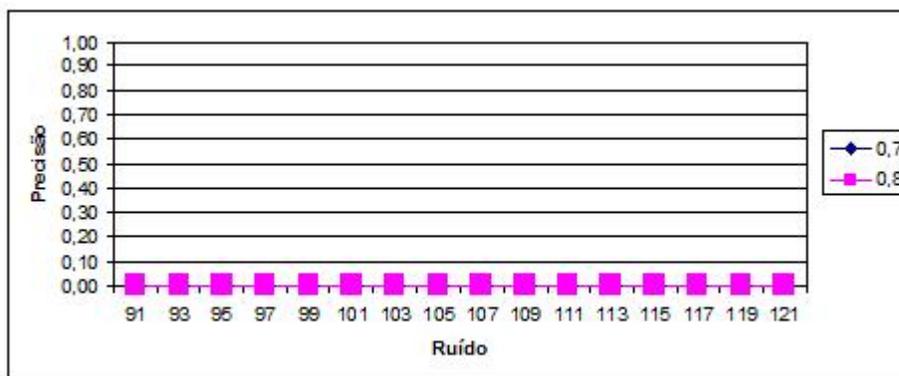


Figura 5.32: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.4.6 Escala 3.0; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

Na Figura 5.33 observa-se que no limiar 0.8 obteve o melhor caso em 4% de toda a base reidentificada corretamente e no pior caso nenhuma pessoa reidentificada. E no limiar 0.7 não obteve alguma pessoa reidentificada.

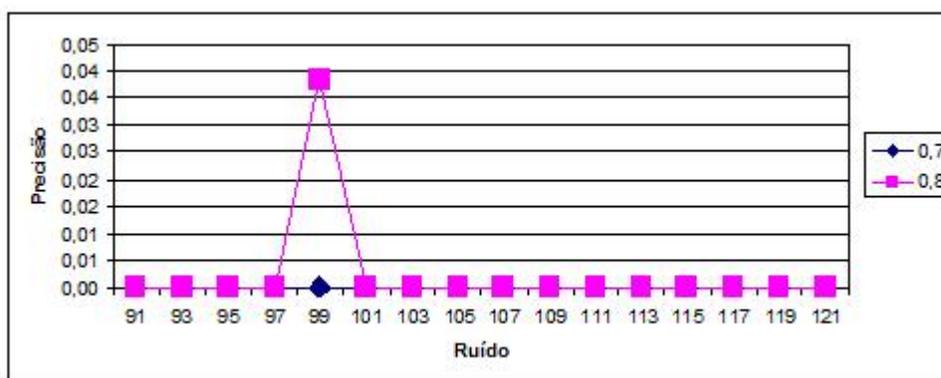


Figura 5.33: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 3.0, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

5.3.4.7 Escala 4.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse

Na Figura 5.34 observa-se que no limiar 0.8 e no limiar 0.7 nenhuma pessoa foi reidentificada corretamente.

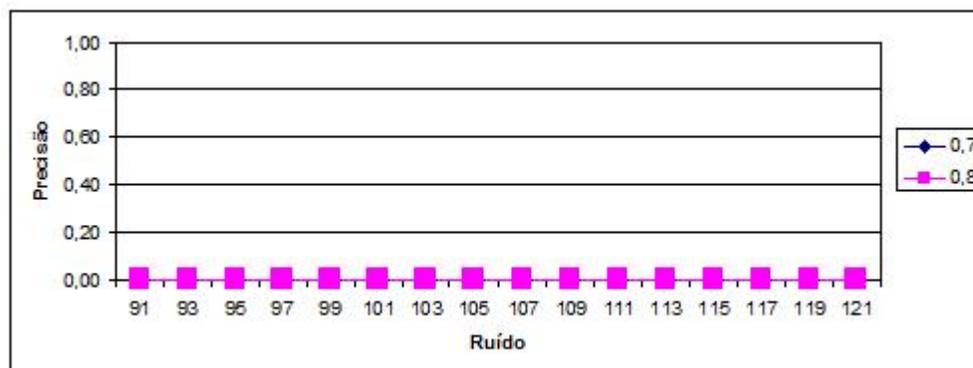


Figura 5.34: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse.

5.3.4.8 Escala 4.5; Método de Casamento usando sinal Laplaciano; Vetor de Característica formado por pontos de interesse e canal H

Na Figura 5.35 observa-se que no limiar 0.8 e no limiar 0.7 nenhuma pessoa foi reidentificada corretamente.

