

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL  
FACULDADE DE INFORMÁTICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**

**RE-IDENTIFICAÇÃO DE  
PESSOAS EM IMAGENS  
ATRAVÉS DE  
CARACTERÍSTICAS  
DESCRITIVAS DE CORES E  
GRUPOS**

**NESTOR ZILLOTTO SALAMON**

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Ciência da Computação na Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Profa. Dra. Soraia R. Musse  
Co-Orientador: Prof. Dr. Julio C. S. Jacques Junior

**Porto Alegre  
2015**



### **Dados Internacionais de Catalogação na Publicação (CIP)**

S159r Salamon, Nestor Ziliotto

Re-identificação de pessoas em imagens através de características descritivas de cores e grupos / Nestor Ziliotto Salamon. – Porto Alegre, 2015.

85 p.

Dissertação (Mestrado) – Faculdade de Informática, PUCRS.

Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Soraia R. Musse.

Co-orientador: Prof. Dr. Julio C. S. Jacques Junior.

1. Informática. 2. Processamento de Imagens.  
3. Reconhecimento de Padrões. I. Musse, Soraia R. II. Jacques Junior, Julio C. S. III. Título.

CDD 006.61


**Ficha Catalográfica elaborada pelo  
Setor de Tratamento da Informação da BC-PUCRS**






## TERMO DE APRESENTAÇÃO DE DISSERTAÇÃO DE MESTRADO

Dissertação intitulada "Re-identificação de Pessoas em Imagens Através de Características Descritivas de Cores e Grupos" apresentada por Nestor Ziliotto Salamon como parte dos requisitos para obtenção do grau de Mestre em Ciência da Computação, aprovada em 13/03/2015 pela Comissão Examinadora:

  
\_\_\_\_\_  
Profa. Dra. Soraia Raupp Musse- PPGCC/PUCRS  
Orientadora

  
\_\_\_\_\_  
Prof. Dr. Júlio César Silveira Jacques Júnior- DOCFIX/PPGCC  
Coorientador

  
\_\_\_\_\_  
Profa. Dra. Isabel Harb Manssour- PPGCC/PUCRS

  
\_\_\_\_\_  
Prof. Dr. William Robson Schwartz- UFMG

Homologada em 23/04/2015, conforme Ata No. 008 pela Comissão Coordenadora.

  
\_\_\_\_\_  
Prof. Dr. Luiz Gustavo Leão Fernandes  
Coordenador.

**PUCRS**

**Campus Central**

Av. Ipiranga, 6681 - P32- sala 507 - CEP: 90619-900  
Fone: (51) 3320-3611 - Fax (51) 3320-3621  
E-mail: [ppgcc@pucrs.br](mailto:ppgcc@pucrs.br)  
[www.pucrs.br/facin/pos](http://www.pucrs.br/facin/pos)



# RE-IDENTIFICAÇÃO DE PESSOAS EM IMAGENS ATRAVÉS DE CARACTERÍSTICAS DESCRITIVAS DE CORES E GRUPOS

## RESUMO

Re-identificação de pessoas e grupos de pessoas em ambientes reais ainda é uma tarefa desafiadora: variações de luminosidade, ângulos de visão e resolução das imagens são, dentre outros, fatores que alavancam esta dificuldade. Este trabalho apresenta uma abordagem para re-identificação de pessoas focada em características *soft*-biométricas. O objetivo é reconhecer uma pessoa combinando descrições de baixo nível e alto nível (quando possível), tais como cores das roupas ou acessórios que carrega e informações de grupos em que se encontra, respectivamente. As características descritivas são manualmente informadas pelo usuário através de cores selecionadas (de um repositório de imagens, de uma imagem do suspeito ou mesmo de uma paleta de cores) e organizadas em um modelo de corpo 2D. Adicionalmente, o usuário pode especificar grupos de duas pessoas a serem buscadas explorando tal informação contextual de agrupamento. Cada modelo 2D é procurado em um banco de imagens usando medidas de distância de cores, segmentadas através de um limiar adaptativo. Por fim, e se aplicável, as assinaturas de cores de cada modelo 2D/pessoa são utilizadas para buscar formações de grupos com tais características.

**Palavras-Chave:** re-identificação de pessoas, análise de grupos, *soft*-biometria, recuperação de imagens.





# PEOPLE RE-IDENTIFICATION IN STILL IMAGES THROUGH COLOR AND GROUP BELONGING FEATURES

## ABSTRACT

The re-identification of people and groups in real environments is still a difficult task due to several changes in illuminance, viewpoints, image resolution and many other challenges. In this work we propose a person re-identification approach applied as a soft-biometric tool. The goal is to recognize a person by combining low-level and high-level (when possible) description of him/her, such as color appearance of clothes or objects he/she is carrying on (low-level) and group belonging (high-level). The input features for each person are manually informed by a user using sample patches from any source (a gallery repository, a picture taken or a color palette selection) and semantically organized in a 2D body model. In addition, groups of two persons (both defined as 2D body models) are specified. Finally, each 2D model is then confronted with databases using a color distance based metric, extended through an adaptive threshold and, if applicable, the color signatures of both persons into the group is used to search for a group composition with such characteristics.

**Keywords:** person re-identification, group analysis, soft-biometric, image retrieval.



## LISTA DE FIGURAS

- Figura 3.1 – Exemplo de imagens que podem ser descritas pelo usuário através das cores selecionadas para buscar e re-identificar os indivíduos nas demais cenas. O casal em (a) carrega sacolas vermelhas e possuem jaquetas predominantemente pretas, com saliências em amarelo e vermelho. Os indivíduos em (b) vestem casacos de cores branca e rosa predominantes, sendo que o segundo porta uma mochila preta. . . . . 43
- Figura 4.1 – Ilustração das etapas do modelo desenvolvido para re-identificação de pessoas e grupos. Na etapa manual (Inicialização), o usuário define a assinatura da pesquisa e o modelo automaticamente retorna, ao final das etapas, um *ranking* com os mais semelhantes indivíduos ou grupos. . . . . 45
- Figura 4.2 – Inicialização e seleção de cores. A seleção de cores na imagem (a) gera o modelo de corpo 2D para a pessoa buscada (*I*), ilustrado em (b). . . . 46
- Figura 4.3 – Detecção de pessoas e divisão do corpo em atributos em uma cena do banco ETHZ (a) e em um subconjunto do banco VIPeR (b) (redimensionados para efeitos de visualização). . . . . 48
- Figura 4.4 – O resultado da segmentação (c) para a cena apresentada em (a). A segmentação utilizou o modelo de cores selecionado em (b) e limiar  $Th_{km}^* = 3$ . 50
- Figura 4.5 – Uma visão geral da abordagem do limiar adaptativo. (a) imagem em análise sub-dividida em atributos; (b) mapa de distâncias  $\Delta E_{94}$  para o atributo *pernas* (parte inferior de (a)) - regiões escuras são as menores distâncias, computadas utilizando o modelo de cor  $T_{20}$  ilustrado na Figura 4.4(b); (c) seleção do limiar adaptativo (linha tracejada vertical); (d) resultado da segmentação utilizando a implementação original ([JJDJ+10]); (e) resultado do algoritmo *SLICO Superpixel* para o atributo *pernas*; (f) a célula com o menor limiar computado (em vermelho) e as células conectadas a ela (em verde); (g) resultado da segmentação com o limiar adaptativo modificado. . . 52
- Figura 4.6 – Ilustração dos erros (*S*) para algumas pessoas candidatas (*P*), computados em relação à pessoa buscada (*I*) ilustrada na Figura 4.2. . . . . 55
- Figura 4.7 – Detecção de grupos: os indivíduos detectados em uma cena do banco ETHZ (delimitados por seus *bounding-boxes* em laranja) e as relações de agrupamento entre si (linhas em vermelho). Pessoas não demarcadas por *bounding-box* não foram encontradas durante a etapa de detecção. 56

Figura 5.1 – Ilustração da re-identificação de um indivíduo utilizando o banco VIPeR. (a) imagem de entrada com as regiões selecionadas pelo usuário (câmera A) para geração do modelo de cores a ser buscado. (b-f) os 5 primeiros resultados - os menores erros na câmera B - com a associação correta na segunda posição do <i>ranking</i> (c). . . . .	60
Figura 5.2 – Curva CMC para o subconjunto de 316 imagens do banco VIPeR. . .	61
Figura 5.3 – Resultados ilustrativos da re-identificação no banco de imagens VIPeR. Na primeira coluna, a seleção feita pelo usuário é mostrada. Os destaques em amarelo denotam o <i>ranking</i> da correta re-identificação. . . . .	62
Figura 5.4 – Ilustração dos resultados para re-identificação de um indivíduo utilizando o banco ETHZ. (a) imagem de entrada com as regiões selecionadas pelo usuário (câmera A) para geração do modelo de cores a ser buscado. (b-d) os 3 primeiros resultados - os menores erros na câmera B - com a associação correta na segunda posição do <i>ranking</i> (c) (redimensionados para efeitos de visualização). . . . .	65
Figura 5.5 – Resultados ilustrativos da re-identificação de indivíduos no banco de imagens ETHZ. Na primeira coluna, a seleção feita pelo usuário é mostrada. Os destaques em amarelo mostram a correta re-identificação e sua respectiva posição no <i>ranking</i> . . . . .	66
Figura 5.6 – Ilustração de um resultado quando duas pessoas são buscadas individualmente e como um grupo. (a-b) ilustra as regiões selecionadas pelo usuário (câmera A), além da posição no <i>ranking</i> de cada pessoa quando buscadas individualmente contra o banco de imagens (213 indivíduos, câmera B). (c) ilustra o grupo re-identificado formado pelos mesmos indivíduos e sua posição no <i>ranking</i> de grupos (dentre os 141 grupos do banco, na câmera B). . . . .	67
Figura 5.7 – Resultados ilustrativos da re-identificação de grupos no banco de imagens ETHZ. Na duas primeiras colunas, as seleções feitas pelo usuário para cada indivíduo do grupo é mostrada. Os destaques em amarelo mostram a correta re-identificação e sua respectiva posição no <i>ranking</i> . . .	68
Figura A.1 – As <i>listas de equivalências</i> . A mulher (no <i>bounding-box</i> esquerdo) e o homem (no <i>bounding-box</i> direito) compartilham o mesmo ID (159 e 160, respectivamente) em três cenas/quadros (a-c). O casal também foi detectado como um grupo nestes 3 quadros (a-c), compartilhando do mesmo ID (93). . . . .	83

## LISTA DE TABELAS

- Tabela 5.1 – Resultados para o banco VIPeR: taxa de re-identificação cumulativa (em %) para melhores posições no *ranking* (mais semelhantes) dentre as 316 imagens/pessoas. As últimas duas linhas mostram a melhoria obtida com a modificação efetuada no limiar adaptativo em comparação com a sua forma original ([JJDJ+10]), conforme descrito na Subseção 4.3.2. A primeira linha mostra os resultados da abordagem estado da arte ([ZOW13a]). 60
- Tabela 5.2 – Comparação da taxa cumulativa de re-identificação (em %) no subconjunto do banco VIPeR com 316 imagens, utilizando quatro pares de espaços de cores/medidas de similaridade. A maior taxa foi obtida com a combinação *Lab* e  $\Delta E_{94}$ . . . . . 63
- Tabela 5.3 – Re-identificação de indivíduos no banco ETHZ: taxa de re-identificação cumulativa (em %) para melhores posições no *ranking* (mais semelhantes) dentre os 213 indivíduos. . . . . 65
- Tabela 5.4 – Resultados obtidos no subconjunto do ETHZ considerando as melhorias na identificação do grupo *versus* a re-identificação de indivíduos (melhores classificações cumulativas (em %) para os 141 grupos e 213 indivíduos). A primeira linha mostra os resultados para a re-identificação de grupos, enquanto a segunda linha sumariza a classificação média quando buscadas individualmente as pessoas de cada grupo (sem a informação contextual do grupo). . . . . 67



## LISTA DE SIGLAS

2D – Duas dimensões

3D – Três dimensões

CIE – *International Commission on Illumination*

CMC – *Cumulative Matching Characteristic*

FAST – *Features from Accelerated Segment Test*

FBI – *Federal Bureau of Investigation*

GLOH – *Gradient Location and Orientation Histogram*

HOG – *Histogram of Oriented Gradient*

HSV – *Hue, Saturation, Value*

LBP – *Local Binary Patterns*

NTSC – *National Television System Committee*

PAL – *Phase Alternating Line*

PCA – *Principal Component Analysis*

PLS – *Partial Least Squares*

RGB – *Red, Green, Blue*

SDALF – *Symmetry-Driven Accumulation of Local Features*

SIFT – *Scale Invariant Feature Transform*

SURF – *Speeded Up Robust Features*

SVM – *Support Vector Machines*





# SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>19</b>
1.1	QUESTÃO DE PESQUISA .....	22
1.2	OBJETIVO GERAL .....	23
1.3	OBJETIVOS ESPECÍFICOS .....	23
<b>2</b>	<b>REFERENCIAL TEÓRICO</b> .....	<b>25</b>
2.1	DETECÇÃO E DESCRIÇÃO DE CARACTERÍSTICAS .....	25
2.2	TÉCNICAS DE APRENDIZADO E CLASSIFICAÇÃO .....	28
2.3	SEGMENTAÇÃO DE IMAGENS COLORIDAS .....	31
2.4	ESPAÇOS DE CORES E MÉTRICAS DE DISTÂNCIA DE SIMILARIDADE .....	32
<b>3</b>	<b>TRABALHOS RELACIONADOS</b> .....	<b>37</b>
3.1	CONTEXTO DESTE TRABALHO NO ESTADO DA ARTE .....	42
<b>4</b>	<b>MODELO</b> .....	<b>45</b>
4.1	INICIALIZAÇÃO - SELEÇÃO DE CORES E CONSTRUÇÃO DO MODELO DE CORPO 2D .....	45
4.2	DETECÇÃO DE PESSOAS E DIVISÃO DO CORPO EM ATRIBUTOS .....	47
4.3	SEGMENTAÇÃO DE CORES COM LIMIAR ADAPTATIVO .....	48
4.3.1	SEGMENTAÇÃO PELA MÉTRICA DE DISTÂNCIA $\Delta E_{94}$ .....	49
4.3.2	LIMIAR ADAPTATIVO MODIFICADO .....	51
4.4	ANÁLISE E <i>RANKING</i> DE INDIVÍDUOS .....	53
4.5	DETECÇÃO, ANÁLISE E <i>RANKING</i> DE GRUPOS .....	55
<b>5</b>	<b>RESULTADOS OBTIDOS</b> .....	<b>59</b>
5.1	RE-IDENTIFICAÇÃO DE PESSOAS COM O BANCO VIPER .....	59
5.2	COMPARAÇÃO DE ESPAÇOS DE CORES NA RE-IDENTIFICAÇÃO DE PESSOAS COM O BANCO VIPER .....	63
5.3	RE-IDENTIFICAÇÃO DE PESSOAS E GRUPOS COM O BANCO ETHZ .....	63
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b> .....	<b>69</b>
	<b>REFERÊNCIAS</b> .....	<b>71</b>

	<b>APÊNDICE A – Processo de montagem dos bancos de imagens e seleção de entradas . . . . .</b>	<b>81</b>
A.1	DEFINIÇÃO DAS IMAGENS UTILIZADAS BANCO VIPER . . . . .	81
A.2	DEFINIÇÃO DO SUBCONJUNTO DE IMAGENS NO BANCO ETHZ . . . . .	81
A.3	CONSTRUÇÃO DAS LISTAS DE EQUIVALÊNCIAS . . . . .	82
A.4	MODELOS DE CORES SELECIONADOS PELO USUÁRIO . . . . .	83
	<b>APÊNDICE B – Lista de publicações obtidas e submetidas . . . . .</b>	<b>85</b>
B.1	ARTIGOS PUBLICADOS . . . . .	85
B.2	ARTIGOS SUBMETIDOS E SOB REVISÃO . . . . .	85

## 1. INTRODUÇÃO

A constante necessidade de investimentos em segurança, juntamente com a diminuição dos custos de equipamentos tecnológicos, tem alavancado o interesse em câmeras de segurança - privadas e públicas, abrindo diversas possibilidades para o avanço das pesquisas em visão computacional para análise das imagens nelas obtidas. No Brasil, em meados de 2014, a realização da Copa do Mundo de futebol trouxe significativos investimentos em segurança para as cidades-sede. A cidade de Porto Alegre, por exemplo, teve seu sistema de monitoramento expandido para 2,1 mil câmeras espalhadas por seu perímetro e região metropolitana<sup>1</sup>. Caso algum técnico observe ou seja informado da ocorrência de uma anormalidade, pode-se querer encontrar o indivíduo responsável dentre as imagens capturadas por uma ou mais destas câmeras. As técnicas de visão computacional podem ser extremamente úteis neste caso.

Uma abordagem aplicável é a identificação dos indivíduos que passam pelas câmeras, ou seja, a definição sem qualquer conhecimento prévio de um identificador único para cada pessoa encontrada. Esta abordagem permite ao observador do sistema de segurança em um estádio da copa, por exemplo, relacionar o identificador atribuído  $x$  ao torcedor de ingresso número 999. Outra abordagem é a re-identificação do suspeito, ou seja, a atribuição do mesmo identificador para todas instâncias do mesmo suspeito em diferentes câmeras ou cenas. A re-identificação é mais útil neste caso de anormalidade: baseado em alguma característica descritiva ou imagem prévia do suspeito, pode-se procurar onde ele se movimentou antes, durante e depois do acontecimento. Ambas tarefas são comumente executadas de forma manual - levando demasiado tempo e demandando alto custo para análise das imagens [BGS14]. Uma possível solução é a automatização das tarefas utilizando visão computacional.

As características do suspeito buscadas na re-identificação podem ser divididas em duas categorias: biométricas e *soft*-biométricas. Havendo uma foto ou vídeo do indivíduo, uma alternativa de re-identificação é a utilização de reconhecimento facial ou análise de padrões em sua locomoção. Tais características fisiológicas e comportamentais são classificadas como características biométricas e são de grande valia para encontrar a exata identidade de tal indivíduo, porém precisam de vídeos ou imagens em alta resolução para serem aplicadas. Caso não haja imagens anteriores - ou as imagens capturadas sejam de baixa qualidade para a análise biométrica - é possível fazer uso de características físicas que diferenciam os seres humanos entre si, tais como gênero, altura ou estilo do cabelo e ainda características descritivas temporárias como cores das roupas utilizadas e objetos portados. Neste segundo caso, a re-identificação do indivíduo é feita através de características *soft*-biométricas [WCC05].