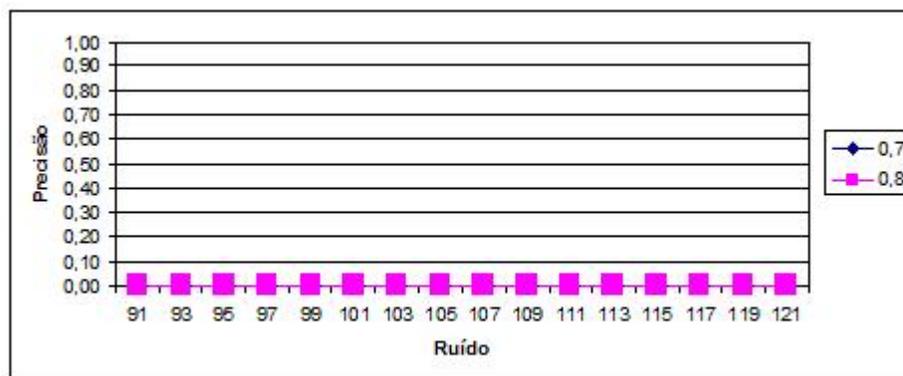


Figura 5.35: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e vetor de característica formado pelos pontos de interesse e a cor.

5.3.4.9 Escala 4.5; Método de Casamento usando sinal Laplaciano e a cor; Vetor de Característica formado por pontos de interesse

Na Figura 5.36 observa-se que no limiar 0.8 e no limiar 0.7 nenhuma pessoa foi reidentificada corretamente.

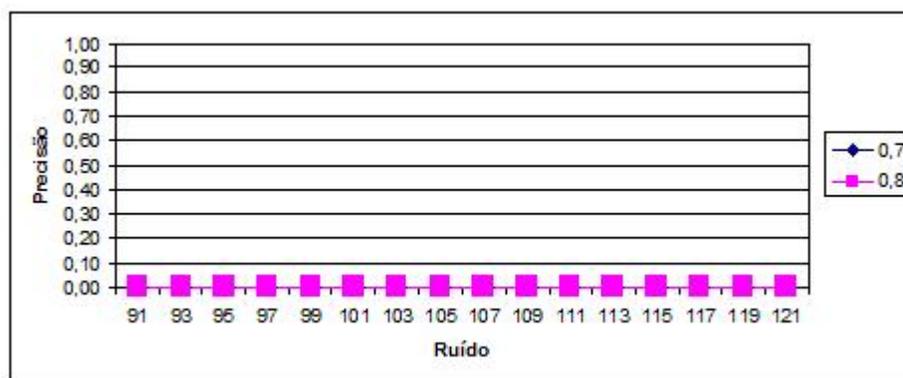


Figura 5.36: Apresentação dos resultados com o ruído variando de 91 a 121 com escala 4.5, usando no casamento o sinal Laplaciano e a cor e vetor de característica formado pelos pontos de interesse.

Portanto, com o aumento da escala para 3.0 e 4.5 os resultados não foram satisfatórios. Por isso, o uso de ruídos altos em imagens de pessoas para consulta e nos objetos contidos na imagem com escalas altas torna-se difícil de reidentificá-los. Mas, dentre todos os resultados o limiar 0.8 com o método de casamento usando sinal Laplaciano e o vetor de característica formado por pontos de interesse apresenta-se os melhores resultados.

5.4 Grupo 2: Avaliação da Metodologia com Vídeos Reais

O Grupo 2 de experimentos avalia a precisão da etapa de reconhecimento para cada vetor de característica formado pelas características obtidas nos vídeos

reais. Esse grupo avalia a precisão das imagens de consulta que consistem do fornecimento dos dados por meio da seleção da câmera sobre a base de vídeos reais. Esse grupo mostra que a partir de uma imagem de um ângulo da pessoa é possível obter sua reidentificação em qualquer câmera com diferentes ângulos de visão da mesma com taxas acima de 80% de precisão. A seguir para melhor compreensão, serão apresentadas as informações sobre as bases utilizadas nesse experimento, e avaliação da precisão do reconhecimento sobre os diferentes vetores apresentados na seção 4.2.5.3.

5.4.1 Descrição da Base Pública de Vídeos Reais

Em Cantata (2000) é apresentada uma grande quantidade de bases públicas de vídeos sobre pessoas, carros e imagens médicas. Dessa fonte foram obtidas as bases utilizadas neste trabalho, como:

- base1: Obtida em Actions (2005). Essa base possui altas irregularidades nas seqüências dos vídeos de cada ação. Ela possui 10 vídeos de pessoas andando em diferentes cenários com diferentes fundos não uniformes e realizando diversos movimentos. Esses vídeos possuem baixa sensibilidade com parciais oclusões e deformações não-rígidas em cada movimento conforme apresentado na Figura 5.37. Essa base possui 7133 imagens pré-classificadas nos movimentos e na pessoa que os realiza. No processo de aquisição de imagens foi setado o *fps* (*frames per second*) para 1, já que a quantidade da base é pequena.



Figura 5.37: Esses quadros são seqüências extraídas dos vídeos da base ACTIONS, os quais apresentam diferentes ações e diversas deformações.

- base2: Obtida em Caviar (2004). Essa base apresenta uma quantidade de vídeos gravados em diferentes cenários que incluem pessoas se movimentando dentro de um shopping. Esses vídeos possuem duas visões panorâmicas, onde uma é no corredor e a outra é na frente de uma loja.

Essas 2 câmeras possuem sobreposição no campo de visão. A resolução é 384 por 288 pontos com 25 quadros por segundo e comprimidos em MPEG2. Na Figura 5.38, são apresentados exemplos de imagens dessa base. Diferente da base1, essa base não é pré-classificada nas pessoas possuindo 72466 imagens. No processo de aquisição de imagens foi setado o *fps* (*frames per second*) para 25, já que a quantidade da base é grande.



Figura 5.38: Essas imagens são seqüências obtidas dos vídeos da base CAVIAR. Essas medidas apresentadas nas figuras foram descartadas neste trabalho.

5.4.2 Descrição da Base de Consulta

As imagens utilizadas para consulta foram extraídas das bases públicas de vídeos. Cada imagem representa um ângulo de visão de cada pessoa. Os experimentos de formação do vetor de características irão analisar a reidentificação de qualquer pessoa em diferentes ângulos de visão pelas câmeras da rede utilizando apenas a imagem de um ângulo de visão da pessoa, em vez de usar como consulta um modelo de cada pessoa com os seus respectivos ângulos de visualização como utilizado no trabalho desenvolvido em Hamdoun et al. (2008). Nas Figuras 5.39 e 5.40 são apresentadas as consultas de cada base.



Figura 5.39: Essas imagens de consulta representam as pessoas da base1. São imagens extraídas de visão frontal da pessoa com pouca deformação.



Figura 5.40: Essas imagens de consulta representam as pessoas da base2. São imagens extraídas de visão frontal e traseira da pessoa com pouca deformação. E com visão panorâmica.

5.4.2.1 Avaliação do Reconhecimento do Método de Casamento do Sinal Laplaciano com o Vetor de Características formado pelos Pontos de Interesse

O reconhecimento irá determinar a lista de reidentificações de uma imagem de consulta. Nesse trabalho o reconhecimento considera no mínimo 2 pontos correspondentes para a imagem do quadro de um vídeo ser considerado similar a imagem de consulta, ou seja, a imagem do quadro pertence a lista de reidentificações. E o limiar como apresentado na seção 4.2.6 serve para filtrar os pontos vizinhos mais próximos pela menor distância. Este trabalho foi analisado sobre quatro limiares 0.5, 0.6, 0.7 e 0.8. E o reconhecimento utilizado foi o apresentado na seção 4.2.6.

Essa avaliação verificou a precisão do vetor de características formado por pontos de interesse sobre as bases públicas de vídeo utilizando as imagens de consulta de cada base, o seu resultado é apresentado na Figura 5.41.

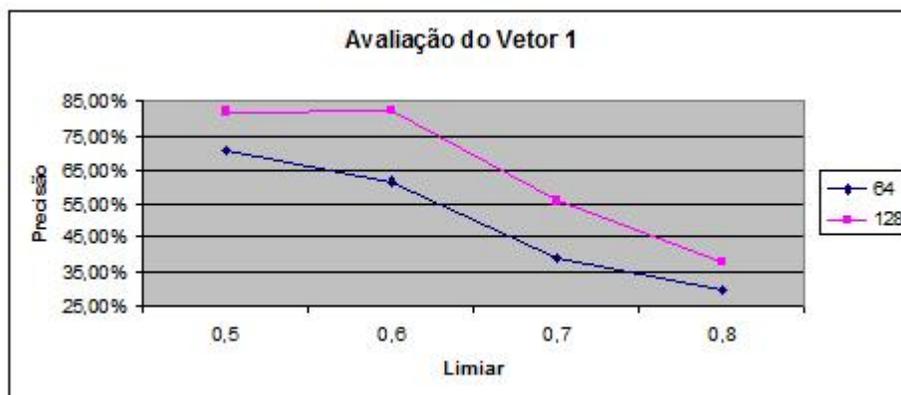


Figura 5.41: Esse gráfico apresenta a precisão do vetor 1 sobre o reconhecimento utilizando os parâmetros de limiar e dimensões do descritor.

De acordo com a Figura 5.2 pode-se visualizar que no vetor de 64 dimensões a função é decrescente sobre o limiar, ou seja, a medida que aumenta o limiar, a taxa de erro também aumenta levando a precisão a diminuir. Entretanto, entre o limiar 0.5 e o 0.6 no vetor com 128 dimensões, a função é crescente e depois do 0.6 ela transforma-se em uma função decrescente, isto mostra que o aumento da taxa de erro acontece quando o limiar é maior que 0.6. Pode-se visualizar também que a medida que aumenta o limiar as 2 funções (vetor de 64 dimensões e o vetor de 128 dimensões) sempre serão paralelas. De acordo com os resultados obtidos em Hamdoun et al. (2008) que foram de 82% de precisão, este trabalho obteve 82,2% de precisão utilizando uma única imagem para reidentificação e o descritor com 128 dimensões, obtendo desta forma uma reidentificação mais eficiente que a apresentada em Hamdoun et al. (2008), onde utiliza uma estrutura KD-Tree que é formada por 21 imagens de cada pessoa e um vetor de 64 posições. Abaixo pode-se observar exemplos difíceis de reidentificar, pois apresentam mudanças em seus vestimentos e em alguns casos um pouco de oclusão.