---

<sup>1</sup><http://glo.bo/1ki7s5S>

Vezzani e sua equipe [VBC13] definem re-identificação como a tarefa de atribuir o mesmo identificador para todas as instâncias de um indivíduo detectado em uma série de imagens e vídeos, inclusive após uma lacuna significativa de espaço ou tempo. Segundo Bedagkar-Gala e Shah [BGS14], a re-identificação é indispensável para estabelecer a atribuição consistente de identificadores em múltiplas câmeras ou na mesma câmera para re-estabelecer trajetórias perdidas.

Esta re-identificação de pessoas em imagens vem possibilitando uma gama de aplicações nos últimos anos, principalmente no ramo de vigilância visual, motivando pesquisadores a buscar novas técnicas para endereçar o problema e automatizar tarefas de análise, hoje feitas, em grande parte dos casos, manualmente por técnicos de vigilância. Duas áreas bastante difundidas na literatura científica apresentam trabalhos que contribuem significativamente para a (re)identificação de pessoas: reconhecimento/detecção facial [VJ04, ZCPR03, HAMJ02] e detecção de pedestres [OPS+97, EG09, MG06].

Porém, quando se tratando de ambientes públicos, pode haver grande densidade de pessoas, e faces ou corpos provavelmente não aparecerão por completo. A pessoa buscada pode estar de costas para a câmera, a oclusão gerada pela grande quantidade de pessoas pode deixar visível somente parte do corpo ou até o próprio posicionamento da câmera pode fazer com que a re-identificação seja prejudicada.

Em 2013, um importante meio de comunicação de artigos científicos, o *IEEE Spectrum*, divulgou a notícia<sup>2</sup> que sistemas de identificação facial não foram um fator crucial na captura dos dois suspeitos do atentado ocorrido na maratona de Boston em abril daquele ano. O sistema de identificação facial não identificou os suspeitos mesmo que suas imagens estivessem cadastradas em bancos de imagens - ambos imigraram legalmente ao país. A identificação pela análise da face foi prejudicada pela qualidade das imagens capturadas pelas câmeras de vigilância (normalmente contendo baixa resolução, muitas vezes mal focalizadas, podendo inclusive capturar imagens por ângulos incomuns). O que parece ter sido um fator crucial nesse caso, que estimulou a obtenção de muitas dicas para se chegar aos suspeitos, foi a decisão do FBI (*Federal Bureau of Investigation*) de liberar publicamente fotos dos suspeitos, então não identificados, usando bonés preto e branco.

Por esse ponto de vista, pode-se considerar que características *soft*-biométricas associadas às pessoas (roupas, objetos que carregam, pose ou postura que assumem, etc.), que não propriamente suas características biométricas, bem como tais informações para pessoas ao seu redor, também podem ser utilizadas no processo de re-identificação. Segundo Jain e equipe [JDN04], características *soft*-biométricas são aquelas que provém alguma informação sobre o indivíduo, mas carecem de distintividade suficiente para diferenciar dois indivíduos. Apesar de não identificar unicamente como uma impressão digital (biométrica), as características *soft*-biométricas podem ser de grande valia na interpretação da informação e re-identificação do indivíduo.

---

<sup>2</sup><http://spectrum.ieee.org/riskfactor/computing/networks/face-recognition-failed-to-find-boston-bombers>

De acordo com Datta e sua equipe [DJLW08], a interpretação da informação visual é difícil de ser caracterizada. Ainda mais difícil é desenvolver estratégias para que uma máquina aprenda a reconhecer e interpretar automaticamente essa informação. Apesar disso, nos últimos anos, diversas tentativas foram realizadas para fazer com que sistemas computacionais conseguissem aprender, compreender, indexar e rotular imagens representando uma grande quantidade de conceitos semânticos [DJLW08, LZLM07]. Ainda de acordo com Datta [DJLW08], a recuperação automática de imagens por conteúdo é qualquer tecnologia que possa ajudar a organizar imagens digitais levando em consideração seu conteúdo visual. Por essa definição, sistemas que vão desde a atribuição de funções de medida de similaridade entre imagens até algoritmos robustos para rotulação de imagens estão inclusos neste contexto. Conforme Datta, a busca por conteúdo em imagens é dividida em três grandes categorias de pesquisa:

- busca por associação: onde não existe a ideia de uma imagem específica, mas conceitos genéricos como a cor ou textura, por exemplo, que são refinados ou modificados iterativamente;
- pesquisa objetiva: onde uma imagem específica é procurada; e
- pesquisa por categoria: onde uma única imagem representativa de uma classe semântica é requerida.

Neste trabalho visa-se re-identificar, por associação, pessoas em ambientes reais, dadas as suas características *soft*-biométricas e informações contextuais de pessoas agrupadas ao seu redor. Uma vítima de assalto relata, por exemplo, que o suspeito estava vestindo uma jaqueta preta e um boné branco e que, após o delito, este foi de encontro ao seu companheiro que usava uma camisa vermelha. É possível recuperar de uma imagem de cena cotidiana informações que possam ser classificadas como “boné branco”, “jaqueta preta” ou “camiseta vermelha”? Além de ser útil para a ciência forense ao procurar por um suspeito nas imagens das câmeras de segurança de vizinhanças ou em um banco de imagens de uma empresa, esta abordagem também é aplicável em casos com imagens de baixa resolução quando reconhecimento biométrico não é possível [VBC13].

Apesar de contornar o problema da oclusão facial, ainda se faz necessária, nesta abordagem *soft*-biométrica, a detecção de pessoas para estruturar seus atributos buscados (e eliminar regiões de cores semelhantes que não estejam relacionadas com os possíveis candidatos). Em ambientes com baixa densidade de pessoas, um detector de pessoas (por exemplo, baseado em HOG [DT05, DSS<sup>+</sup>13]) pode executar esta tarefa. Técnicas de subtração de fundo (abordadas em [SV14]) são amplamente utilizadas para detectar movimentos e também poderiam ser utilizadas; todavia, podem falhar com alta densidade de pessoas e não são aplicáveis a imagens estáticas. Segundo Mazzon e equipe [MTC12], quando cenários com alta densidade de pessoas são tratados, uma solução plausível para contornar o problema de oclusão é detectar a parte superior do corpo. O modelo cabeça-ombro

[WZM13], por exemplo, pode ser utilizado para extrair informações contextuais como roupas ou estilo de cabelo, sendo de grande valor para (re)identificação de pessoas, especialmente quando detecção facial por si só não provê informações suficientes [XACT11].

O problema de pesquisa deste trabalho consiste em investigar a utilização de características *soft*-biométricas na re-identificação de pessoas. Em outras palavras, o modelo desenvolvido não objetiva re-identificar automaticamente determinada pessoa, e sim reconhecê-la baseado em atributos manualmente extraídos que determinam, por exemplo, cores de roupas e objetos, além de informações sobre outros indivíduos que possam estar agrupados na cena. Esta abordagem de re-identificação também pode ser associada ao reconhecimento de indivíduos onde quer que a busca de determinado suspeito seja executada e todas as instâncias correspondentes sejam recebidas [VBC13]. Adicionalmente, a exploração das características de grupos, por exemplo, ao buscar “uma pessoa com camisa vermelha ao lado de outra com camisa azul”, pode ser de grande valia para tratar ambiguidades e casos onde ocorrem oclusões ou variações nas aparências e ângulos de visão [ZGX09]. Não obstante, dois pontos comuns na re-identificação de pessoas ainda são desafiadores neste trabalho: influência de inúmeros fatores do mundo real como sombras, ruídos de imagens, oclusões totais ou parciais e configurações de câmeras, além de fatores relacionados à própria natureza humana como variedade de aparências, posturas, roupas e forte semelhança entre diferentes pessoas [LHG12].

A apresentação deste trabalho está organizada de forma que: a questão de pesquisa é definida na Seção 1.1; as Seções 1.2 e 1.3 descrevem, respectivamente, os objetivos gerais e específicos deste trabalho; o Capítulo 2 discorre sobre as técnicas para detecção e aprendizado de características utilizados na detecção de objetos e pessoas, bem como abordagens para segmentação de cores que serviram de base teórica para construção do modelo; o Capítulo 3 analisa trabalhos relacionados na identificação e re-identificação de pessoas e detecção e análise de grupos. O restante deste trabalho foca na descrição da implementação e avaliação dos resultados como segue: o Capítulo 4 descreve o modelo desenvolvido neste trabalho para re-identificação de pessoas e grupos baseado em cores; o Capítulo 5 mostra os resultados obtidos utilizando o modelo implementado sobre os bancos de imagens VIPeR [GBT07] e ETHZ [ELS<sup>+</sup>08, ELVG07] para, respectivamente, re-identificação de pessoas e de grupos. Finalmente, o Capítulo 6 apresenta as considerações finais e sugestões para possíveis trabalhos futuros.

## 1.1 Questão de pesquisa

É possível recuperar de imagens estáticas características *soft*-biométricas que possam ser classificados como “boné branco”, “jaqueta preta” ou “camiseta vermelha”, para re-identificar os indivíduos que as portam, baseado em características descritivas de cores?

## 1.2 Objetivo Geral

O objetivo geral deste trabalho consiste em re-identificar pessoas em ambientes reais através de descrições de cores (baixo nível) que representam atributos como vestimentas e objetos portados e, se aplicável, definições de agrupamento ou proximidade (alto nível) com outras pessoas na cena.

Além dos desafios gerais para detecção de objetos/pessoas tais como iluminação ou resolução das imagens, outros aspectos a serem analisados envolvem definições de espaço e distâncias de cores, bem como tratativas para agrupamento de pessoas e oclusões. Faz-se, então, necessário um estudo aprofundado nas áreas de visão computacional, envolvendo métodos de detecção de características e de pessoas, teoria das cores, definições de atributos *soft*-biométricos e análise de grupos de pessoas em ambientes com até moderada densidade.

## 1.3 Objetivos Específicos

Para responder a questão de pesquisa e atingir o objetivo geral, os seguintes objetivos específicos foram definidos:

- pesquisar métodos para detecção de características e pessoas em imagens;
- revisar métodos para segmentação e comparação de cores;
- pesquisar e analisar abordagens já publicadas na área de re-identificação;
- desenvolver e avaliar um módulo para segmentação de cores;
- utilizar métodos de detecção de pessoas para definir os atributos *soft*-biométricos;
- desenvolver um módulo para detecção de grupos;
- integrar os módulos em um modelo capaz de re-identificar pessoas baseado em descrições de cores de atributos;
- definir um banco de imagens para testes;
- definir um banco de imagens de cenas cotidianas para detecção de pessoas e grupos;
- avaliar os resultados obtidos na re-identificação de indivíduos de forma qualitativa;
- avaliar os resultados obtidos na re-identificação de grupos; e
- escrever artigos com os resultados obtidos durante o desenvolvimento.





## 2. REFERENCIAL TEÓRICO

De acordo com Zhan e sua equipe [ZMR<sup>+</sup>08] e Jacques Junior e sua equipe [JJMJ10], há um considerável número de técnicas sendo utilizadas para detecção, contagem, rastreamento e identificação de pessoas, em ambientes com alta ou baixa densidade, que podem apresentar oclusões e agrupamentos das mesmas. Grande parte destes métodos foca na detecção do modelo de pessoa, no seu rastreamento a partir do modelo ou face ou, ainda, na identificação de poses, contorno de cabeça ou movimentos e comportamento para grupos. Uma análise dos trabalhos mais representativos focados em re-identificação de pessoas será apresentada no Capítulo 3. Antes disso, para melhor compreensão das técnicas utilizadas na área - da descrição de características para detecção de objetos e pessoas à segmentação de cores, um estudo teórico foi realizado e é apresentado neste capítulo. A revisão é baseada no trabalho de Jia e Zhang [JZ08] para detecção de humanos em imagens estáticas e estendida com métodos para detecção de objetos, finalizada por uma relação de métodos de segmentação e espaços de cores que podem ser utilizados para relacionar as características dos indivíduos buscados.

O conhecimento das diferentes etapas em cada abordagem foi substancial para consolidação dos trabalhos relacionados, além do foco e direcionamento do estudo das técnicas aplicáveis.

### 2.1 Detecção e Descrição de Características

A detecção de características (*features*) consiste em encontrar propriedades relevantes da imagem que possam servir ao interesse do usuário, sejam elas cantos, bordas, gradientes ou padrões. Estas características podem, ainda, ser representadas através de descritores que, por sua vez, nada mais são que identificadores representativos da características em questão. A literatura apresenta vários métodos que encontram e/ou descrevem conjuntos de características em imagens em escalas de cinza ou coloridas que podem levar à detecção de pessoas [PP00, VJ01].

O primeiro método a se sobressair na literatura para detecção de características foi proposto por Harris e Stephens em 1988 [HS88]. O *Harris* (ou *Plessey*) *Corner Detector*, detecta cantos e bordas com uma matriz de Harris (matriz de autocorrelação) dos valores da imagem, sobre a qual as variações são analisadas através de vetores próprios (*eigenvectors*, vetores que resumem as propriedades da matriz). Aplicações como explorações de imagens baseada em constantes geométricas [ZDFL95] ou geração de descritores para busca de imagens [SM97] utilizam este método. Porém, para aplicações mais complexas, o método precisa de ajustes, pois não trata variações na escala das imagens.

Encontrando essa limitação, Lindeberg [Lin98] propôs uma nova abordagem baseada em matrizes Hessian (para cálculo da assinatura do espaço-escala) e operadores Laplacianos (para diferenciais de escala), permitindo seleção automática de escala. Mikolajczyk e Schmid [MS02, MS04] propuseram abordagens para detecção de características inserindo elementos da geometria afim (origem/extensão/ângulo), estendendo com independência de escala os modelos de Harris e Lindeberg.

Em 1998, Papageorgiou e sua equipe [POP98] propuseram um método de detecção de características baseado em *Haar wavelets* - uma sequência de funções de “forma quadrada” que compõe uma família de ondas, baseando-se na diferença de intensidade em regiões da imagem que, por usarem somente o valor dos *pixels* da imagem, pode ser rapidamente computado. Diferentes padrões e tamanhos de *wavelets* caracterizam regiões distintas, maiores ou menores, verticais ou horizontais, etc. Um estudo mais completo do uso de *wavelets* foi feito pelos mesmos autores utilizando o modelo como base para detecção de faces, carros e pessoas [PP00]. Porém, o ápice da utilização da técnica foi quando Viola e Jones [VJ04] propuseram o algoritmo para detecção facial em tempo real amplamente difundido na literatura. Viola e Jones utilizaram *Haar-like features* (adaptadas das *Haar wavelets*) para detecção das regiões de interesse e o fazem em tempo real devido à utilização das suas Imagens Integrais - estruturas de dados aliadas a um algoritmo rápido e eficiente para gerar a soma de valores em subconjuntos de *pixels* em uma imagem [VJ01]. Lienhart e Maydt [LM02] estenderam as *Haar-like features* rotacionando as *wavelets*, permitindo detecção de características em diferentes ângulos.

No ano seguinte, Lowe [Low99] propôs um algoritmo de detecção de objetos independentes de escala, rotação e translação e parcialmente independentes de iluminação ou projeção 3D. Considerando somente a busca das regiões de interesse, uma procura sobre diferenças máximas e mínimas da função Gaussiana (*DoG*) aplicada em duas passadas sobre a imagem é executada removendo baixos contrastes, destacando tais regiões de interesse, que serão representadas através do descritor de características proposto: o *Scale Invariant Feature Transform* (SIFT). Uma extensão do SIFT para 3D foi proposta por Allaire e sua equipe [AKB<sup>+</sup>08], validando a aplicabilidade ao encontrar características em imagens de radioterapia. Mikolajczyk e equipe [MS05] também propõem uma extensão do SIFT, baseada em localização gradiente e histograma (GLOH), incrementando sua performance.

Inspirado no SIFT, o SURF (*Speeded Up Robust Features*) foi proposto por Bay e equipe em 2006 [BTVG06]. Independente de escalas e rotações no reconhecimento de objetos, o descritor representa as regiões de interesse detectadas através de matrizes *Hessian* (as curvaturas das funções) calculadas sobre Imagens Integrais. Usando uma aproximação de derivada de segunda ordem de Gauss para obter os valores da matriz *Hessian* em determinado ponto e escala, a determinante balanceada denota as características detectadas. Aplicações usando SURF são vistas na área médica para o registro de imagens [LZA11]

e no problema de detecção facial [DSH<sup>+</sup>09], sendo ainda estendidas para implementações em dispositivos móveis [YC12].

Análise de Componentes Principais (*Principal Component Analysis* - PCA) é um método estatístico que analisa uma tabela de dados e observa a correlação entre variáveis independentes. O principal objetivo do PCA é a redução de dimensionalidade [Jol05] - de descritores de características, por exemplo. Uma aplicação do método é a detecção de padrões em imagens, como no trabalho de Sirovich e Kirby [SK87], onde os autores introduziram o conceito de *eigenpictures* - coleção de imagens de faces que formam um conjunto suporte, refinado usando PCA. Turk e Pentland [TP91] decomposeram imagens de faces em características chamadas *eigenfaces* - conjuntos de *eigenvectors* - utilizando PCA para redução dimensional em seu reconhecedor facial. Desde então, PCA e suas adaptações são amplamente utilizados em aplicações que detectam características faciais [KJK02, YZFY04]. Le e Satoh [LS05] propuseram uma técnica simples e eficiente para seleção de características baseado em PCA, escolhendo-as nos eixos mais próximos ao PCA e diminuindo significativamente o tempo de computação. Ke e Sukthankar [KS04] propuseram o PCA-SIFT, incrementando os descritores do SIFT, usando PCA sobre gradientes normalizados e proporcionando otimizações para etapas seguintes - como a definição baseada nos descritores das mais prováveis localizações do objeto proposta por Zickler e Efros [ZE07].

Dalal e Triggs [DT05] propuseram um método para descrição de características baseado em histogramas de gradientes. Do inglês, *Histogram of Oriented Gradient* (HOG), o método divide a imagem em uma densa grade de células, cada célula contendo um histograma de gradientes orientados. Para cada *pixel* da célula, o vetor do gradiente é calculado e convertido para um ângulo que, ponderado pela magnitude do gradiente, influenciará na orientação da célula. Ainda, as células podem ser agrupadas em blocos para normalizar o contraste. HOG se mostrou eficiente na descrição de características para detecção de humanos [DTS06, SRBB06]. Posteriormente, Zhu e sua equipe [ZYCA06] inseriram o conceito de Integral de Gradientes ([Por05]) para tornar o cálculo do HOG mais eficiente. Schwartz e equipe [SKHD09] também propuseram uma abordagem para detecção de humanos utilizando HOG, adicionando informações de cores e texturas e reduzindo a dimensão de seus descritores com *Partial Least Squares* (PLS).

Rosten e Drummond [RD06] propuseram o FAST (*Features from Accelerated Segment Test*), um detector de cantos que utiliza um círculo como máscara de teste e é capaz de processar vídeos PAL (*Phase Alternating Line*) em tempo real. A implementação baseia-se na premissa de que deverá haver, dentro do círculo de teste,  $X$  *pixels* conectados cujos valores são mais claros ou mais escuros que um limiar determinado pelo *pixel* central. O algoritmo de aprendizado de máquina ID3 ([Qui86]) é, então, utilizado para definir quais comparações de *pixel* serão executadas, fator crítico na velocidade do método. Tal fato implica que nem todas as configurações de *pixel* serão testadas na detecção de cantos. Miar e

sua equipe [MHB<sup>+</sup>10] propõem um detector de cantos com a mesma base do FAST, fazendo uma busca mais genérica procurando pela árvore de decisão ótima em um maior espaço de configuração, inferindo a ordem de comparação dos *pixels* adicionando variáveis como “similar” e “não-mais-escuro”, além das “mais-escuro” e “mais-claro” já utilizadas. Rosten e sua equipe ainda propõem outras modificações sobre o FAST usando diferentes heurísticas e removendo constantes, permitindo que o método encontre cantos em vídeos PAL com menos de 5% de utilização do processamento [RPD10].

Padrões binários locais (*Local binary patterns* - LBP) são, segundo Ahonen e equipe [AHP06], um dos melhores descritores quando avaliadas texturas. O conceito de LBP foi proposto em 1996 por Ojala e equipe [OPH96], onde a implementação atribui rótulos para os *pixels* das imagens. Definido um *pixel* central, os *pixels* (3x3) da vizinhança são calculados em função do *pixel* central (limiar) e ponderados para que, quando agrupados, a soma seja o descritor da unidade de textura. Ojala e equipe estenderam os LBPs em 2002 [OPM02] para trabalhar com diferentes escalas de texturas e rotações. LBPs têm escopo bastante amplo em visão computacional, sendo utilizados no reconhecimento facial [AHP06], avaliação de expressões faciais [SGM09] e detecção de pedestres [WHY09].

Em se tratando da comparação de características, a utilização de somente seus descritores no cálculo das diferenças gera significativo ganho de tempo em buscas ou nas etapas de aprendizado e avaliação. Para comparativos de performance e eficiência de alguns detectores de características e descritores mais utilizados, recomenda-se a leitura dos trabalhos de Juan e Gwun [JG09] e Mikolajczyk e Schmid [MS05].

Conforme apresentado nesta seção, há várias maneiras na literatura para encontrar características candidatas a objetos de interesse em uma imagem. Estas características podem, então, ser aprendidas e classificadas para determinar se representam, por exemplo, uma pessoa, como de interesse neste trabalho.

## 2.2 Técnicas de Aprendizado e Classificação

Técnicas de aprendizado de máquina (*machine learning*) são métodos que aprendem dados e os utilizam para tomar decisões, ao invés de seguir um fluxo específico de um método enumerativo ou função. Estas técnicas estão diretamente ligadas à técnicas de classificação, ou seja, visam analisar dados e treinar classificadores para distinguir se um objeto pertence ou não a determinada classe.

Em um detector de cores simples, por exemplo, pode-se gerar um conjunto de imagens vermelhas e não vermelhas (exemplos positivos e exemplos negativos, respectivamente). O algoritmo de aprendizado será treinado com estas imagens para montar um conceito da cor vermelha. Posteriormente, é possível informar uma nova imagem e um

classificador buscará as informações aprendidas para retornar a probabilidade de, dado este treinamento, a entrada ser da cor vermelha. Este detector de cores é um exemplo de modelo discriminativo. Ao contrário dos modelos generativos que calculam a probabilidade conjunta baseado nos conhecimentos recém obtido dos dados não rotulados, os modelos discriminativos calculam diretamente a probabilidade condicional, ou seja, a probabilidade de  $Y$  ser vermelho dado que  $X$  é vermelho. Ng e Jordan [NJ01] fazem um estudo comparativo entre estes dois métodos.

Modelos discriminativo são inerentemente utilizados em técnicas de aprendizado supervisionado, onde dados rotulados são necessários para aprender o modelo - a definição manual dos exemplos positivos e negativos da cor vermelha exemplifica este aprendizado. Quando não há necessidade de dados rotulados para o aprendizado, o aprendizado é chamado não supervisionado - o PCA se encaixa também como uma técnica de aprendizado desta categoria, uma vez que reduz as dimensionalidades sem a necessidade de rotulação dos dados. Para reconhecimento de objetos e pessoas, baseado em imagens ou descritores, os modelos discriminativos se destacam na literatura [VJ01, DT05, WGDD12, BHW11].

Redes Neurais Artificiais (*Artificial Neural Networks*) são metodologias computacionais que executam análise multifatorial dos dados [DD01]. Em outras palavras, são modelos computacionais capazes de aprender e reconhecer padrões, cujo conceito foi cunhado baseado nas interligações dos neurônios cerebrais. A história remonta à criação do termo por McCulloch e Pitts [MP43], com um modelo matemático e algorítmico de redes neurais. Em 1949, Hebb [Heb49] criou a hipótese de aprendizado neural que mais tarde viria a ser simulada em computadores por Farley e Clark [FC54]. Uma rede neural é composta de elementos processadores interconectados que respondem paralelamente a um conjunto de entradas para elas dadas, usando observações para encontrar uma resposta ótima dentro da classe, sendo a optimalidade definida pela ponderação da função de custo pré-definida. Os diferentes modelos de Redes Neurais utilizam tanto aprendizado supervisionado quanto não supervisionado. Suas principais aplicações são encontradas na detecção de padrões numéricos [FMI83] e detecção facial [AJKA10, RBK98].

Máquinas de vetores suporte, do inglês *Support Vector Machines* (SVM), são modelos de algoritmos para aprendizado supervisionado utilizados na detecção de padrões e análise de dados. Dado um conjunto de imagens rotuladas entre duas categorias, o modelo aprenderá as características das instâncias de cada classe, montando o conceito que as separam. Proposto em 1995 por Cortes e Vapnik [CV95], o modelo constrói um espaço de decisão linear firmado em um mapeamento não linear dos vetores de entrada, definindo uma borda de separação baseada no *kernel* especificado e nos dados aprendidos. Em comparação a Redes Neurais, SVMs não minimizam o erro artificial mas maximizam o limite de decisão para melhor separar as classes. SVMs são bastante utilizados em sistemas de classificação de pedestres [MPP01, PP00] e, recentemente, Prosser e equipe [PZG<sup>+</sup>10] desenvolveram o *RankSVM*, uma abordagem para re-identificação de pessoas que evita

problemas de escalabilidade dos modelos tradicionais de *ranking* com SVM. Zhan e sua equipe [ZBMM06] propuseram um híbrido de SVM e Redes Neurais para trabalhar com múltiplas classes de problemas.

*Boosting* é um meta-algoritmo de aprendizado de máquina que visa reduzir a tendência do dado possuir características sobressalentes que deterioreem outras igualmente válidas. O conceito surgiu com uma afirmação para a possibilidade de um conjunto de algoritmos fracos de aprendizado poderem formar um consistente algoritmo de aprendizado [Sch90]. Várias implementações de algoritmos de *boosting* são encontradas na literatura, juntando diferentes algoritmos de aprendizado de máquina. Um dos mais utilizados na detecção de objetos é o *AdaBoost* [FS95]. Criado em 1995 por Freund e Schapire, o aprendizado do *AdaBoost* se torna consistente através da combinação linear ponderada de classificadores fracos. Da forma que utilizado no trabalho de Viola e Jones [VJ01] na detecção objetos, a cada ciclo um classificador é utilizado para cada característica e o classificador fraco com menor erro é chamado; a ponderação de seu valor incrementa os exemplos incorretos e decrementa o valor dos exemplos corretos, focando no aprendizado dos exemplos que possam ter escapado à classificação correta. Recentemente, Trzcinski e sua equipe [TCL13] propuseram o *BinBoost*, um conceito de descritor de características aprendido através de *boosting*.

As técnicas de classificação podem ser consideradas uma instância das técnicas de aprendizado supervisionado, uma vez que os dados rotulados do treinamento prévio são utilizados para definir a qual classe o objeto pertence. Um classificador baseado em SVM, por exemplo, divide em duas classes os exemplos treinados e define à qual a nova amostra pertence. O classificador *AdaBoost* é um meta-algoritmo adaptável que pode utilizar incrementavelmente diferentes tipos de aprendizado. Todavia, segundo Jia e Zhang [JZ08], este classificador não é eficiente para detecção de objetos quando utilizado um único e consistente treinamento: para o problema de detecção de pessoas, por exemplo, grande parte das características detectadas são exemplos negativos. Neste sentido, Viola e Jones [VJ01] introduziram a estrutura de classificadores em cascata.

O *Cascade AdaBoost* é uma estrutura de classificadores em cascata utilizando *AdaBoost*, sendo uma das estruturas mais utilizadas hoje em dia para detecção de objetos. Durante o treinamento, cada camada da cascata rejeita exemplos inválidos e somente os válidos são repassados para a próxima camada. Como vários exemplos falsos podem ser rejeitados com classificadores simples, é comum que a complexidade dos classificadores aumente a cada camada (e que a quantidade de exemplos diminua). Cabe ressaltar que, no modelo geral de cascata, os classificadores de cada camada podem ser baseados em diferentes métodos de aprendizado.

Outra estratégia para incrementar a classificação de exemplos é o *Bootstrapping*, aplicada no reconhecedor de faces de Sung e Poggio [SP98]. Nesta técnica, o conjunto de treinamento de cada fase de aprendizado é incrementado com exemplos de não interesse:

falso-positivos de classificadores já executados são coletados de um conjunto aleatório de padrões não válidos e outro classificador é alimentado com este conjunto. O aprendizado termina quando não mais houver ganho de performance entre os classificadores.

Exemplos que utilizam algumas das técnicas apresentadas de detecção de características, aprendizado e classificação para detectar objetos e pessoas já são bem difundidos na literatura: Dalal e Triggs usando *HOG-SVM/Bootstrapping* [DT05], Viola e Jones usando *Haar-Adaboost/Cascade* [VJ01] e Jia e Zhang usando *HOG-AdaBoost/Cascade* [JZ08].

Em uma abordagem *soft-biométrica*, as características detectadas e aprendidas podem ser relacionadas com, por exemplo, as cores da camiseta de um indivíduo a ser reidentificado. O particionamento da imagem de acordo com as cores de interesse é, então, a próxima área a ser aprofundada teoricamente.

## 2.3 Segmentação de Imagens Coloridas

Trabalhar com certas operações em imagens coloridas é um desafio que vem sendo explorado há décadas. Vários algoritmos de manipulação de imagens são executados sobre imagens em escala de cinza, dada a relativa facilidade de trabalhar com limitação de cores. Em outras palavras, separar as regiões com as cores de interesse pode facilitar o trabalho. Tal processo de separação pode ser efetuado através da segmentação de cores.

No contexto de operações em imagens, segmentação é uma das operações mais conhecidas. Segmentação de imagens consiste em particionar imagens por regiões disjuntas e homogêneas. É uma operação de baixo nível, porém executa papel de suma importância em aplicações como reconhecimento, interpretação e representação de imagens, haja vista a considerável diminuição no espaço de busca que pode ser proporcionada. Em 1994, Skarbek e sua equipe [SKV<sup>+</sup>94] já apresentavam uma classificação de técnicas para abordar o problema da segmentação de imagens coloridas: i) segmentação baseada em *pixels*, ii) baseada em área, iii) baseada em bordas e iv) baseada em física.

Na segmentação baseada em *pixels*, pode-se ter: i) técnicas baseadas em histogramas, onde picos são identificados e intervalos trabalhados; ii) segmentação por agrupamento de dados, onde os *pixels* são agrupados por representatividade para serem classificados; e iii) segmentação por agrupamento difuso, onde os agrupamentos são montados com funções difusas e, no processo inverso, agrupamentos concisos são feitos com regiões conectadas máximas.

Na segmentação baseada em área, duas outras categorias se sobressaem: i) crescimento de regiões, onde regiões iniciais são dadas e uma estratégia de ocupação

busca os vizinhos semelhantes; e ii) divisão e conquista, onde regiões não uniformes são divididas até que se tornem uniformes para serem agrupadas por funções heurísticas.

A segmentação por bordas tem sido amplamente utilizada em imagens em tons de cinza. Para imagens coloridas, abordagens globais e locais podem ser utilizadas, estas últimas envolvendo técnicas de otimizações para diferentes áreas de conhecimento. Seja global ou local, a segmentação por bordas é geralmente feita por gradientes, definindo uma medida que engloba a variação em todos os canais de cores ou calculando o gradiente para cada canal e combinando-as posteriormente.

A quarta abordagem é a segmentação baseada em Física. Talvez a abordagem mais complexa, esta consiste em segmentar a imagem pelo limite de objetos e materiais e não por influências de luzes e sombras (que afetam as cores). Esta abordagem previne erros que podem ocorrer nas demais técnicas pela mudança drástica na iluminação, por exemplo. Apesar do embasamento matemático dos modelos físicos ser semelhante aos dos demais modelos, estes podem se diferenciar na reflexão dos objetos (por exemplo, metal, plástico, madeira, etc.). Em suma, modelos físicos tentam distinguir variações nos materiais das variações de iluminação.

Técnicas dentro destas categorias ainda continuam sendo aperfeiçoadas e utilizadas até os dias atuais. Hoje em dia, ainda, o aumento da capacidade de processamento dos computadores vem tornando possível a implementação de outras técnicas que levam em conta a complexidade de imagens coloridas em diversos espaços de cores [CJSW01]. Vantaram e Saber [VS12] apresentaram recentemente uma nova avaliação do estado da arte em segmentação de imagens, incluindo técnicas como a utilização de Redes Neurais para encontrar padrões - permitindo levar em consideração informações espaciais, a utilização de SVMs para classificação de propriedades específicas das imagens, a segmentação através de descritores de imagens (HOG, LBP), dentre outras.

Para algoritmos representativos em cada classificação, refere-se, em adicional, os trabalhos de Cheng e equipe [CJSW01] e Lucchese e Mitray [LM01].

Não há um algoritmo que resolva todos os problemas de segmentação. Uma avaliação mais aprofundada dos dados a serem trabalhados pode requerer diferentes técnicas (de diferentes categorias) de segmentação. Outro ponto a ser levado em consideração é como a segmentação será realizada em relação à cor de referência, ou seja, como a medida de distância/similaridade entre as cores será calculada.

## **2.4 Espaços de cores e métricas de distância de similaridade**

Ao trabalhar com imagens coloridas, um fator de suma importância é o espaço de cores: o modelo matemático abstrato que formaliza a descrição das cores. Para ima-



gens obtidas por câmeras digitais, o espaço de cores mais comum é o *RGB* (*Red, Green, Blue*). Outro modelo primário é o *XYZ*, onde *Y* é iluminação/brilho, e *Z* e *X* formam um plano contendo todas as possibilidades de cromaticidade dado *Y*. Há modelos baseados na percepção humana de cores, como o *HSV* (*Hue, Saturation, Value*) - obtido através de transformadas do *RGB* - e o *Lab* (*Lightness, a* para verde-vermelho e *b* para azul-amarelo) - obtido através do espaço de cores *XYZ*. *YIQ* e *YUV* são modelos de cores utilizados em televisores americanos (*NTSC*) e europeus (*PAL*), respectivamente, também transformados a partir do *RGB*. Algoritmos de manipulação de imagens podem funcionar somente em determinados espaço de cores e justificam as conversões entre os espaços. Uma abordagem mais detalhada sobre estes e outros espaços de cores é apresentada no trabalho de Tkalcic e Tasic [TT03].

Para comparar cores dentro do mesmo espaço de cores, esbarra-se na extensa amplitude do conceito de cor semelhante. Uma cor *RGB* (180, 2, 27), por exemplo, é uma tonalidade de vermelho escuro. Computacionalmente, pode-se imaginar que ao modificar somente as componentes *GB*, a cor será um “vermelho semelhante”. Mas não é o que de fato ocorre: o *RGB* (180, 150, 80) é uma tonalidade de bege. Esta característica é oriunda da percepção humana das cores: uma resposta do sistema receptivo de cada indivíduo, do olho e do cérebro, a um estímulo de cor - a reflexão ou transmissão da fonte de luz por determinado material [BVM08]. Ou seja, a interpretação humana não é uma simples diferença de componentes para definir a semelhança entre duas cores, sendo necessárias outras definições de medidas para distância de similaridade entre as cores.

A vasta quantidade de espaços de cores definidos, além das diversas medidas de distâncias elaboradas, permite que operações envolvendo cores apresentem diferentes resultados conforme a combinação utilizada. Os quatro espaços de cores mais utilizados no problema de re-identificação - *RGB*, *HSV*, *HS* e *Lab* - podem ter a medida de similaridade entre suas cores calculada com diferentes métricas de distâncias.