Figura 5.42: Imagem apresentando uma reidentificação da consulta, onde o ambiente apresenta uma grande quantidade de pessoas e ruído.

Na Figura 5.42, observa-se diferenças no vestimento entre a consulta e o quadro da câmera e principalmente diferenças no ângulo, onde na consulta a pessoa está visualizada de frente e no quadro da câmera a pessoa está de costa dificultando muito pois a visão de costa não possui características visíveis a ponto de reidentificar facilmente. Mas, o sistema conseguiu mesmo assim reidentificar os pontos de interesse obtidos da calça da imagem de consulta. Essa figura utilizou o vetor de 128 dimensões e o limiar 0.8.



Figura 5.43: Imagem apresentando uma reidentificação da consulta onde aparece somente a pessoa que se deseja reidentificar.

Na Figura 5.43, é apresentada a reidentificação da consulta com o vetor

de 128 dimensões e usando limiar o 0.5. Observa-se nessa figura diferenças no vestimento de tal forma que dificultasse até para um operador reidentificá-lo. Entretanto, o sistema conseguiu reidentificá-lo através dos pontos de interesse da camisa extraídos na consulta. Percebe-se, também que se trabalha-se com a cor poderia retornar pontos como falsos positivos no caso o próprio chão.



Figura 5.44: Imagem apresentando uma reidentificação da consulta onde aparece um conjunto de pessoas próximo a pessoa que se deseja reidentificar.

Na Figura 5.44, é apresentada a reidentificação da consulta com o vetor de 128 dimensões e usando o limiar 0.6 e 0.7. Observa-se entre a consulta e o quadro da câmera diferenças no vestimento e na disposição da pessoa. No entanto, o sistema reidentificou os pontos de interesse do rosto e um ponto da camisa apresentados na consulta.



Figura 5.45: Imagem apresentando uma reidentificação da consulta onde aparece um conjunto de pessoas próximo a pessoa que se deseja reidentificar e essa pessoa possui oclusão.

Na Figura 5.45, é apresentada a reidentificação da consulta com o vetor de 128 dimensões e utilizando como limiar o 0.6, pode-se observar que existe oclusão do objeto de consulta. Para um operador poderia ter dificuldades em encontrar essa pessoa visualizando cada vídeo. Mas, o sistema conseguiu reidentificar este caso através dos pontos de interesse obtidos na consulta.

5.5 Grupo 3: Avaliação da Complexidade dos Algoritmos

5.5.1 Integral de Imagens

A complexidade da integral da imagem é $\theta(m * n)$ onde m consiste na quantidade de linhas e n a quantidade de colunas da imagem.

5.5.2 Detecção dos Pontos de Interesse

A complexidade da detecção dos pontos de interesse é $\theta(o * i * r_m * r_n)$ onde o é a quantidade de oitavos, como apresentado na seção 5.2.2 a quantidade de oitavos é 3, i é a quantidade de camadas dos oitavos, como apresentado

na seção 5.2.2 a quantidade de camadas de oitavos é 4, e r_m e r_n são as quantidades de respostas para cada *pixel* na imagem onde o r_m varia entre o tamanho do filtro usado e a altura da imagem e o r_n varia entre o tamanho do filtro usado e a largura da imagem. Onde pode ser simplificado para $\theta(r_m * r_n)$ já que a quantidade de oitavos e camadas de oitavos é constante.

5.5.3 Orientação

A orientação possui complexidade $O((\pi + 1) * ta)$ onde ta é o tamanho do ângulo do eixo x de cada ponto de interesse.

5.5.4 Descritor dos pontos de interesse

O descritor é $\theta(l)$, onde l é a quantidade de pontos de interesse detectados, pois a extração das características são cálculos matemáticos usando 4 loops dando um custo total de 625 para cada ponto de interesse.

5.5.5 Reconhecimento do Método de Casamento do Sinal Laplaciano com o Vetor de Características formado pelos Pontos de Interesse

A complexidade do reconhecimento depende da quantidade de pontos q do quadro da câmera e a quantidade de pontos c da consulta. Desta forma, a complexidade do reconhecimento é $O(q * c)$. Enquanto que no trabalho desenvolvido em Hamdoun et al. (2008) utiliza no reconhecimento uma kd-tree, onde a complexidade é $O(n \log n)$ Este algoritmo é mais eficiente que o algoritmo de reconhecimento desenvolvido em Hamdoun et al. (2008) pois a quantidade de pontos c trabalhada é menor do que em Hamdoun et al. (2008), pois os pontos são obtidos de apenas uma imagem de consulta neste trabalho enquanto que no trabalho desenvolvido em Hamdoun et al. (2008) usa 21 imagens de uma consulta. E como apresentado na seção 5.4 a precisão foi a mesma obtida em Hamdoun et al. (2008).