Nas três componentes do *HSV* - matiz, saturação e valor - a diferença entre cores pode ser obtida através da distância Euclidiana de três dimensões. De acordo com Fisher [Fis99], a partir de um valor *RGB*, para levar em conta a instabilidade da cor quando convertida para *HSV*, a cor resultante deve ser avaliada no hexacone *HSV*. Logo, a distância entre duas cores  $h = (vs \cos(2\pi h), vs \sin(2\pi h), v)$  e  $h' = (v's' \cos(2\pi h'), v's' \sin(2\pi h'), v')$  pode ser calculada através da Equação 2.1. Para o espaço minimizado, com somente as componentes *HS*, na mesma representação do hexacone *HSV*, a diferença entre duas cores  $hs = (s \cos(2\pi h), s \sin(2\pi h))$  e  $hs' = (s' \cos(2\pi h'), s' \sin(2\pi h'))$  é obtida pela Equação 2.2.

$$D(h, h') = \sqrt{(vs \cos(2\pi h) - v's' \cos(2\pi h'))^2 + (vs \sin(2\pi h) - v's' \sin(2\pi h'))^2 + (v - v')^2}. \quad (2.1)$$

$$D(hs, hs') = \sqrt{(s \cos(2\pi h) - s' \cos(2\pi h'))^2 + (s \sin(2\pi h) - s' \sin(2\pi h'))^2}. \quad (2.2)$$

Para o espaço de cores *RGB*, uma métrica que calcula a diferença entre duas cores é a distância de Mahalanobis [Mah36]. Sendo  $\mu$  o vetor médio da referência e  $C$  a matriz de covariância, a distância para a cor  $c$  é dada pela Equação 2.3.

$$D(c) = \sqrt{(c - \mu)^T C^{-1} (c - \mu)}. \quad (2.3)$$

Para cálculo da diferença de cores no espaço *Lab*, a medida  $\Delta E_{94}$  é bastante referenciada na literatura por levar em consideração a percepção humana dentro do *Lab* [Goo12]. Dado duas cores *Lab*  $L_1 = (L_1, a_1, b_1)$  e  $L_2 = (L_2, a_2, b_2)$ , a distância  $\Delta E_{94}$  é calculada por

$$\Delta E_{94}^*(L_1, L_2) = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H S_H}\right)^2}, \quad (2.4)$$

onde

$$\begin{aligned} \Delta L^* &= L_1^* - L_2^*, \\ \Delta C_{ab}^* &= C_1^* - C_2^*, \\ \Delta H_{ab}^* &= \sqrt{\Delta a^{*2} + \Delta b^{*2} - \Delta C_{ab}^{*2}}, \end{aligned} \quad (2.5)$$

$$C_1^* = \sqrt{a_1^{*2} + b_1^{*2}}, \quad C_2^* = \sqrt{a_2^{*2} + b_2^{*2}}, \quad (2.6)$$

$$\Delta a^* = a_1^{*2} - a_2^{*2}, \quad \Delta b^* = b_1^{*2} - b_2^{*2}, \quad (2.7)$$

$S_L, S_C, S_H$  são parâmetros de ponderação que ajustam as diferenças do CIE em relação à localização do padrão CIE1976 (sendo  $S_L = 1, S_C = 1 + K_1 C_1^*, S_H = 1 + K_2 C_1^*$ ), e  $k_L, k_C$  e  $k_H$  são parâmetros específicos para a aplicação.

A variedade de espaço de cores e métricas para avaliação de distância entre estas é bastante explorada na literatura, tanto computacionalmente ([GH92], [Fis99]) quanto perceptualmente ([ITM01]). Du e equipe [DAL12] apresentam uma avaliação de diferentes espaços de cores - quando analisados como descritores - no problema de re-identificação, onde o *RGB* normalizado tem melhor taxa de acertos entre os espaços não combinados. Porém, esta não é uma conclusão definitiva aplicável a todas soluções e abordagens. A escolha do espaço e métrica mais adequados ainda são dependentes da aplicação e necessitam ser avaliadas pontualmente.

A seguir será apresentada uma compilação de trabalhos relacionados que utilizam, dentre outras, as técnicas e conceitos até aqui vistos, em abordagens para re-identificação de pessoas. Complementar a estes, serão buscadas, neste trabalho, características de adereços e vestimentas que possam induzir à pessoa de interesse.



### 3. TRABALHOS RELACIONADOS

Uma tarefa fundamental para um sistema de vigilância multi-câmera distribuído é associar pessoas entre câmeras com diferentes ângulos de visão e em diferentes posições e tempos. Assim Gong e equipe [GCLH14] definem o problema de re-identificação. Deixar o monitoramento (e associação) ser feito por humanos pode ser errôneo, custoso e demorado [BGS14]. Este capítulo elenca e analisa alguns trabalhos já publicados na literatura que automatizam o problema de re-identificação de pessoas e utilizam, dentre outras, as técnicas de detecção de características e rotulação/aprendizado até aqui vistas. Também são relacionadas técnicas aplicadas a grupos de pessoas neste problema.

Segundo Bedagkar-Gala e Shah [BGS14] em sua recente análise de abordagens e tendências para o problema de re-identificação, esta tarefa ainda é difícil de ser executada automaticamente sem intervenção humana, uma vez que computadores precisam extrair e re-identificar descritores (como face, roupas, altura) dos indivíduos, algo que os seres humanos podem fazer com naturalidade no dia a dia. Não obstante, como mencionado no trabalho de Mazzon e equipe [MTC12], câmeras disjuntas (tal como dispostas câmeras de segurança por uma cidade) tornam a re-identificação de pessoas um problema ainda mais desafiador, já que mudanças na pose, escala e iluminação modificam a aparência das pessoas. Vezzani e sua equipe [VBC13] apresentaram recentemente uma revisão do estado da arte em re-identificação de pessoas analisando abordagens que visam contornar estes desafios. No mesmo trabalho [VBC13], os autores propuseram uma taxonomia multidimensional para classificar as diferentes abordagens em re-identificação de pessoas, levando em conta a configuração de câmera (calibradas, sobrepostas, disjuntas, etc.), o conjunto de amostra (múltiplas ou única imagem - *multi-shot* ou *single-shot*), a assinatura (texturas, cores, formas, etc.), modelo de corpo (2D, 3D, sem modelo), aprendizado de máquina (*i.e.* distâncias, transformação de cores) e cenário de aplicação (rastreamento, recuperação de imagem, etc.). Ainda, segundo Cheng e equipe [CCS<sup>+</sup>11], quando levado em consideração o somente o aprendizado de máquina, as abordagens podem ser sub-classificadas como aplicações baseadas em aprendizado ou aplicações diretas - quando um conjunto de dados é utilizado para treinamento ou quando extraem diretamente as características das imagens, respectivamente.

Haja vista as dificuldades na automatização da re-identificação de pessoas ([BGS14, MTC12]), alinhadas com as diferentes possibilidades e categorias para endereçar o problema ([VBC13, CCS<sup>+</sup>11]), propostas de solução bastante abrangentes são encontradas na literatura.

O cerne dos métodos de re-identificação de pessoas através de características *soft*-biométricas consiste na aparência dos indivíduos. Schwartz e Davis [SD09] propuseram uma abordagem discriminativa baseada em aprendizado de aparências na qual a

assinatura de cada indivíduo é composta por diversos descritores de características: matrizes de co-ocorrência são responsáveis pela descrição de texturas, HOG ([DT05]) captura bordas e gradientes e *rankings* de intensidade para cada canal *RGB* descrevem as cores. Tais descritores são aprendidos através de exemplos positivos para a pessoa buscada (*multi-shot*) alinhados com contra exemplos (as pessoas restantes na base de entrada) e dimensionalmente reduzidos utilizando PLS para possibilitar a classificação. Hirzer e sua equipe [HBRB11] utilizam uma combinação de abordagens descritivas e discriminativas para re-identificar pessoas em câmeras disjuntas. Através da similaridade das características descritivas - aprendidas em diferentes quadros rotulados e representadas por descritores de região de covariância, um *ranking* inicial de re-identificação é estabelecido. Se a correspondência correta não estiver em uma posição alta no *ranking*, o algoritmo gerará uma classificação discriminativa utilizando *Haar features* e características de covariância sobre cores, aprendidas sem rotulação (sobre os exemplos já refinados) usando técnicas de *boosting* ([VJ01]).

Abordagens de re-identificação baseadas em aprendizado e classificação de características são encontradas com facilidade na literatura. Zhou e equipe [ZQJ+14] propuseram uma abordagem na qual a re-identificação é resultante de um *ranking* não-linear com vetores de diferenças, construído sobre um classificador binário com SVM ([CV95]) através da diferença do histograma *HSV* entre os candidatos. Zheng e equipe [ZGX13] apresentam uma abordagem cujas distâncias relativas maximizam a probabilidade de pares corretos - com distâncias pequenas - através do aprendizado das melhores distâncias de similaridade. Zhao e equipe [ZOW13a] relatam que a maioria dos métodos existentes faz a correspondência de imagens de pedestres através da comparação direta de características não alinhadas, oriundas da variação dos ângulos de visão e mudanças de poses, além de remover atributos como uma mochila ou um boné por não serem considerados partes do corpo. Neste mesmo trabalho, o problema da re-identificação de pessoas foi formulado como uma correspondência de saliências, em roupas ou acessórios, aprendidas supervisionadamente e ponderadas de acordo com suas capacidades discriminativas. Ao explorar a distribuição pareada das saliências entre imagens de pedestres em uma estrutura unificada de aprendizado com *RankSVM*, a abordagem tornou-se o estado da arte na re-identificação *single-shot*.

Zhao e equipe também propuseram, em um segundo trabalho [ZOW13b], um modelo para re-identificação de pessoas baseado no aprendizado não supervisionado de saliências, onde as características distintivas são extraídas sem necessitar rótulos no treinamento. A abordagem foi motivada pela constatação de que grande parte dos trabalhos nesta área utiliza de modelos de aprendizado supervisionado, o que requer dados rotulados para treinamento. Dados estes que em, por exemplo, uma mudança de câmera, precisam ser rotulados e gerados novamente, tornando impraticáveis aplicações em larga escala com diferentes câmeras. Schwartz [Sch12] relata que a reconstrução dos dados de treinamento

quando imagens são alteradas/adicionadas pode ser feito com somente uma parte dos dados, porém não se evita a reconstrução.

A abordagem de Farenzena e equipe [FBP<sup>+</sup>10] considera a aparência da pessoa a ser re-identificada através da análise direta de três atributos: i) cromaticidade global da imagem, ii) distribuição espacial das cores em regiões estáveis e iii) presença de recorrentes temas com alta entropia. A abordagem faz a separação dos *pixels* de fundo e de primeiro plano para obter a silhueta da pessoa em análise (na abordagem *single-shot*, as imagens são divididas com máscaras de silhueta, enquanto na *multi-shot* técnicas de subtração de fundo são utilizadas) e extrair as características da pessoa em primeiro plano dividindo simetricamente a silhueta obtida. Para cada parte dessa divisão, descritores de cores (*HSV*), *Maximally Stable Color Regions* e *Recurrent High-Structured Patches* - representando, respectivamente, os três atributos analisados - são atribuídos e comparados de acordo com o conjunto de amostra (*single-shot* ou *multi-shot*) utilizado.

Haja vista a grande quantidade de descritores de características e diferentes análises utilizadas (correspondência direta ou aprendida), alguns autores trabalham com ponderação relativas para diferentes atributos/características. Liu e equipe [LGL14] propuseram uma abordagem que aprende adaptativamente a ponderar descritores de cores *HSV*, *RGB*, *YCbCr* ou filtros *Gabor* e *Schmid*. A abordagem pode ser complementar a aplicações já disponibilizadas, porém depende da quantidade e qualidade dos exemplos não rotulados utilizados no aprendizado das características. Li e equipe [LZW13] basarem sua abordagem na ideia de que diferentes métricas visuais devem melhor ser aprendidas quando provenientes de diferentes conjuntos de candidatos. Em um conjunto de busca inicial, subconjuntos para cada candidato são selecionados através dos seus vizinhos mais próximos. Sobre o conjunto de busca resultante (filtrado), para cada indivíduo, as métricas de distância são aprendidas e ponderadas. Dado uma imagem a ser buscada, a comparação é feita através das métricas otimizadas para cada candidato e seus vizinhos mais próximos. As análises são feitas através de descritores de cores HOG ([DT05]), LBP ([OPH96]), SIFT ([Low99]) e filtros *Gabor*.

O problema da variação de iluminação oriunda dos diferentes ângulos de visão foi tratado por Li e Wang [LW13]. Os autores propuseram um modelo para trabalhar com câmeras de diferentes ângulos de visão através da similaridade de métricas transformadas para diferentes visualizações. A abordagem utiliza aprendizado supervisionado para, dado um par de imagens de diferentes ângulos de visão a ser comparado (todos contra todos), alinhá-los através do projeção para um espaço de características comuns e então combiná-los com métricas otimizadas localmente, baseadas nos descritores LBP ([OPH96]), histogramas *HSV* e filtros *Gabor*.

A quantidade de possíveis candidatos na cena também é um fator a ser considerado. Para lidar com situações de multidões densas, Mazzon e equipe [MTC12] propõem um modelo para re-identificação de pessoas que leva em conta a aparência, a localiza-

ção espacial das câmeras e os potenciais caminhos que o indivíduo pode percorrer. O trabalho extrai características de aparência de um modelo de representação definido como uma faixa vertical ao redor da cabeça do indivíduo, estimada usando um detector de cabeças ([EG09]). Dutra e equipe [DSS<sup>+</sup>13] utilizam esquemas de indexação baseados em listas invertidas para reduzir o número de possíveis candidatos: em um primeiro estágio, os candidatos têm suas imagens divididas em blocos - cada qual com seus descritores HOG ([DT05]) e predominância *RGB* - e um dicionário é montado com descritores randomicamente selecionados, chamados *codewords*; na etapa de aprendizado, uma lista invertida (que permite mapear os descritores extraídos aos seus identificadores/indivíduos) é criada para cada bloco extraído; os descritores do bloco são confrontados com as *codewords*, populando a lista invertida com os identificadores que geraram as *codewords* mais próximas. Por fim, para buscar um indivíduo, Dutra e equipe dividem a imagem de entrada em blocos, comparando seus descritores às *codewords* que retornarão, da lista invertida, os indicadores dos candidatos mais prováveis, dentre os quais uma média de covariância *Riemannian* ([PFA06]) realizará a re-identificação.

Utilizar conhecimento humano para especificar parâmetros e atributos discriminativos é outra abordagem já explorada na literatura. Layne e equipe [LHG12] propuseram um modelo para re-identificação que aprende uma seleção e ponderação de atributos semânticos (tais como estilo do cabelo, tipo de sapato e roupa) para descrever uma pessoa, inspirado nos procedimentos usados por profissionais no ramo de vigilância. Neste caso, um desafio é a acuracidade na detecção dos atributos relacionados às partes inferiores do corpo quando em multidões, onde as pessoas estão oclusas de diversas maneiras. Por outro lado, como mencionado pelos próprios autores, a combinação e ponderação dos atributos pode prover pistas discriminativas significantes para identificação, além de complementar representações de características de baixo nível. A abordagem de Cheng e equipe [CCS<sup>+</sup>11] toma por base como seres humanos fazem a re-identificação: usuários foram submetidos a testes para ligar duas imagens de indivíduos, enquanto monitorado em quais atributos dispndiam maior atenção realizando a correspondência. O estudo demonstrou que a comparação era feita parte a parte, e não do indivíduo por inteiro. Cheng e equipe então definiram as partes correspondidas através de *Pictorial Structures* - estruturas que representam o corpo dos indivíduos em uma configuração deformável, capturando aparência local de cada parte do corpo. Para cada parte, as cromaticidades (histogramas *HSV*) e distribuição espacial das cores foram utilizadas para criar as assinaturas dos indivíduos. Por fim, as assinaturas de cada parte para as pessoas buscadas são confrontadas com todas as imagens do conjunto de busca. Para trabalhar com abordagens onde há mais de uma imagem para cada indivíduo a ser buscado (*multi-shot*), Cheng e equipe propuseram as *Custom Pictorial Structures*, responsáveis pelo aprendizado não supervisionado dos atributos para melhorar a re-identificação de cada parte.



Diferente dos trabalhos de Cheng e equipe [CCS<sup>+</sup>11] e Layne e equipe [LHG12], a divisão do corpo em atributos de forma empírica, sem análise humana, também é explorada na literatura, variando de acordo com cada abordagem. Bak e equipe [BCBT10] utilizam uma modificação do detector de pessoas baseado em HOG ([DT05]) para dividir em 15 regiões (distribuídas pela silhueta aprendida) o corpo da pessoa encontrada. Em tal abordagem discriminativa, as regiões são agrupadas em 5 partes do corpo (cabeça, tronco, pernas e braços direito e esquerdo) e um descritor da covariância de cada parte é utilizado para medir a similaridade entre as cores normalizadas (e seus gradientes). As partes detectadas do corpo, somadas ao corpo por inteiro, são avaliadas diretamente por suas dissimilaridade em uma estrutura de pirâmide - do corpo inteiro aos atributos. Wei e equipe [WMZ<sup>+</sup>14] detectam os possíveis candidatos e dividem o corpo em 8 regiões - como cabeça, braço direito, braço esquerdo, etc. - atribuindo um descritor SIFT ([Low99]) para cada região, porém sem utilizar cores.

Ressalta-se que, independente da abordagem, quando baseada em descritores obtidos automaticamente, ao menos uma imagem do suspeito é necessária para realizar a extração ou aprendizado das assinaturas de busca.

Quando se tratando de ambientes públicos, segundo Zheng e equipe [ZGX14], pessoas comumente andam em grupos, seja com pessoas que conhecem ou entre desconhecidos. Associar as pessoas através dos grupos que elas pertencem pode trazer duas vantagens: i) a associação de grupos após de grande espaço ou tempo pode ser extremamente útil para entender e inferir associações de longo termo e o comportamento holístico do grupo no espaço público e ii) pode prover contexto visual que auxilia vitalmente na associação de indivíduos uma vez que a aparência de um indivíduo sozinho frequentemente sofre alterações drásticas causadas por mudanças de ângulos de visão ou iluminação. Neste segundo caso, entende-se que o contexto do grupo pode ajudar a re-identificação dos indivíduos.

Apesar da detecção, contagem e análise comportamental de agrupamentos de pessoas já terem sido amplamente estudadas na literatura ([AZ08, KGT05, JJMJ10]), o trabalho de Zheng e equipe [ZGX09] foi o pioneiro na utilização de grupos de pessoas no contexto de re-identificação, usando imagens capturadas de múltiplas câmeras não sobrepostas. Uma desvantagem da abordagem é a utilização de algoritmos de subtração de fundo para remover *pixels* indesejados ao fundo, o que não é aplicável em imagens estáticas/únicas. Recentemente, os mesmos autores propuseram um complemento [ZGX14] do trabalho pioneiro que foca na influência dos grupos comparado com re-identificação individual. Uma combinação das métricas *Center Rectangular Ring Ratio-Occurrence* e *Block-Based Ratio-Occurrence* sobre os descritores (SIFT [Low99] com *RGB*) de grupos e indivíduos mostra a melhoria na re-identificação de pessoas quando combinadas com o contexto de grupos. Porém, para apresentar uma classificação dos resultados mais semelhantes, a abordagem requer de técnicas de aprendizado das características mais discrimi-

nativas, ponderadas utilizando *RankSVM*. Em ambos os trabalhos, como mencionado pelos autores, o foco consiste na avaliação dos descritores de grupos propostos e uma detecção automática de grupos se faz necessária na prática.

### 3.1 Contexto deste trabalho no estado da arte

A abordagem proposta neste trabalho visa explorar alguns conceitos dos trabalhos relacionados neste capítulo e técnicas estudadas no Capítulo 2, além de incluir novas estratégias, com o objetivo de re-identificar pessoas e grupos de pessoas através de um modelo semi-automático, em uma abordagem independente de aprendizado de características ou informação temporal. Deverá ser um diferencial a possibilidade de re-identificação através de, por exemplo, uma descrição do “suspeito de ‘boné branco’ e ‘jaqueta preta’ que andava junto do indivíduo de ‘camisa vermelha’ ”, de acordo com a necessidade do usuário que definirá a relação das cores e atributos buscados. Inclusa nesta descrição do indivíduo, ressalta-se a contribuição na utilização da informação contextual de grupos em imagem estáticas. Esta abordagem se encaixa na taxonomia de Vezzani e equipe [VBC13] da seguinte maneira:

- Configuração de câmera: utilizando *câmeras disjuntas não calibradas*, uma vez que o objetivo é encontrar uma pessoa em determinado banco de imagens dada sua respectiva assinatura de cores (em caso de aplicação em bancos com câmeras calibradas, tal informação não é utilizada).
- Conjunto de amostra: utilizando *única imagem*, já que informação temporal não é utilizada (sem subtração de fundo) e, principalmente, porque a entrada de dados pode ser montada a partir de uma imagem qualquer contendo a cor de interesse (não é obrigatória imagem anterior do suspeito para que sua assinatura seja definida).
- Assinatura: *cor*. A característica principal usada pra medir a similaridade entre o modelo de entrada e a pessoa candidata é a informação de cor (que pode ser acrescida da associação de grupo);
- Modelo de corpo/pessoa: *modelo de corpo 2D*. As características de entrada (assinatura de cores) são semanticamente organizadas em um modelo de corpo 2D para construir uma coerente representação dos atributos (vestimentas e objetos) da pessoa.
- Aprendizado de máquina: *distance metric-based*. Apesar de não utilizar o aprendizado propriamente dito, o *ranking* dos resultados é feito através do modelo de segmentação de cores baseado em distância de similaridade utilizando um limiar calculado automaticamente.

- Cenário de aplicação: *recuperação de imagem*, uma vez que todas imagens correspondentes à assinatura buscada deverão ser classificadas ordenadamente (*ranking*).

Ainda, segundo a sub-classificação de Cheng [CCS<sup>+</sup>11] quanto ao aprendizado, esta abordagem é classificada como direta, uma vez que as características dos indivíduos são definidas pelo usuário e extraídas sem treinamentos para comparação.

Diferentemente dos trabalhos automáticos (*i.e.* [FBP<sup>+</sup>10, SD09]) ou com aprendizado (*i.e.* [ZQJ<sup>+</sup>14, ZGX13]) mencionados até agora, a entrada do modelo foca na interação com o usuário, o qual manualmente definirá as características descritivas do indivíduos ou grupos que deseja re-identificar. A não necessidade de aprendizado dos descritores a serem buscados possibilita que o modelo seja distribuído como aplicação para o usuário final, que poderá fazer consultas em seu banco de imagens da forma que lhe for necessário. Nesta etapa manual, as cores salientes e/ou predominantes dos atributos *soft-biométricos* (calças ou camisetas, por exemplo) da pessoa buscada (e, se for o caso, das demais pessoas nas proximidades) são definidas utilizando imagens-exemplo. Por cor saliente, entende-se a cor que diverge das demais na cena. Já cor predominante, é aquela que ocupa a maior parte do atributo. A Figura 3.1 ilustra os atributos salientes (a) nas regiões em amarelo e vermelho das jaquetas dos indivíduos em primeiro plano na cena; em (b) as cores predominantes (casacos branco e rosa) das duas pessoas à esquerda são exemplificadas. Tanto em (a) quanto (b), as sacolas e a mochila carregadas pelos indivíduos também podem ser selecionadas para busca, além da própria relação de proximidade/agrupamento entre as pessoas.



Figura 3.1 – Exemplo de imagens que podem ser descritas pelo usuário através das cores selecionadas para buscar e re-identificar os indivíduos nas demais cenas. O casal em (a) carrega sacolas vermelhas e possuem jaquetas predominantemente pretas, com saliências em amarelo e vermelho. Os indivíduos em (b) vestem casacos de cores branca e rosa predominantes, sendo que o segundo porta uma mochila preta.

As características descritivas (*i.e.* [ZOW13b, ZOW13a]) e os atributos semânticos (*i.e.* [LHG12, BCBT10]) são então definidos, respectivamente, pelas cores selecionadas e

o modelo de corpo 2D por elas definido para ser buscado. As etapas que dão continuidade ao modelo são automáticas: i) indexação dos possíveis candidatos através de um detector de pessoas ([DT05]), ii) cálculo das distâncias de cores - baseado nas cores do modelo de corpo 2D - que geram um valor de erro para cada pessoa detectada, iii) elaboração do *ranking* com os menores erros calculados - as pessoas mais similares e, se requisitado, iv) detecção das ocorrências de grupos e geração do *ranking* com grupos mais prováveis, para desambiguar candidatos similares e aumentar a taxa de re-identificação ([ZGX14]).

Das etapas aprofundadas teoricamente no Capítulo 2, o modelo utiliza cores para detecção e descrição de características, porém não faz uso de técnicas de aprendizado e classificação. A segmentação utilizada é baseada em *pixels* buscando por regiões homogêneas e a avaliação de espaços de cores e medidas de similaridade entre cores são realizadas especificamente para esta implementação durante a avaliação dos resultados.

O Capítulo 4 descreve em detalhes o modelo desenvolvido para re-identificação de pessoas através de características descritivas de cores e grupos.

## 4. MODELO

Este capítulo descreve o modelo proposto para re-identificar pessoas em ambientes reais através de descrições de cores que representam atributos *soft*-biométricos, como vestimentas e objetos portados. Se aplicável, definições de agrupamento ou proximidade com outras pessoas na cena também podem ser definidas para auxiliar na re-identificação.

O modelo semi-automático é inicializado permitindo ao usuário selecionar cores homogêneas (salientes e/ou predominantes, indiferentemente) que descrevem a(s) pessoa(s) buscada(s). Esta etapa pode ser executada através da seleção de regiões de interesse em uma imagem - seja em uma foto disponível do suspeito ou de um repositório qualquer - ou através de uma paleta de cores. As cores selecionadas são associadas ao modelo de corpo 2D definido por três atributos (*cabeça*, *tronco* e *pernas*) para que a busca no banco de imagens seja executada automaticamente e reconheça/re-identifique as melhores correspondências para o indivíduo ou grupo referido. O fluxo das etapas para execução do modelo é ilustrado na Figura 4.1.

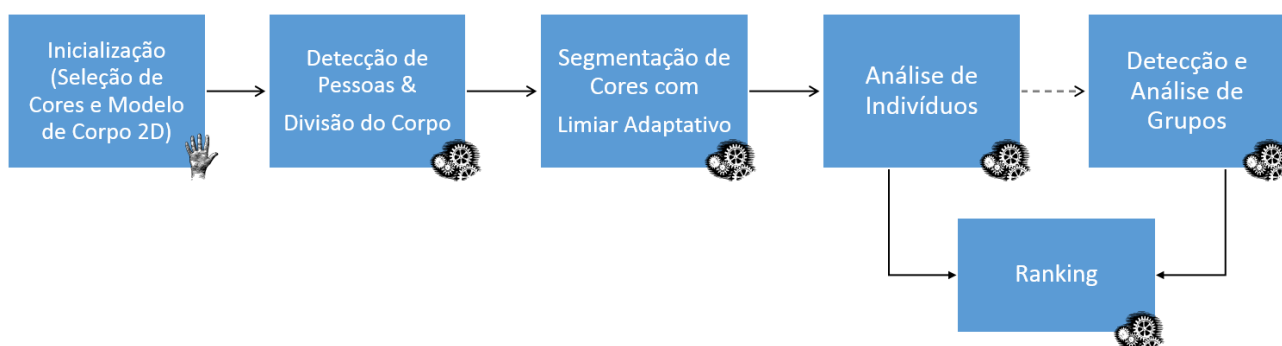


Figura 4.1 – Ilustração das etapas do modelo desenvolvido para re-identificação de pessoas e grupos. Na etapa manual (Inicialização), o usuário define a assinatura da pesquisa e o modelo automaticamente retorna, ao final das etapas, um *ranking* com os mais semelhantes indivíduos ou grupos.

As próximas seções apresentam em detalhes cada etapa da Figura 4.1.

### 4.1 Inicialização - Seleção de cores e construção do modelo de corpo 2D

A primeira etapa do modelo consiste na construção dos atributos da pessoa a ser buscada - a aparência das roupas do suspeito, por exemplo - através da seleção de cores salientes ou predominantes. Para cada pessoa buscada  $l$ , é possível selecionar até  $n$  (experimentalmente,  $n = 3$ ) cores para cada atributo. Os atributos definidos para esta

abordagem foram: *cabeça*, *tronco* e *pernas*. A seleção das cores é feita em imagens de exemplo onde o usuário seleciona as  $n$  regiões salientes ou predominantes que ele julgar melhor descreverem as vestimentas e/ou acessórios do suspeito. A cor média de cada região selecionada irá gerar o modelo de cor para o atributo, representado por  $T_{km}$  (onde  $k = [0, 1, 2]$  para *cabeça*, *tronco* e *pernas*, respectivamente, e  $m = [0, 1, \dots, n - 1]$  é o indexador para cada cor associada ao atributo).

Trabalhar com cores sempre levanta a questão crucial de qual espaço de cores deve ser utilizado. Diversos espaços de cores já foram utilizados anteriormente [LM01] e a recomendação varia para cada aplicação. Foi optado, inicialmente, pela utilização do *Lab*, uma vez que este apresenta cores mais uniformemente espaçadas que no *RGB* ou *HSV* [CG99] e leva em consideração a percepção humana na diferença de cores [Goo12]. Durante os experimentos, a escolha do *Lab* foi ainda validada através de uma análise descrita em detalhes na Seção 5.2, comparando diferentes espaços de cores/distâncias de similaridade específicos para esta aplicação. O modelo foi então definido para operar sobre este espaço de cores, como será visto em detalhes na Seção 4.3.

A Figura 4.2 ilustra a etapa de inicialização. Na Figura 4.2(a), dois tons de azul são selecionados para a região do *tronco* ( $T_{10}$  e  $T_{11}$ , representada pelos retângulos vermelhos) assim como uma cor predominante para as *pernas* ( $T_{20}$ , representado pelo retângulo verde); nenhuma cor foi selecionada para o atributo *cabeça* ( $T_0$ ). A Figura 4.2(b) ilustra o modelo de corpo 2D gerado para a pessoa buscada ( $I$ ), construído com a cor *Lab* média de cada seleção. É importante ressaltar que o modelo é flexível e, portanto, esta etapa poderia ter sido executada selecionando as cores de uma paleta de cores ou de quaisquer outras imagens em uma galeria, o que possibilitaria a busca pelo suspeito baseado na descrição de cores sem uma imagem prévia.

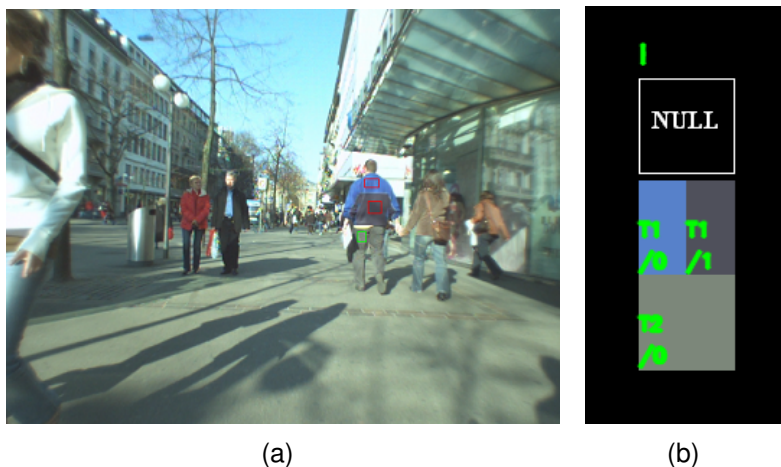


Figura 4.2 – Inicialização e seleção de cores. A seleção de cores na imagem (a) gera o modelo de corpo 2D para a pessoa buscada ( $I$ ), ilustrado em (b).

Finalizada a inicialização manual, o modelo irá executar automaticamente as etapas de busca no banco de imagens para retornar um *ranking* dos indivíduos mais semelhantes ao modelo 2D construído.

## 4.2 Detecção de pessoas e divisão do corpo em atributos

Na segunda etapa do modelo, uma lista de todos os possíveis candidatos - todas as pessoas detectadas no banco de imagens - é montada. Cada pessoa candidata  $P$  da lista terá, então, sua imagem dividida de acordo com os atributos especificados: *cabeça*, *tronco* e *pernas*.

Dada a flexibilidade do modelo, há várias abordagens para montar esta lista, sendo a única restrição que, ao menos, uma caixa delimitadora (*bounding-box*) ao redor das pessoas seja o resultante. Exemplos de métodos possíveis são: detecção de pessoas ([DT05, SKHD09]), detecção de cabeças ([EG09]) e detecção cabeça-ombro ([WZM13, JJRM14]) - estes dois últimos estimando a caixa delimitadora.

Neste trabalho, levando em consideração os bancos de imagens utilizados, duas abordagens para preencher a lista de possíveis candidatos foram utilizadas:

- para cada imagem (cena) no banco de imagens em que o suspeito será buscado (ou para cada quadro de um vídeo), as pessoas são detectadas através de um detector de características baseado em HOG ([DT05], na forma tal qual implementado na biblioteca *OpenCV*<sup>1</sup>);
- em um banco de imagens cujas pessoas já sejam definidas pelo recorte (*crop*) da imagem, ou seja, somente as pessoas são disponibilizadas sem o contexto da cena, as pessoas são automaticamente adicionadas à lista - assume-se que já foram detectadas em um estágio prévio.

As abordagens acima referem-se, respectivamente, aos bancos de imagens ETHZ<sup>2</sup> ([ELS<sup>+</sup>08, ELVG07]) e VIPeR<sup>3</sup> ([GBT07]), ambos bastante referenciados na literatura no que tange problemas de re-identificação.

Para cada pessoa candidata  $P$  na lista, independente do banco de origem, o modelo divide a estrutura de seus corpos em três partes, de acordo com os atributos aqui utilizados: i) *cabeça*, cujo tamanho é definido por 17% da altura da pessoa; ii) *tronco* representado por 33% da altura e iii) *pernas*, 50%. A definição de altura está relacionada com a altura do *bounding-box* (ou do recorte) da pessoa durante sua inserção na lista. Estas medidas foram empiricamente definidas e também estão relacionadas aos bancos de

<sup>1</sup><http://opencv.org/>

<sup>2</sup>Disponível em <http://www.vision.ee.ethz.ch/~aess/dataset/> e <http://www.vision.ee.ethz.ch/~aess/iccv2007/>

<sup>3</sup>Disponível em <http://vision.soe.ucsc.edu/node/178>



imagens utilizados, onde as pessoas são vistas por uma câmera lateral, deixando o modelo dependente da visão da câmera. Todavia, tais valores - e abordagem para inserção na lista - podem ser facilmente modificados para utilização de outros bancos de imagens e diferentes ângulos de visão. A Figura 4.3 ilustra a divisão dos atributos para cada pessoa detectada em uma cena do banco ETHZ (a) e para cada recorte de pessoa de um subconjunto do banco VIPeR (b).



Figura 4.3 – Detecção de pessoas e divisão do corpo em atributos em uma cena do banco ETHZ (a) e em um subconjunto do banco VIPeR (b) (redimensionados para efeitos de visualização).

Tendo todas as pessoas candidatas ( $P$ ) listadas e os modelos  $T_{km}$  de cada atributo da pessoa buscada ( $I$ ), a etapa seguinte fará a relação de distância entre as cores dos atributos dos candidatos e do modelo descritivo.

### 4.3 Segmentação de cores com limiar adaptativo

O cerne do modelo reside nesta etapa de segmentação. É aqui onde todos os candidatos  $P$  que possuem atributos cujas cores sejam semelhantes às selecionadas para o modelo 2D na inicialização (atributos  $T_{km}$ ) serão destacados. Apesar de parecer uma etapa simples de segmentação, vários desafios da área se fazem presentes - iluminação, oclusão, resolução, etc. - e influenciam diretamente nas cores visualizadas e resultados.

A métrica de distância de similaridade escolhida para executar esta tarefa é chamada  $\Delta E_{94}$  (ou CIE94) [Goo12]. A métrica objetiva, levando em consideração a percepção humana, retornar uma distância entre cores dentro do espaço de cores  $LCh$  que, por sua vez, tem as componentes  $Ch$  derivadas do  $ab$  do  $Lab$ . A seguir será detalhado o papel do  $\Delta E_{94}$  na segmentação das imagens.



#### 4.3.1 Segmentação pela métrica de distância $\Delta E_{94}$

Em 1931, o CIE (*International Commission on Illumination*) padronizou sistemas de cores de acordo com a fonte de luz, observador e metodologia utilizada para derivar os valores que descrevem cores. Estes sistemas possuíam limitações de cromaticidades e, então, em 1976, o CIE1976 ( $L^*a^*b^*$ ) surgiu como um dos espaços de cores recomendados para considerar a percepção humana. Também conhecido como CIELAB ou simplesmente *Lab*, o espaço é um padrão internacional onde as cores são perceptivelmente mais uniformes que cores no *RGB* ou *HSV* [CG99]. Desta maneira, a diferença perceptível entre duas cores no *Lab* poderia ser aproximada através de distâncias Euclidianas. Todavia, sendo o ser humano mais sensível a certas cores que outras, métodos para calcular a distância entre duas cores com mais exatidão perceptual precisam levar em consideração este fator. A primeira métrica para resolver este problema foi o CIE94, ou  $\Delta E_{94}$  [Goo12]. Conforme já apresentado na Seção 2.4, dado duas cores *Lab*, a distância de similaridade  $\Delta E_{94}$  entre elas é calculada por

$$\Delta E_{94}^* = \sqrt{\left(\frac{\Delta L^*}{k_L S_L}\right)^2 + \left(\frac{\Delta C_{ab}^*}{k_C S_C}\right)^2 + \left(\frac{\Delta H_{ab}^*}{k_H S_H}\right)^2}. \quad (4.1)$$

Os parâmetros utilizados neste modelo foram definidos como segue.  $S_L$ ,  $S_C$ ,  $S_H$ , os parâmetros de ponderação que ajustam as diferenças do CIE em relação à localização do padrão CIE1976, foram valorados tal que  $S_L = 1$ ,  $S_C = 1 + K_1 C_1^*$ ,  $S_H = 1 + K_2 C_1^*$ ). Uma vez que a entrada da busca é representada pela cor descritiva da roupa dos candidatos nos atributos *tronco* e *pernas* (ou cores de acessórios como boné ou mochila), os parâmetros  $k_L$ ,  $k_C$  e  $k_H$  foram definidos como usado em aplicações têxteis:  $k_L = 2$ ,  $k_C = 1$ ,  $k_H = 1$ ,  $K_1 = 0,048$  e  $K_2 = 0,014$  [Cho14].