5.6 Grupo 4: Avaliação da Comunicação entre Nós

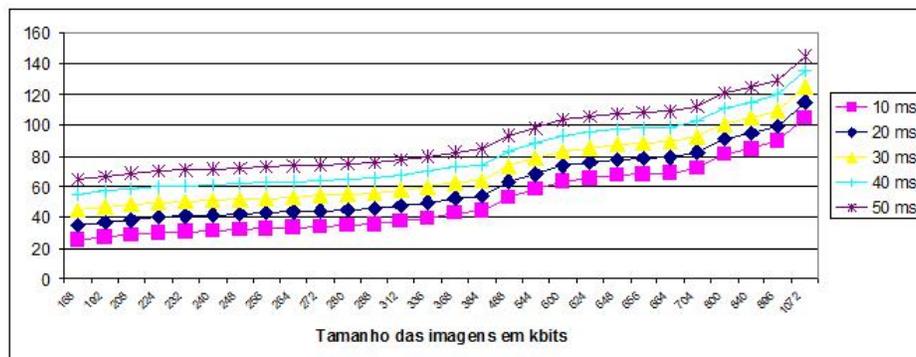
Esse grupo de experimentos avalia, sob o ponto de vista teórico, o comportamento da comunicação levando-se inicialmente em consideração o tamanho das imagens, o tempo de espera e o tamanho da banda entre 2 nós da rede e depois a quantidade de multi-saltos para uma rede padronizada.

5.6.1 Parâmetros da rede

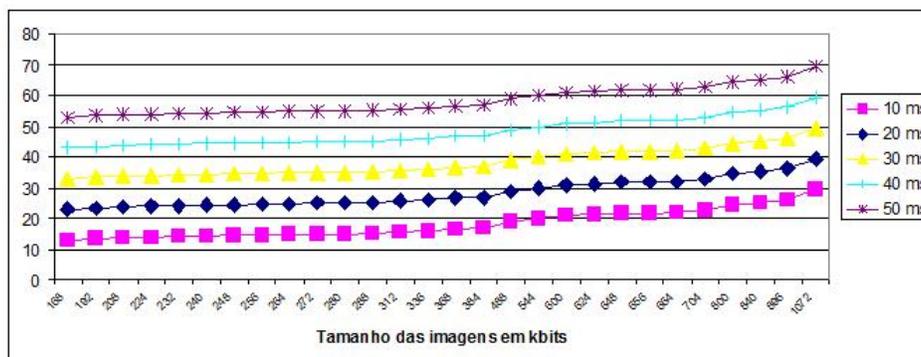
O sistema de comunicação utiliza uma rede em malha com protocolo TCP (*Transport Process Control*) e com os seguintes parâmetros:

- tempo de espera: 10, 20 e 30 ms.
- tamanho da banda: 11, 54 e 108 Mbps.

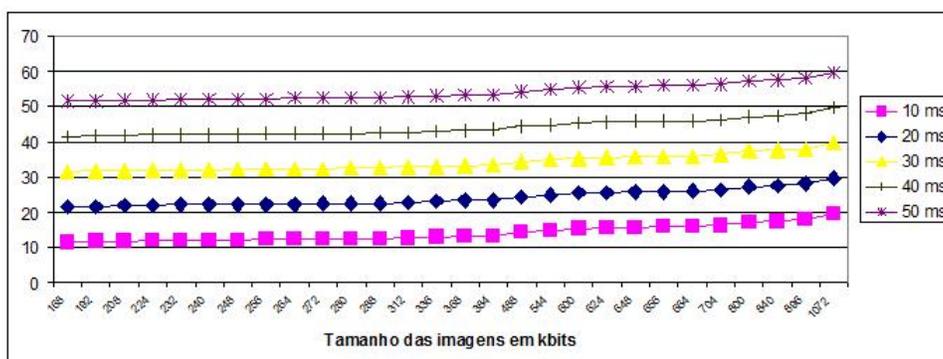
Será utilizada nessa avaliação a fórmula para calcular o tempo de envio de uma informação, conforme apresentado na seção 2.1.1.1. A informação utilizada para a comunicação consiste em cada imagem da base de imagens sintéticas.



(a) 11 Mbps



(b) 54 Mbps



(c) 108 Mbps

Figura 5.46: As Figuras 5.6.1, 5.6.1 e 5.6.1 apresentam o tempo consumido em um envio entre dois nós para cada imagem enviada ou recebida variando o tempo de espera entre 10 ms a 50 ms.