Uma vez que imagens digitais são normalmente obtidas no espaço *RGB* - e assim o são as imagens dos bancos utilizados, uma conversão para *Lab* é necessária para comparação. Não foi definida uma conversão direta *RGB-Lab*, o que implica em uma conversão *RGB-XYZ* e outra *XYZ-Lab*. Dado um valor *RGB* normalizado, a conversão é feita através da Equação 4.2. Os valores da matriz de conversão *XYZ* referem-se ao *RGB-Padrão* (*sRGB*), sem correção de *gamma*.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0,4124 & 0,3576 & 0,1805 \\ 0,2126 & 0,7152 & 0,0722 \\ 0,0193 & 0,1192 & 0,9505 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (4.2)$$

Tendo o valor da cor no espaço *XYZ*, este é utilizado na obtenção do valor *Lab* correspondente através da Equação 4.3.

$$\begin{aligned}
L^* &= 116f(Y/Y_n) - 16, \\
a^* &= 500[f(X/X_n) - f(Y/Y_n)], \\
b^* &= 200[f(Y/Y_n) - f(Z/Z_n)],
\end{aligned}
\tag{4.3}$$

onde  $X_n$ ,  $Y_n$  e  $Z_n$  são os valores de referência do *whitepoint* D65 (0,9505, 1,0 e 1,0890, respectivamente) e os declínios infinitos com  $t = 0$  são evitados pela função  $f$ , conforme Equação 4.4:

$$f(t) = \begin{cases} t^{1/3} & \text{se } t > (\frac{6}{29})^3, \\ \frac{1}{3}(\frac{29}{6})^2 t + \frac{4}{29} & \text{caso contrário.} \end{cases}
\tag{4.4}$$

Por fim, a segmentação propriamente dita é executada: para cada pessoa  $P$  na lista de candidatos, todos os *pixels* dentro de cada atributo (*cabeça*, *tronco* e *pernas*) são confrontados - usando a distância  $\Delta E_{94}$  - a seus respectivos atributos nos modelos de cores  $T_{km}$  definidos pelo usuário para a pessoa buscada ( $I$ ). Cada parte do corpo  $k$  gerará um mapa de distância  $D_{km}$ . *Pixels* com distâncias menores que um limiar pré-definido (limiar  $Th_{km}^*$ ) são mantidos; caso contrário, são ignorados. A Figura 4.4(c) ilustra o resultado de uma segmentação utilizando o modelo de cores selecionados na Figura 4.2(b), na cena da Figura 4.2(a) (repetidos para melhor visualização nas Figuras 4.4(b) e 4.4(a), respectivamente). O limiar utilizado foi  $Th_{km}^* = 3$ , selecionado experimentalmente para ilustração.

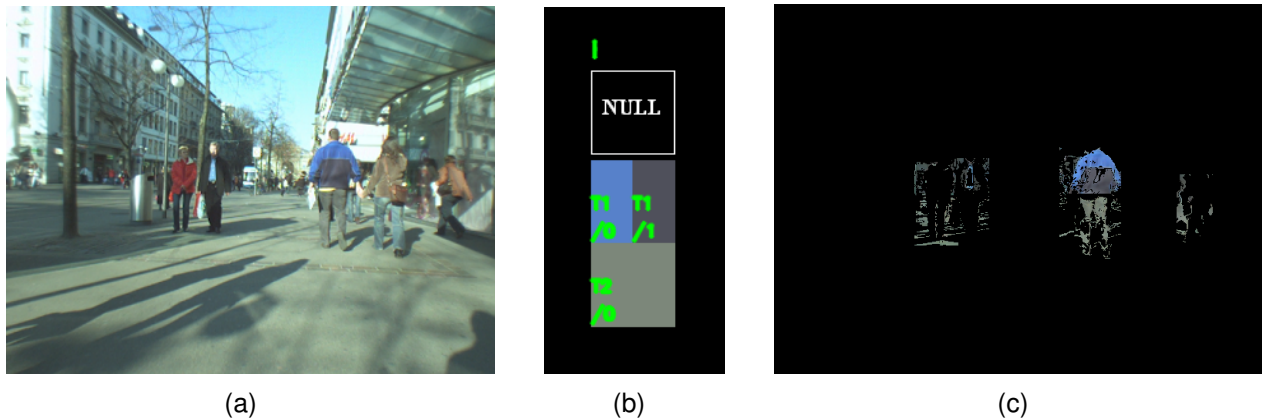


Figura 4.4 – O resultado da segmentação (c) para a cena apresentada em (a). A segmentação utilizou o modelo de cores selecionado em (b) e limiar  $Th_{km}^* = 3$ .

Para validação da escolha do *Lab* como espaço padrão de cores e  $\Delta E_{94}$  como métrica de distância de similaridade, uma comparação dos espaços de cores e métricas de distâncias específicos para esta aplicação - elencados na Seção 2.4 - foi avaliada durante os experimentos e será apresentada na Seção 5.2.

Foi observado ainda, durante o desenvolvimento do trabalho, que a qualidade da imagem pode influenciar diretamente no limiar  $Th_{km}^*$  adotado, fazendo a escolha deste limiar um fator determinante no sucesso ou falha da etapa de segmentação. Para automatizar a

escolha do limiar - e também evitar valores escolhidos manualmente - foi desenvolvida uma abordagem para calcular adaptativamente o limiar  $Th_{km}^*$  através da análise do histograma  $D_{km}$ , como especificado a seguir.

#### 4.3.2 Limiar adaptativo modificado

A abordagem para cálculo do limiar adaptativo apresentada a seguir foi baseada na implementação de Jacques Junior e equipe [JJDJ<sup>+</sup>10], com o auxílio do autor na condição de co-orientador deste trabalho. Esta abordagem utiliza o mapa de distância  $D_{km}$  que definirá o limiar da segmentação, partindo do princípio que o objeto a ser segmentado é o que possui a menor distância, ou seja, o objeto cuja cor é mais similar ao modelo da cor  $T_{km}$ . Na implementação original ([JJDJ<sup>+</sup>10]), Jacques Junior e equipe fazem uso de um modelo baseado em histogramas para calcular o limiar desejado, dado uma cor de referência e uma região de busca. Os autores assumem que *pixels* de fato relacionados com a cor desejada de segmentação irão apresentar menores distâncias (gerando um pico perto da origem do histograma), enquanto *pixels* de outras estruturas tendem a apresentar valores maiores (gerando uma calda ou picos menores). O limiar desejado deverá estar entre o primeiro (normalmente o maior) máximo local e o primeiro mínimo local. Contudo, como mencionado pelos autores na implementação original, apesar da escolha do primeiro mínimo local parecer adequada, há casos em que o histograma é monotonicamente decrescente e não há mínimo local. Logo, ao invés de buscar pelo mínimo local, eles buscam por um ponto no histograma que seja suficientemente plano.

Mais precisamente, com  $h(D_{km})$  denotando o histograma suavizado do mapa de distâncias  $D_{km}$  e  $F_1$  e  $F_2$  sendo as posições do primeiro máximo local e primeiro mínimo local, respectivamente, o limiar desejado pode ser obtido através da Equação 4.5.

$$Th_{km} = \min\{D_{km} | F_1 < D_{km} < F_2 \wedge h''(D_{km}) > 0 \wedge |h'(D_{km})| \leq \alpha\}, \quad (4.5)$$

onde  $\alpha$  é o “limiar de achatamento” (setado experimentalmente em  $0.5774 \equiv 30^\circ$ ). A segunda derivada é incluída para evitar a seleção de pontos com baixa derivativa próximos ao máximo local (onde  $h''(D_{km}) < 0$ ), para que o limiar seja selecionado depois do ponto de inflexão, conforme exemplificado na Figura 4.5(c).

Uma desvantagem desta abordagem na sua forma original ocorre quando a região utilizada para computar o histograma é grande o suficiente para incluir diversos *pixels* com distâncias pequenas e ligeiramente diferentes dos quais se deseja segmentar. Em outras palavras, valores indesejados podem ser inclusos na classe de *pixels* que são buscados, como ilustrado na Figura 4.5(a-d) para o atributo *pernas* de um específico candi-

dato, fazendo com que quase a totalidade da região seja segmentada como de interesse (Figura 4.5(d)).

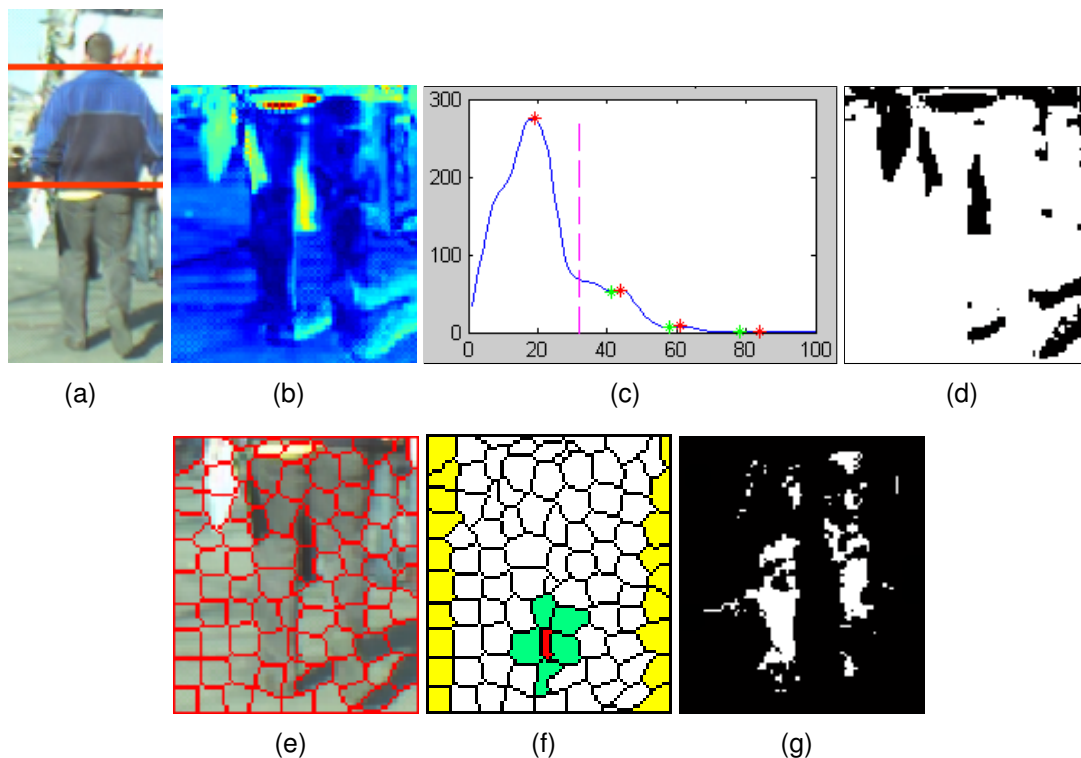


Figura 4.5 – Uma visão geral da abordagem do limiar adaptativo. (a) imagem em análise sub-dividida em atributos; (b) mapa de distâncias  $\Delta E_{94}$  para o atributo *pernas* (parte inferior de (a)) - regiões escuras são as menores distâncias, computadas utilizando o modelo de cor  $T_{20}$  ilustrado na Figura 4.4(b); (c) seleção do limiar adaptativo (linha tracejada vertical); (d) resultado da segmentação utilizando a implementação original ([JJDJ+10]); (e) resultado do algoritmo *SLICO Superpixel* para o atributo *pernas*; (f) a célula com o menor limiar computado (em vermelho) e as células conectadas a ela (em verde); (g) resultado da segmentação com o limiar adaptativo modificado.

Para contornar este problema na segmentação global do atributo, a abordagem foi modificada no presente trabalho: propõe-se subdividir o mapa de distâncias em pequenas células para, então, computar o limiar local de cada célula e suas adjacentes, usando a implementação baseada em histograma da forma original. A divisão da região de interesse (cada parte do corpo) em pequenas células é feita com auxílio do algoritmo *SLICO Superpixel*<sup>4</sup>, proposto por Achanta e equipe [ASS<sup>+</sup>12]. A ideia é computar o limiar para cada célula usando suas respectivas distâncias e, em adicional, as distâncias das células a ela conectadas (como ilustrado na Figura 4.5(f), coloridas em vermelho e verde, respectivamente). A hipótese agora assumida é de que quando o mapa de distâncias é dividido em pequenas células, haverá, ao menos, uma célula na qual a distância desejada está isolada, gerando o pico desejado próximo à origem do histograma computado - então relacionado ao melhor limiar. Em suma, diferentes limiares são computados - de acordo com o número de

<sup>4</sup><http://ivrg.epfl.ch/research/superpixels>

células, e então a célula com menor valor de limiar computado é retornada, como definido na Equação 4.6. Vale ressaltar que esta abordagem será computada separadamente para cada mapa de distâncias  $D_{km}$ , gerando um limiar  $Th_{km}^*$  para cada atributo  $k$  e modelo de cor  $m$ .

$$Th_{km}^* = K \min_{i=1 \text{ to } p_k} Th_{km}(i), \quad (4.6)$$

onde  $p_k$  é o número de células geradas e  $K$  é o fator de escala adotado (setado empiricamente para  $K = 2$ ), usada para dar flexibilidade ao limiar adaptativo. Ainda, para lidar com imagens ruidosas e variações na iluminação, considera-se pico máximo (calculando  $F_1$  e  $F_2$ ) onde há um valor máximo precedido (à esquerda) por um valor inferior em  $\delta$  (onde  $\delta = 0,5$ , escolhido com base nos experimentos executados).

O *SLICO Superpixel* é aplicado para cada atributo  $k$  da pessoa candidata  $P$  em sua respectiva imagem *RGB*. O número de células *superpixels*  $p_k$ , para cada atributo  $k$ , é uma fração da área  $A_k$  por ele ocupada (experimentalmente definido por  $p_k = A_k 0,015$ ). A Figura 4.5(e) ilustra o resultado do algoritmo *SLICO Superpixel* aplicado para o atributo *pernas* (mostrado na Figura 4.5(a), região inferior). A Figura 4.5(f) ilustra a célula selecionada (em vermelho) com o menor limiar computado, cercada pelas células adjacentes (em verde), usadas durante o cálculo do limiar. O resultado obtido na segmentação utilizando a modificação aqui proposta para o limiar adaptativo é mostrado na Figura 4.5(g).

Adicionalmente, notou-se que as partes do corpo relacionadas aos atributos desejados (*cabeça*, *tronco* e *pernas*) usualmente não estão conectadas às bordas verticais de seus respectivos *bounding-boxes* (assim como, usualmente, o *bounding-boxes* contém *pixels* de fundo). Para minimizar a segmentação indesejada de *pixels* de fundo, foram ignoradas as células do *Superpixel* que estão conectadas às bordas verticais durante o cálculo do limiar adaptativo (pré-processamento) e também ignorados os *pixels* dentro delas após a segmentação (pós-processamento) - ilustrados na Figura 4.5(f) em amarelo. Ainda, para prevenir problemas causados por uma segmentação ruim dos resultados, a segmentação é considerada nula (*NULL*) quando a área dos *pixels* segmentados for menor que 1% da área  $A_k$  - relacionada ao atributo  $k$  em análise, uma vez que é demasiado pequena para representar um atributo nesta abordagem.

De posse dos resultados da segmentação para cada atributo de cada candidato, o modelo prossegue com a etapa de análise e *ranking* dos candidatos.

#### 4.4 Análise e *ranking* de indivíduos

O resultado da etapa de segmentação é uma imagem binária para cada parte do corpo/atributo  $k$  e cor selecionada  $m$ . Tal imagem é usada para calcular o valor do erro

médio  $E_{km}$ , o qual é definido sobre o mapa de distância  $D_{km}$ , como uma distância média, considerando a localização dos *pixels* segmentados relacionados ao atributo específico  $k$  e a cor selecionada  $m$ . O valor do erro médio calculado  $E_{km}$  é então usado para computar a medida de erro total de um candidato, que definirá sua posição no *ranking* de similaridade com o modelo buscado, como será mostrado a seguir.

A medida de erro total  $S(I, P)$  para determinada pessoa  $P$  da lista de candidatos, comparado à pessoa de referência  $I$  (conjunto dos atributos inicializados pelo usuário), é computada através da Equação 4.7.

$$S(I, P) = \sum_{s=0}^{s'-1} E_{0s} + \sum_{t=0}^{t'-1} E_{1t} + \sum_{u=0}^{u'-1} E_{2u} + \sum_{k'=0}^{k-1} W_{k'}, \quad (4.7)$$

onde  $s'$ ,  $t'$  e  $u'$  são o número de cores selecionadas (pelo usuário) para cada atributo  $k'$  (*cabeça*, *tronco* e *pernas*, respectivamente), e  $W_{k'}$  é o valor da penalização para o atributo  $k'$  quando alguma das cores selecionadas não é encontrada (ou seja, quando ocorre uma segmentação *NULL*), definido conforme a Equação 4.8. O valor de tal penalidade é igual à zero ( $W_{k'} = 0$ ) para o específico atributo  $k'$  quando o usuário não atribui nenhuma cor a ele ou quando todas as cores selecionadas foram encontradas.

$$W_{k'} = \begin{cases} 2v\mu_{k'} & \text{se } B = \text{verdadeiro} \\ 2n\mu_{k'} & \text{se } B = \text{falso,} \end{cases} \quad (4.8)$$

onde

$$\mu_{k'} = \frac{1}{z'} \sum_{z=1}^{z'} Th_{k'}^*(z), \quad (4.9)$$

$z'$  é o número de limiares adaptativos calculados para o atributo específico  $k'$  (considerando  $m$  cores selecionadas com segmentação válida, não nula).  $B = \text{verdadeiro}$  significa que o usuário selecionou  $m$  cores (até  $n = 3$ ) para o atributo  $k'$  e ao menos uma cor foi encontrada, e  $v$  é o número de cores não encontradas no atributo (por exemplo,  $v = 1$  se o usuário selecionou 3 cores para o atributo *tronco* e a segmentação encontrou somente 2 cores com área não nula);  $B = \text{falso}$  implica que o usuário selecionou  $m$  cores para o atributo específico e nenhuma foi encontrada (por exemplo, o usuário selecionou uma cor para o atributo *pernas* e a segmentação resultou em uma área nula (*NULL*)). Desta maneira, garante-se que quando nenhuma cor é encontrada para o atributo em questão, o erro será maior que em qualquer outra situação onde pelo menos uma cor for encontrada.

Por fim, o candidato  $P$  com o menor  $S$  será o mais semelhante ao modelo buscado  $I$ . A Figura 4.6(b-f) mostra o erro  $S$  para alguns candidato da cena (a). A primeira posição no *ranking* - a que apresenta menor erro ( $S = 4.29$  na Figura 4.6(d)) - foi corretamente associada à pessoa vestindo jaqueta com dois tons de azul previamente selecionado pelo usuário em outra cena/câmera (Figura 4.2).



Figura 4.6 – Ilustração dos erros ( $S$ ) para algumas pessoas candidatas ( $P$ ), computados em relação à pessoa buscada ( $I$ ) ilustrada na Figura 4.2.

#### 4.5 Detecção, análise e *ranking* de grupos

Como previamente mencionado, o modelo proposto pode incorporar informações sobre outra pessoa caminhando ao lado da (ou cruzando pela) pessoa buscada, se aplicável à situação. Esta abordagem pode ser de grande utilidade em casos onde várias pessoas candidatas são similares ao modelo de referência buscado. Se a pessoa buscada caminha próximo ao, por exemplo, seu amigo, o qual possui características discriminativas sobressalentes, esta informação pode desambiguar a busca e melhorar os resultados.

Foi definido como grupo um par de indivíduos em uma cena (imagem/quadro que contém  $p$  candidatos detectados -  $P_1, P_2, \dots, P_p$ ) com uma distância  $d$  entre si menor que um limiar  $2Tg$ , onde  $d$  é a distância Euclidiana entre eles (calculada utilizando o centro de seus respectivos *bounding-boxes*) e  $Tg$  é a menor largura de *bounding-boxes* dentre os indivíduos analisados. Esta “menor largura” foi escolhida para prevenir classificações errôneas de grupos causadas por problemas de perspectivas como uma pessoa bastante próxima à câmera (com maior largura) agrupada com outra longe da câmera (e largura pequena). Em outras palavras, o objetivo é utilizar o tamanho das pessoas detectadas como limiar onde uma pessoa deve estar “distante um corpo de largura” da outra pessoa para caracterizar um grupo, independente da direção em que andam. Como as pessoas normalmente não caminham tão próximas na rua, os grupos foram detectados utilizando o limiar  $2Tg$ , o que significa que as pessoas podem estar “distantes dois corpos de largura” e formarem um grupo. Vale ressaltar que este limiar pode ser facilmente alterado para suportar situações com maiores ou menores densidades de pessoas. O resultado da detecção de grupos é ilustrado na Figura 4.7, onde a relação de agrupamento entre os indivíduos detectados na cena é mostrada através da ligação em vermelho.

Para re-identificar um grupo, o usuário necessita selecionar as cores dos atributos para dois indivíduos de referência buscados ( $I_1$  e  $I_2$ ). Estas cores serão utilizadas durante o



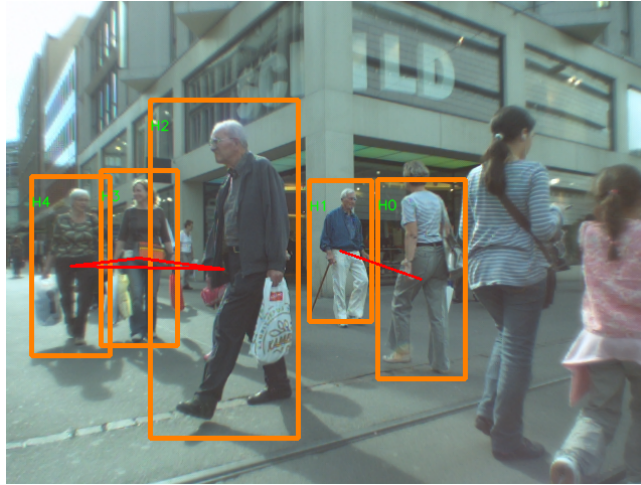


Figura 4.7 – Detecção de grupos: os indivíduos detectados em uma cena do banco ETHZ (delimitados por seus *bounding-boxes* em laranja) e as relações de agrupamento entre si (linhas em vermelho). Pessoas não demarcadas por *bounding-box* não foram encontradas durante a etapa de detecção.

modelo já descrito para computar dois valores de erro ( $S_{g1}$  e  $S_{g2}$ ), associado a cada grupo detectado, conforme Equações 4.10 e 4.11.

$$S_{g1} = S(I_1, P_1) + S(I_2, P_2), \quad (4.10)$$

$$S_{g2} = S(I_1, P_2) + S(I_2, P_1), \quad (4.11)$$

onde  $S(I_1, P_1)$  é o erro obtido para o modelo buscado  $I_1$  em relação ao candidato  $P_1$ , seguindo a mesma ideia para  $S(I_1, P_2)$ ,  $S(I_2, P_1)$  e  $S(I_2, P_2)$  com o objetivo de permitir troca de posições dentro do grupo. Ao final desta etapa, o par de indivíduos com menor valor  $S_g$  é o grupo mais semelhante aos modelos buscados.

O Algoritmo 4.1 descreve o pseudo-código utilizado para detecção de grupos.

#### Algoritmo 4.1 – Re-identificação de Grupos

- 1: Selecionar as cores para a pessoa  $I_1$
- 2: Selecionar as cores para a pessoa  $I_2$
- 3: **for** cada candidato  $P_i$  **do**
- 4:   computar  $d$  contra demais candidatos  $P_j$  da cena
- 5:   **if**  $d \leq 2Tg$  **then**
- 6:     eles formam um grupo
- 7:   **end if**
- 8: **end for**
- 9: **for** cada grupo **do**
- 10:   computar  $S_{g1}$  e  $S_{g2}$
- 11: **end for**
- 12: Ordenar crescentemente os grupos através da medida  $S_g$



Calculando os erros - ou, inversamente, a medida de similaridade - entre modelos e candidatos, seja individualmente ou por grupos, o modelo aqui especificado provê um método que realiza a re-identificação de pessoas a partir da descrição das cores de seus atributos. No próximo capítulo serão apresentados resultados de um estudo de caso para a re-identificação de indivíduos e grupos, respectivamente nos bancos VIPeR e ETHZ, que validam a eficácia da abordagem.



## 5. RESULTADOS OBTIDOS

Este capítulo apresenta os resultados experimentais obtidos com o modelo descrito. O modelo é validado através de dois diferentes cenários: i) utilizando o banco de imagens VIPeR ([GBT07]) para identificação de indivíduos e ii) aplicando a abordagem para re-identificação de pessoas e de grupos no banco de imagens ETHZ ([ELS<sup>+</sup>08, ELVG07]). O resultados são sumarizados na forma de taxa de re-identificação cumulativa e, ao fim de cada experimento, resultados ilustrativos são apresentados.

Para permitir futuras comparações com o presente trabalho, as informações detalhadas sobre a montagem dos bancos de imagens (subconjuntos e quadros utilizados, por exemplo) estão descritas no Apêndice A, juntamente com as localizações das detecções de grupos e pessoas com seus respectivos identificadores. No mesmo apêndice, as seleções de cores feitas pelo usuário na etapa manual de entrada - que geram os modelos de corpo 2D dos indivíduos a serem buscados - estão relacionadas na Seção A.4.

### 5.1 Re-identificação de pessoas com o banco VIPeR

O primeiro cenário de testes foi avaliado utilizando o banco de imagens VIPeR [GBT07]. Bastante referenciado na avaliação de problemas de re-identificação, o VIPeR é composto por imagens de baixa resolução de pedestres recortados (*crop*), com significativas variações de câmera, poses, iluminação e algumas com oclusão e ruídos de fundo. Possui 632 pares de pedestres - sendo cada par um pedestre visto através de duas câmeras (câmera A e câmera B) em um ambiente externo - e a maioria dos pares apresenta variações no ângulo de visão maiores que 90°. Para este experimento, foi assumido que a imagem de cada pessoa no banco de imagens é o resultado da etapa de detecção de pessoas descrita na Seção 4.2.

Seguindo o mesmo protocolo de avaliação adotado por Zhao e equipe [ZOW13a], foram selecionados randomicamente 50% dos indivíduos contidos no VIPeR (candidatos a serem buscados, 316 imagens), capturados na câmera A. Para cada um destes indivíduos, foi instruído ao usuário que selecionasse ao menos uma cor para o atributo *tronco* e uma para o *pernas* - o atributo *cabeça* foi definido como opcional devido à baixa resolução das imagens. A assinatura de cada pessoa, composta pelas cores médias de cada atributo, foi confrontada contra os mesmos 50% das imagens, porém utilizando os indivíduos capturados pela câmera B. Em outras palavras, as imagens da câmera A são usadas como conjunto de entrada e as mesmas imagens na câmera B como conjunto de busca, de tal forma que toda imagem de entrada seja confrontada com todas imagens na busca. Para a lista das imagens selecionadas, refere-se ao Apêndice A, Seção A.1.

A Figura 5.1(a) ilustra as regiões selecionadas pelo usuário para criar o modelo de cores de um indivíduo (a partir da câmera A) enquanto a Figura 5.1(b-f) ilustra as 5 primeiras posições no *ranking* de mais semelhantes (capturados pela câmera B). A correta re-identificação ocorreu na segunda posição, Figura 5.1(c). Vale ressaltar que a seleção das cores poderia ser feita em qualquer imagem prévia ou definindo uma cor *RGB*, porém foi utilizada a cor média da região selecionada sobre a imagem da câmera A para melhor dinâmica na seleção da entrada e avaliação dos resultados.



Figura 5.1 – Ilustração da re-identificação de um indivíduo utilizando o banco VIPeR. (a) imagem de entrada com as regiões selecionadas pelo usuário (câmera A) para geração do modelo de cores a ser buscado. (b-f) os 5 primeiros resultados - os menores erros na câmera B - com a associação correta na segunda posição do *ranking* (c).

A curva CMC (*Cumulative Matching Characteristic*) para este cenário completo (com 316 imagens) é ilustrada na Figura 5.2. A curva mostra a quantidade cumulativa de re-identificações (em %) até determinado *ranking*. A Tabela 5.1 sumariza os resultados obtidos neste cenário com o limiar adaptativo modificado (Seção 4.3), com o limiar original ([JJDJ+10]) e compara-os com o algoritmo do estado da arte ([ZOW13a], apresentado no Capítulo 3). Na tabela, a quantidade de re-identificações corretas (em %) até determinada posição no *ranking* é representada (r1, re-identificações na primeira posição; r5 até a quinta posição; r10 até a décima posição).

Tabela 5.1 – Resultados para o banco VIPeR: taxa de re-identificação cumulativa (em %) para melhores posições no *ranking* (mais semelhantes) dentre as 316 imagens/pessoas. As últimas duas linhas mostram a melhoria obtida com a modificação efetuada no limiar adaptativo em comparação com a sua forma original ([JJDJ+10]), conforme descrito na Subseção 4.3.2. A primeira linha mostra os resultados da abordagem estado da arte ([ZOW13a]).

Abordagem / <i>Ranking</i>	r1	r5	r10
Resultados no algoritmo do estado da arte ([ZOW13a])	30,16	≈ 55	≈ 63
Resultados obtidos com limiar adaptativo modificado	12,02	25	34,81
Resultados obtidos com limiar adaptativo original ([JJDJ+10])	3,16	14,24	20,57

Como apresentado na Tabela 5.1, a abordagem proposta não tem performance melhor que o estado da arte ([ZOW13a]) neste primeiro cenário. Apesar da modificação no

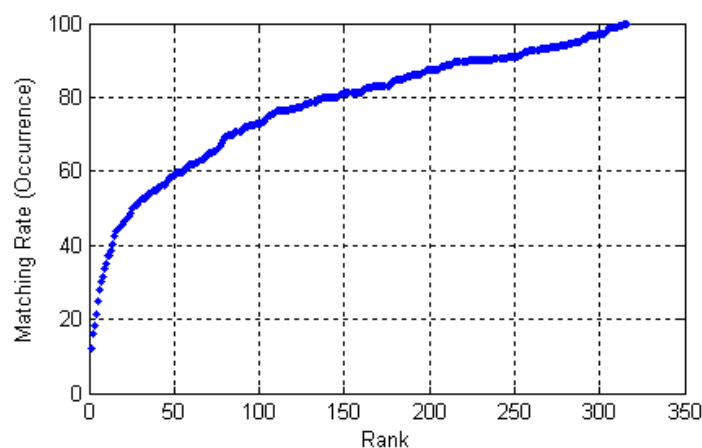


Figura 5.2 – Curva CMC para o subconjunto de 316 imagens do banco VIPeR.

limiar adaptativo ter melhorado os resultados em comparação ao original, a abordagem de segmentação baseada somente em cores dos atributos no modelo 2D não é competitiva com o estado da arte quando manipulando imagens de baixa resolução e grandes variações de ângulo de visão, com o objetivo de re-identificar um indivíduo sozinho. Um fator determinante para a diferença dos resultados pode residir na saliência das cores: enquanto no estado da arte o cerne da busca são as cores salientes - as que se destacam na assinatura do indivíduo, o modelo aqui implementado dá liberdade para o usuário selecionar cores salientes ou predominantes e requer a seleção de ao menos uma cor para o atributo *tronco* e uma para o *pernas*. Selecionar uma cor predominante que seja relativamente comum apenas para validar a seleção do atributo pode gerar erros que não ocorreriam se somente o atributo saliente fosse selecionado.

Para ilustração dos resultados, a Figura 5.3 exemplifica outras re-identificações realizadas neste cenário. A seleção feita pelo usuário na imagem de entrada é ilustrada na primeira coluna, seguida pelas colunas correspondentes aos *rankings* resultantes da primeira a quarta posição. A correta re-identificação é destacada com fundo amarelo e repetida na última coluna quando entre as quatro primeiras do *ranking*. Caso esta não esteja dentre as primeiras quatro posições, a correta re-identificação é apresentada na última coluna.

É importante ressaltar que diferentemente da abordagem proposta por Zhao e equipe [ZOW13a], onde um aprendizado não-supervisionado é necessário para a detecção das saliências, a abordagem aqui apresentada pode re-identificar o indivíduo a partir de características de baixo nível (cores) baseado em imagens de qualquer galeria ou a definição de uma cor *RGB*, não sendo obrigatória a imagem do indivíduo nem utilizado aprendizado. Ainda, quando trabalhando com cenas de ambientes por inteiro, ou seja, com diversas pessoas na cena, é possível fazer uso da informação contextual relacionada à distância (proximidade) entre pessoas buscadas. É neste cenário que o corrente modelo apresenta suas

melhores taxas de re-identificação: como será mostrado no cenário da Seção 5.3, resultados indicam que a re-identificação de pessoas é significativamente melhorada utilizando a característica de alto nível de grupos de pessoas.









Seleção	Rank 1	Rank 2	Rank 3	Rank 4	Rank Encontrado
					 <b>(5)</b>
					 <b>(1)</b>
					 <b>(2)</b>
					 <b>(11)</b>
					 <b>(73)</b>

Figura 5.3 – Resultados ilustrativos da re-identificação no banco de imagens VIPeR. Na primeira coluna, a seleção feita pelo usuário é mostrada. Os destaques em amarelo denotam o *ranking* da correta re-identificação.

## 5.2 Comparação de espaços de cores na re-identificação de pessoas com o banco VIPeR

Antes de focar na re-identificação de grupos, uma questão pendente necessita ser avaliada: o espaço de cores *Lab*, combinado com a medida de similaridade  $\Delta E_{94}$ , é realmente a escolha mais adequada para este modelo?

Para responder tal questão foi repetido o experimento de re-identificação sobre a base VIPeR descrito no cenário anterior, agora com os espaços de cores e medidas de distâncias/similaridade de cores apresentados na Seção 2.4. As mesmas regiões selecionadas foram utilizadas, sendo efetuada somente a conversão das cores médias selecionadas. Os resultados na taxa de re-identificação são sumarizados na Tabela 5.2, através da quantidade de re-identificações cumulativas (em %) por *ranking*. Maior valor representa melhor taxa de acerto, confirmando a escolha do espaço *Lab* e distância  $\Delta E_{94}$  como mais adequados para esta aplicação.

Tabela 5.2 – Comparação da taxa cumulativa de re-identificação (em %) no subconjunto do banco VIPeR com 316 imagens, utilizando quatro pares de espaços de cores/medidas de similaridade. A maior taxa foi obtida com a combinação *Lab* e  $\Delta E_{94}$ .

Abordagem / <i>Ranking</i>	r1	r5	r10	r20
<i>HSV</i> e distância Euclidiana	8,86	17,4	23,73	34,49
<i>HS</i> e distância Euclidiana	9,17	22,15	29,43	40,18
<i>RGB</i> e distância Mahalanobis	9,17	21,83	30,37	43,35
<i>Lab</i> e distância $\Delta E_{94}$	<b>12,02</b>	<b>25</b>	<b>34,81</b>	<b>45,57</b>

É importante salientar que a utilização do limiar adaptativo modificado e o subconjunto específico avaliado (única execução com 316 imagens do VIPeR) inferem diretamente nos resultados. Portanto, trata-se de uma avaliação pontual para fins de comparações e validações dentro do escopo deste trabalho. O espaço de cores *Lab* e a distância de similaridade  $\Delta E_{94}$  foram, então, utilizados nos demais cenários deste trabalho.