Na Figura 5.6.1 é mostrado que a medida que o tempo de espera aumenta, o tempo para envio de uma informação também aumenta. Isto acontece porque o tempo de espera é somado ao tamanho da imagem dividido pela largura de banda. Pode-se observar que entre uma rede com largura de banda de 11 Mbps, é duas vezes mais demorado o tempo de envio do que em uma rede com largura de banda de 54 Mbps. E observa-se também que mesmo duplicando a largura de banda de 54 Mbps para 108 Mbps, o tempo de envio é o mesmo até o valor 488 onde o tempo ultrapassa de 20 ms diferente do 108 Mbps que o limite máximo é de 20 ms. A medida que aumenta a largura de banda a função vai se aproximando de uma função constante. Com isso pode-se concluir que o melhor caso seria utilizar uma rede de 108 Mbps, com tempo de espera de 10 ms. Entretanto, essa rede é cara. Para uma rede barata pode ser utilizada uma rede com banda de 54 Mbps já que o tempo de espera é no máximo 10 ms a mais que a banda de 108 Mbps e com o mesmo tempo de espera.

A vantagem em utilizar uma rede de malha consiste em um nó distribuir a informação que recebe entre seus nós vizinhos. Com isso um nó pode enviar uma resposta utilizando os nós vizinhos para chegar a outro nó. Cada nó utilizado para chegar a um nó destino é chamado de salto. Desta forma, o tempo total consumido para uma consulta por segundo consiste no tempo consumido com o processamento local da reidentificação somado ao tempo consumido dos saltos para enviar a informação utilizando um caminho de nós. Esse tempo total deve ser menor que um segundo para garantir o retorno da informação em tempo real para o operador. Pois, o tempo para obter as imagens de cada câmera consiste em 1 s, e para não obter informações desatualizadas, é necessário que o resultado alcance o nó da central de vigilância no máximo em 1 s.

5.6.2 Avaliação de Multi-Saltos

Como visto na seção anterior teoricamente a configuração de rede aconselhável para utilizar neste trabalho seria com a largura de banda 54 Mbps e com

o tempo de espera de 10 ms. Essa avaliação verifica a quantidade de saltos que podem ser realizados nesta configuração de rede para o envio de uma informação de 134 KB de informação para enviar, considerando a informação com maior quantidade de *KB* da seção anterior. Na Figura 5.47 apresenta os resultados.

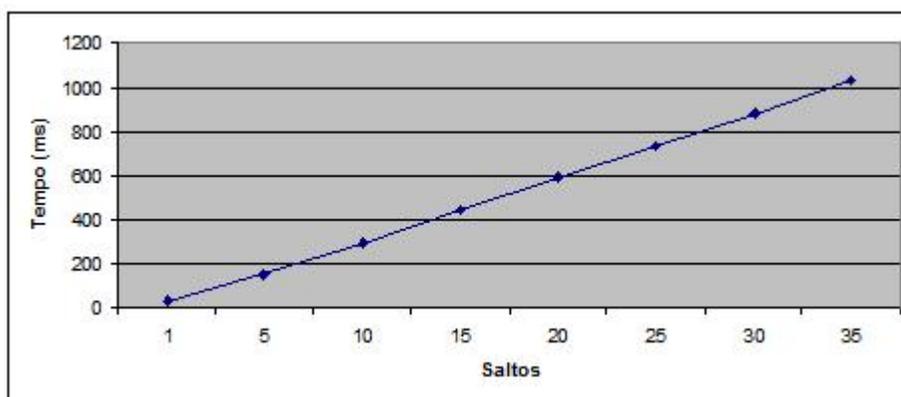


Figura 5.47: Apresentação dos tempos de envios variando os saltos.

Na Figura 5.47 observa-se que não existe a necessidade de uma organização entre os nós da rede somente um limite de saltos que devem ser seguidos onde este limite é de 35 saltos entre o nó do centro de vigilância e a câmera mais distante. Onde como visto na Figura 5.47 com 35 saltos é obtido um tempo de envio de 1 s.

Capítulo 6

Conclusão

Este trabalho abordou o problema de reidentificação de pessoas em uma rede de câmeras. Para solucionar este problema, foi necessário desenvolver e adaptar um conjunto de técnicas relacionadas com as áreas de Visão Computacional, Reconhecimento de Padrões e Sistemas Distribuídos.

A implementação da metodologia para reidentificação de pessoas levou em consideração o uso da aquisição das imagens, a detecção de pontos de interesse, extração das características de cada ponto e o reconhecimento destas características. Tais características são a cor e o descritor de cada ponto de interesse, combinados em um espaço *wavelet*. A cor representa o canal H de cada ponto descrito no modelo HSV. O descritor dos pontos de interesse é invariante na rotação, escala, brilho e no contraste.

Desenvolveu-se um estudo aprofundado dentre os quais estão relacionados com os problemas existentes com os Sistemas de Visão Computacional Distribuída, como oclusões, ruídos, baixa resolução da imagem e alta taxa de quadros por segundo.

Por meio dos experimentos em imagens sintéticas analisou-se a robustez da metodologia sobre a escalabilidade de pessoas, variação na escala das imagens de consulta e na adição do ruído entre 91 a 121, desta forma o uso do descritor de pontos de interesses representados em 128 dimensões e o método de casamento usando o sinal Laplaciano foram considerados os mais

robustos.

Por meio dos experimentos reais aplicados em bases públicas de vídeo se avaliou a precisão do método de correspondência de padrões e dos vetores de características representados em 64 e 128 dimensões. Onde o uso do descritor de pontos de interesses representados em 128 dimensões foi o mais preciso.

Realizou-se ainda uma avaliação teórica da comunicação entre câmeras. Mostrou-se que o uso de uma rede em malha com 54 Mbps e com tempo de espera no mínimo de 10 ms é considerado de custo barato e que o ambiente real pode ser formado por câmeras que possuam até 35 saltos entre a câmera mais distante e a central de vigilância.

Podemos concluir que os resultados obtidos nos experimentos são satisfatórios e que este trabalho apresentou maior eficiência do que os trabalhos relacionados e uma boa precisão na reidentificação.

As principais limitações deste trabalho consistem na falta de experimentos reais em rede de câmeras e o uso de câmeras constituídas de um processamento local.

6.1 Trabalhos Futuros

Apesar do método proposto auxiliar na tarefa de reidentificação de pessoas baseado no descritor dos pontos de interesse, outras evidências podem ser propostas, por meio da inclusão de novos atributos, como por exemplo, a reidentificação de pessoas através de pontos selecionados por regiões nas imagens.

Extensões do trabalho proposto podem ser a expansão para um buscador de vídeos de pessoas *online*, a reidentificação de modelos 3D ou até mesmo de outros objetos de interesse como automóveis e objetos perdidos.

A implantação do sistema no ambiente real assim como a inserção de uma estrutura topológica que representasse a localização das câmeras, uma atividade que deverá ser complementada no futuro.

6.2 Considerações Finais

Através dos resultados obtidos pelos grupos de experimentos tornou-se viável a reidentificação de uma pessoa em um ambiente monitorado por uma rede de câmeras utilizando uma característica inteligente formada por pontos de interesses. Esse ambiente consiste de câmeras com poder de processamento local e interligadas em uma rede sem fio em malha com um tamanho de banda de 54 Mbps.

A aplicação dessa abordagem em cenários reais que necessitem de vigilância são altamente recomendáveis, pois trata-se de uma abordagem eficiente e inovadora na área de segurança em grandes ambientes, como aeroportos, hospitais, shoppings, vias públicas, centros comerciais, estádios e também pode ser usado como um recurso auxiliar ao sistema de proteção e vigilância da Amazônia. Este trabalho agrega a funcionalidade de reidentificação que é desprovida pela maioria dos sistemas atuais.

O desenvolvimento deste trabalho é complementar ao projeto de pesquisa em Visão Computacional Distribuída sendo de grande valor a sua contribuição para as pesquisas atuais que podem vir a contribuir para o desenvolvimento científico e tecnológico regional.

Referências Bibliográficas

- Actions (2005). Actions as space-time shapes. [Online; acessado 10-Agosto-2008].
- Akyildiz, I. F., Wang, X., and Wang, W. (2005). Wireless mesh networks: a survey. *Computer Networks*, 47:445–487.
- Arth, C., Leistner, C., and Bischof, H. (2007). Object reacquisition and tracking in large-scale smart camera networks. In *ICDSC07*, pages 156–163.
- Bay, H., Tuytelaars, T., and Van Gool, L. (2006). Surf: Speeded up robust features. In *Proceedings of European Conference on Computer Vision*, pages 404–417.
- Beaudet, P. (1978). Rotationally invariant image operators. In *ICPR78*, pages 579–583.
- Black, J., Ellis, T., and Makris, D. (2004). Wide area surveillance with a multi camera network. *IEE Seminar Digests*, 2004(10426):21–25.
- Bradski, G. (2000). The OpenCV Library. *Dr. Dobb's Journal of Software Tools*.
- Brown, M. and Lowe, D. (2002). Invariant features from interest point groups. In *British Machine Vision Conference*, pages 656–665.

- Bruno Müller, J. and de Oliveira Anido, R. (2004). Distributed real-time soccer tracking. In *VSSN '04: Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*, pages 97–103, New York, NY, USA.
- Cantata (2000). Datasets for cantata project. [Online; acessado 10-Agosto-2008].
- Caviar (2004). Caviar test case scenarios. [Online; acessado 10-Agosto-2008].
- Collins, R., Lipton, A., and Kanade, T. (1999). A system for video surveillance and monitoring. In *American Nuclear Society 8th Internal Topical Meeting on Robotics and Remote Systems*.
- Collins, R., Lipton, A., Kanade, T., Fujiyoshi, H., Duggins, D., Tsin, Y., Tolliver, D., Enomoto, N., and Hasegawa, O. (2000). A system for video surveillance and monitoring. Technical Report CMU-RI-TR-00-12, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA.
- Evans, C. (2009). Notes on the opensurf library. Technical Report CSTR-09-001, University of Bristol.
- Forsyth, D. A. and Ponce, J. (2002). *Computer Vision: A Modern Approach*. Prentice Hall.
- Fu, K. S. (1982). *Syntactic Pattern Recognition and Applications*. Prentice-Hall, Englewoods Cliffs.
- Gheissari, N., Sebastian, T. B., and Hartley, R. (2006). Person reidentification using spatiotemporal appearance. In *2006 Conference on Computer Vision and Pattern Recognition (CVPR 2006)*, pages 1528–1535.
- Gonzalez, R. C. and Woods, R. E. (2001). *Digital Image Processing*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

- Hamdoun, O., Moutarde, F., Stanciulescu, B., and Steux, B. (2008). Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *ICDSC08*, pages 1–6.
- Hutchinson, S., Hager, G., and Corke, P. (1996). A tutorial on visual servo control. *RA*, 12(5):651–670.
- Lowe, D. (2003). Distinctive image features from scale-invariant keypoints. In *International Journal of Computer Vision*, volume 20, pages 91–110.
- Matsuyama, T. (1999). Cooperative distributed vision: Dynamic integration of visual perception, action, and communication. In *Mustererkennung 1999, 21. DAGM-Symposium*, pages 138–151, London, UK. Springer-Verlag.
- Michelsoni, C., Foresti, G., and Snidaro, L. (2005). A network of co-operative cameras for visual surveillance. *VISP*, 152(2):205–212.
- Mikolajczyk, K. and Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630.
- Morioka, K., Mao, X., and Hashimoto, H. (2006). Global color model based object matching in the multi-camera environment. *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 2644–2649.
- Nakazawa, A., Kato, H., and Inokuchi, S. (1998). Human tracking using distributed vision systems. *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on*, 1:593–596 vol.1.
- Park, U., Jain, A. K., Kitahara, I., Kogure, K., and Hagita, N. (2006). Vise: Visual search engine using multiple networked cameras. *Pattern Recognition, International Conference on*, 3:1204–1207.

- Pham, T. V., Worring, M., and Smeulders, A. W. M. (2007). A multi-camera visual surveillance system for tracking of reoccurrences of people. In *Distributed Smart Cameras, 2007. ICDSC '07. First ACM/IEEE International Conference on*, pages 164–169.
- Pio, J. L. d. S. (2006). Observação Dinâmica Cooperativa. Tese de Doutorado.
- Saito, H., Baba, S., Kimura, M., Vedula, S., and Kanade, T. (1999). Appearance-based virtual view generation of temporally-varying events from multi-camera images in the 3d room. In *Proceedings of Second International Conference on 3-D Digital Imaging and Modeling*, pages 516 – 525.
- Schalkoff, R. J. (1989). *Digital image processing and computer vision: an introduction to theory and implementations*. John Wiley & Sons, Inc., New York, NY, USA.
- Simon, D., Espiau, B., Castillo, E., and Kappalos, K. (1995). Computer-aided design of a generic robot controller handling reactivity and real-time control issues.
- Sivic, J. and Zisserman, A. (2003). Video google: a text retrieval approach to object matching in videos. *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1470–1477 vol.2.
- Stricker, M. and Swain, M. (1994). The capacity and the sensitivity of color histogram indexing. Technical report, Communications Technology Lab.
- Tanenbaum, A. S. (2002). *Computer Networks, Fourth Edition*. Prentice Hall PTR.
- Trucco and Verri, A. (1998). *Introductory Techniques for 3-D Computer Vision*. Prentice Hall.

- Tu, P. H., Doretto, G., Krahnstoever, N. O., Perera, A. A. G., Wheeler, F. W., Liu, X., Rittscher, J., Sebastian, T. B., Yu, T., and Harding, K. G. (2007). An intelligent video framework for homeland protection. In *Proceedings of SPIE Defence and Security Symposium - Unattended Ground, Sea, and Air Sensor Technologies and Applications IX*, volume 6562.
- Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features.
- Wei, G., Petrushin, V. A., and Gershman, A. V. (2004). Multiple-camera people localization in a cluttered environment. *Multimedia Data Mining Workshop*, pages 1–9.
- Witkin, A. P. (1983). Scale-space filtering. In *8th Int. Joint Conf. Artificial Intelligence*, volume 2, pages 1019–1022.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)