## 5.3 Re-identificação de pessoas e grupos com o banco ETHZ

No segundo cenário de avaliação, um subconjunto específico de imagens do banco ETHZ [ELS+08, ELVG07] foi construído para trabalhar com cenas contendo grupos. O banco possui diversas sequências de cenas com pessoas caminhando pelas ruas, sendo cada sequência capturada por um par de câmeras em movimento (doravante referenciadas por câmera A (esquerda) e câmera B (direita)). Esta configuração de câmeras (em heterogêneas cenas) gera diversas variações na aparência dos indivíduos, assim como problemas

de iluminação e oclusão. Para demonstrar o potencial do modelo na re-identificação de grupos, primeiramente foi conduzido um experimento para re-identificação individual (como no cenário anterior) e, posteriormente, a melhoria das re-identificações foi observada no experimento com a informação de grupos.

O subconjunto utilizado foi montado selecionando pares (câmera A e B) de imagens/cenas esparsas em diversas sequências do banco ETHZ. Para cada par de cenas selecionado, a abordagem de detecção de grupos descrita na Seção 4.5 foi aplicada sobre a imagem da câmera B. Os grupos resultantes foram analisados manualmente e descartou-se aqueles cujos indivíduos que o formavam não apareciam na câmera A, afim de garantir que o usuário possa selecionar o modelo de cor do indivíduo na imagem de entrada - mantendo o processo de avaliação do cenário anterior. O subconjunto final totaliza 141 grupos detectados, contendo 213 indivíduos (alguns compartilhando mais de um grupo) em 72 pares de cena.

Em uma aplicação envolvendo usuários há, ainda, outro fator a ser considerado: a correta re-identificação do indivíduo/grupo pode ocorrer em diferente espaço de tempo e/ou câmeras, onde o usuário validará que se trata do mesmo indivíduo buscado porém em outra pose, outro espaço de tempo ou sob outra iluminação, por exemplo. Logo, para este experimento, os 72 pares de cenas foram analisados manualmente em busca de indivíduos e grupos que aparecessem repetidamente em diferentes cenas a fim de montar o conceito de *listas de equivalências*, permitindo que o indivíduo possa ser re-identificado por diferentes câmeras (ou na mesma câmera após significativo espaço de tempo e/ou diferente ambiente) em todas suas ocorrências - sendo a primeira destas re-identificações considerada a correta durante a avaliação.

Nas *listas de equivalências* montadas, dentre os 213 indivíduos no conjunto de busca, 29 indivíduos aparecem pelo menos uma vez mais em outra cena. Destes 29, 2 indivíduos compartilham 6 IDs (aparecem 6 vezes no banco de imagens), 1 indivíduo compartilha 5 IDs, 7 compartilham 3 IDs e 19 indivíduos compartilham 2 IDs. Já se tratando de grupos, o subconjunto contém 11 grupos equivalentes, onde 9 compartilham 2 IDs, 1 compartilha 3 IDs e 1 compartilha 6 IDs.

O detalhamento da construção do subconjunto utilizado, bem como lista de detecções e imagens selecionadas, encontra-se ao Apêndice A, Seção A.2. Já para detalhes e identificadores em cada lista de equivalência, refere-se o Apêndice A, Seção A.3.

Seguindo o mesmo processo do cenário anterior, para cada indivíduo no conjunto de busca (câmera B), foi apresentada ao usuário a cena em que o indivíduo aparecia no conjunto de entrada (câmera A) e solicitado que ao menos uma cor fosse selecionada para cada atributo *tronco* e *pernas - cabeça* continuou opcional. Cada assinatura definida pelo usuário, composta pelas cores médias de cada atributo, foi confrontada com cada um dos 213 indivíduos capturados na câmera B - que formaram a lista de candidatos (Seção 4.2). O limiar adaptativo modificado também foi utilizado neste cenário. A Figura 5.4 ilustra o



resultado da uma detecção de indivíduos no banco ETHZ: o pedestre de camisa azul e calção cinza teve suas regiões selecionadas (a) e foi re-identificado corretamente na segunda posição do *ranking* (c). A Tabela 5.3 compila a taxa (em %) de acertos na re-identificação dos indivíduos através de seus modelos de cores neste cenário.

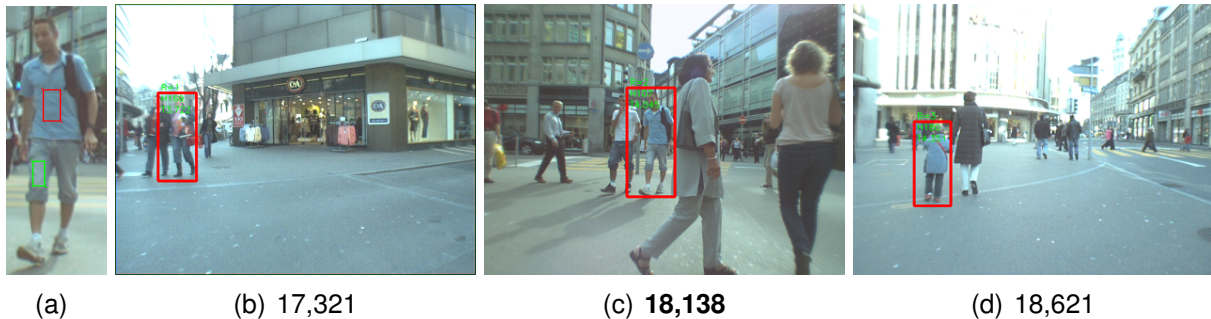


Figura 5.4 – Ilustração dos resultados para re-identificação de um indivíduo utilizando o banco ETHZ. (a) imagem de entrada com as regiões selecionadas pelo usuário (câmera A) para geração do modelo de cores a ser buscado. (b-d) os 3 primeiros resultados - os menores erros na câmera B - com a associação correta na segunda posição do *ranking* (c) (redimensionados para efeitos de visualização).

Tabela 5.3 – Re-identificação de indivíduos no banco ETHZ: taxa de re-identificação cumulativa (em %) para melhores posições no *ranking* (mais semelhantes) dentre os 213 indivíduos.

Abordagem / <i>Ranking</i>	r1	r2	r3	r4
Re-identificação do indivíduo	82,16	90,61	91,55	95,77

Como mostrado na Tabela 5.3, o modelo re-identificou 82,16% dos indivíduos no banco ETHZ em primeiro lugar, enquanto mais de 95% foram detectados até a quarta posição. A Figura 5.5 ilustra outros resultados obtidos nas re-identificações de indivíduos no banco ETHZ. A seleção feita pelo usuário na imagem de entrada é ilustrada na primeira coluna, seguida pelas colunas correspondentes aos *rankings* resultantes da primeira a quarta posição. A correta re-identificação é destacada em amarelo.

Comparações com demais trabalhos de re-identificação não foram possíveis neste cenário. O estado da arte ([ZOW13a]) utiliza os bancos VIPeR e CUHK Campus. Nos demais trabalhos que utilizam o banco ETHZ (*i.e.* [FBP<sup>+</sup>10], [DSS<sup>+</sup>13] ou [WMZ<sup>+</sup>14]), a configuração do banco é baseada na proposta por Schwartz e Davis [SD09], levando em consideração as anotações (*crops*) das pessoas já detectadas e relacionadas em diferentes quadros. Neste trabalho, o subconjunto do banco ETHZ foi construído selecionando cenas/quadros esparsos focando em avaliações *single-shot* e no contexto das pessoas na cena por inteiro, a fim de avaliar distâncias de proximidades/agrupamentos. E, sendo esta avaliação a razão do subconjunto de imagens montado, um segundo experimento foi realizado para validar a melhoria nos resultados quando inclusa tal informação contextual de grupos.



Figura 5.5 – Resultados ilustrativos da re-identificação de indivíduos no banco de imagens ETHZ. Na primeira coluna, a seleção feita pelo usuário é mostrada. Os destaques em amarelo mostram a correta re-identificação e sua respectiva posição no *ranking*.

Utilizando o banco ETHZ e re-aproveitando as cores dos atributos de cada pessoa selecionadas pelo usuário na avaliação anterior, o segundo experimento utiliza a abordagem de detecção e re-identificação de grupos (descrita na Seção 4.5) para medir a taxa de re-identificação (através dos erros totais  $S_g$ ) de cada grupo candidato no conjunto de busca, gerando um *ranking* de semelhança em relação ao grupo buscado. A Figura 5.6(a-b) ilustra dois indivíduos na câmera A com suas respectivas regiões (que geraram os modelos de cores dos atributos) selecionadas pelo usuário. A Figura 5.6(c) mostra o grupo detectado na câmera B, com os indivíduos reconhecidos como um grupo. Abaixo de cada imagem é mostrada a posição no *ranking* obtida quando buscado somente pelos indivíduos (Figura 5.6(a-b)) e quando por grupo (Figura 5.6(c)).

Para avaliar a melhoria obtida com a re-identificação de grupos em relação a re-identificação de indivíduos, calculou-se, para cada grupo, a posição média dos *rankings* de seus membros na re-identificação individual. A Figura 5.6 exemplifica esta avaliação: o indivíduo  $P_1$  mostrado na Figura 5.6(a) foi classificado na primeira posição e, o indivíduo  $P_2$ , mostrado na Figura 5.6(b), na sexta. A posição média dos indivíduos do grupo é 3,5 ( $rank \leq 4$ ). Os mesmos indivíduos, quando buscados como um grupo, foram re-identificados na primeira posição (Figura 5.6(c),  $rank = 1$ ). Fazendo uma analogia para “quantas imagens



Figura 5.6 – Ilustração de um resultado quando duas pessoas são buscadas individualmente e como um grupo. (a-b) ilustra as regiões selecionadas pelo usuário (câmera A), além da posição no *ranking* de cada pessoa quando buscadas individualmente contra o banco de imagens (213 indivíduos, câmera B). (c) ilustra o grupo re-identificado formado pelos mesmos indivíduos e sua posição no *ranking* de grupos (dentre os 141 grupos do banco, na câmera B).

precisam ser manualmente analisadas para visualizar os candidatos buscados”, pode-se afirmar que a informação do grupo ajudou significativamente a encontrar tais indivíduos.

A Tabela 5.4 sumariza as re-identificações de grupo e indivíduos no subconjunto do banco ETHZ, através do modelo apresentado. Devido à diferente abordagem para grupos e ao banco de imagens especificamente montado para este fim, comparações com outros trabalhos não são possíveis. Os trabalhos que avaliam grupos no problema de re-identificação ([ZGX09, ZGX14]) utilizam informação temporal com subtração de fundo para comparar seus descritores e técnicas de aprendizado para ponderar características mais relevantes, necessitando imagem prévia do grupo. Todavia, como ilustrado na Figura 5.6 e reiterado na Tabela 5.4, os experimentos demonstram que dois indivíduos são melhor classificados quando buscados em grupo.

Tabela 5.4 – Resultados obtidos no subconjunto do ETHZ considerando as melhorias na identificação do grupo *versus* a re-identificação de indivíduos (melhores classificações cumulativas (em %) para os 141 grupos e 213 indivíduos). A primeira linha mostra os resultados para a re-identificação de grupos, enquanto a segunda linha sumariza a classificação média quando buscadas individualmente as pessoas de cada grupo (sem a informação contextual do grupo).

Abordagem / <i>Ranking</i>	r1	r2	r3	r4
Re-identificação do grupo	82,26	92,90	96,45	98,58
Re-identificação individual (média) dos membros do grupo	70,92	85,10	93,61	97,12

Como visto nesse cenário, a re-identificação de indivíduos apresentou uma taxa de 95,77% de acertos (Tabela 5.3) até a quarta posição do *ranking* no banco ETHZ, constatando-



se ainda que a informação de alto nível dos grupos pode melhorar a re-identificação, atingindo 98,58% até a mesma posição (Tabela 5.4).

Para melhor ilustração das re-identificações de grupos, a Figura 5.7 mostra resultados adicionais obtidos durante este experimento. As seleções feitas pelo usuário na imagem de entrada (na mesma cena, porém com indivíduos recortados para melhor visualização) são ilustradas na primeira e segunda colunas. Demais colunas correspondem aos *rankings* resultantes da primeira a quarta posição, sendo a correta re-identificação destacada em amarelo.

Grupo		Ranking 1	Ranking 2	Ranking 3	Ranking 4
Seleção 1	Seleção 2				

Figura 5.7 – Resultados ilustrativos da re-identificação de grupos no banco de imagens ETHZ. Na duas primeiras colunas, as seleções feitas pelo usuário para cada indivíduo do grupo é mostrada. Os destaques em amarelo mostram a correta re-identificação e sua respectiva posição no *ranking*.

Ao fim destes cenários, conclui-se que o modelo apresentado neste trabalho é capaz de re-identificar o “suspeito de ‘boné branco’ e ‘jaqueta preta’ que andava junto ao indivíduo de ‘camisa vermelha’ ” buscado a partir da descrição e organização dos atributos *soft-biométricos*.

No próximo capítulo são apresentadas as conclusões finais e sugestões para possíveis trabalhos futuros.

## 6. CONSIDERAÇÕES FINAIS

Este trabalho apresentou um modelo para a re-identificação de pessoas e grupos baseado em características *soft*-biométricas. O modelo utiliza definições manuais de cores - inicializadas por uma imagem do suspeito ou mesmo imagens de qualquer origem - para criar uma assinatura de cores da(s) pessoa(s) buscada(s), organizando-as semanticamente em atributos dentro de um modelo de corpo 2D. As possíveis instâncias do suspeito em outras imagens são classificadas de acordo com as menores diferenças de cores entre seus atributos.

Na etapa inicial, a não necessidade de aprendizado dos descritores a serem buscados (uma vez que informados pelo usuário) possibilita que o modelo seja distribuído como aplicação para o usuário final, que poderá fazer consultas em seu banco de imagens da forma que lhe for necessário, selecionando cores salientes ou predominantes, com ou sem informações de grupos.

Para definir o cálculo de diferença entre as cores, foi elaborado um estudo utilizando quatro espaços de cores e diferentes medidas de distância de similaridade. Os melhores resultados foram obtidos utilizando o espaço *Lab* e a distância  $\Delta E_{94}$ . Elaborou-se também neste trabalho um aperfeiçoamento para a técnica de obtenção do limiar baseado em análise de histograma [JJDJ+10] em conjunto com o autor do trabalho original, automatizando a segmentação das cores semelhantes.

Resultados experimentais demonstram que a abordagem baseada em características descritivas de cores, quando semanticamente organizadas em atributos *soft*-biométricos, podem levar a re-identificação do suspeito buscado. A introdução da informação de grupos corrobora a afirmação de Zheng e equipe [ZGX14], incrementando a taxa de re-identificação e diminuindo a ambiguidade dos resultados.

Responde-se, então, a questão de pesquisa: é possível recuperar de uma imagem informações classificadas como “boné branco”, “jaqueta preta” e “camiseta vermelha”, levando à re-identificação de quem as porta. Tais atributos *soft*-biométricos puderam ser estimados a partir da detecção de pessoas e medidas de distância de cores que indicaram a correspondência entre as descrições. O agrupamento entre os indivíduos também foi fator contribuinte para satisfatória re-identificação dos suspeitos.

Tomando em consideração o estado corrente da pesquisa e o leque aberto para contribuições, observa-se a viabilidade de trabalhos futuros que agregariam valor ao trabalho implementado. Lista-se, por ordem de prioridade estimada, possíveis trabalhos futuros, estando o primeiro item já sob investigação:

- melhorias na segmentação - grande parte dos erros na re-identificação se deve à corrente abordagem considerar *pixels* de fundo durante a segmentação. Sendo pos-

sível eliminar - ou minimizar - a presença destes *pixels*, estima-se que a taxa de re-identificação seja incrementada significativamente;

- grupos com mais de duas pessoas - propriedades associativas na formação de grupos, assim como possíveis análises de diferentes composições destes sem a informação temporal podem ajudar na triagem de resultados;
- utilização de texturas - tornar possível que o usuário selecione uma textura ao invés de uma cor média pode melhorar a taxa de re-identificação; novos desafios na seleção e na comparação da textura se farão presentes nesta implementação; e
- permitir a entrada de dados por descrição textual - estudar a viabilidade de implementar uma entrada de dados interpretando informações textuais como, por exemplo “camiseta vermelha”, introduzindo elementos de inteligência artificial à abordagem.

Conclui-se que, de modo geral, o modelo definido pôde atingir os objetivos deste trabalho e, por ser modularmente customizável, abre diversas oportunidades para expansão e melhorias futuras.

No Apêndice B são apresentadas as publicações obtidas e submetidas durante o curso de mestrado e desenvolvimento deste trabalho.

## REFERÊNCIAS BIBLIOGRÁFICAS

- [AHP06] Ahonen, T.; Hadid, A.; Pietikainen, M. "Face description with local binary patterns: Application to face recognition", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28–12, 2006, pp. 2037–2041.
- [AJKA10] Agarwal, M.; Jain, N.; Kumar, M.; Agrawal, H. "Face recognition using eigen faces and artificial neural network", *International Journal of Computer Theory and Engineering*, vol. 2–4, 2010, pp. 1793–8201.
- [AKB+08] Allaire, S.; Kim, J. J.; Breen, S. L.; Jaffray, D. A.; Pekar, V. "Full orientation invariance and improved feature selectivity of 3d sift with application to medical image analysis". In: *Computer Vision and Pattern Recognition Workshops, 2008. CVPRW'08. IEEE Computer Society Conference on*, 2008, pp. 1–8.
- [ASS+12] Achanta, R.; Shaji, A.; Smith, K.; Lucchi, A.; Fua, P.; Susstrunk, S. "SLIC Superpixels Compared to State-of-the-art Superpixel Methods", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34–11, 2012, pp. 2274 – 2282.
- [AZ08] Arandjelovic, O.; Zisserman, A. "Crowd detection from still images." In: *BMVC, 2008*, pp. 1–10.
- [BCBT10] Bak, S.; Corvee, E.; Brémond, F.; Thonnat, M. "Person re-identification using spatial covariance regions of human body parts". In: *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 2010, pp. 435–440.
- [BGS14] Bedagkar-Gala, A.; Shah, S. K. "A survey of approaches and trends in person re-identification", *Image and Vision Computing*, vol. 32–4, 2014, pp. 270–286.
- [BHW11] Brown, M.; Hua, G.; Winder, S. "Discriminative learning of local image descriptors", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33–1, 2011, pp. 43–57.
- [BTVG06] Bay, H.; Tuytelaars, T.; Van Gool, L. "Surf: Speeded up robust features". In: *Computer Vision–ECCV 2006*, Springer, 2006, pp. 404–417.
- [BVM08] Busin, L.; Vandenbroucke, N.; Macaire, L. "Color spaces and image segmentation", *Advances in imaging and electron physics*, vol. 151, 2008, pp. 65–168.
- [CCS+11] Cheng, D. S.; Cristani, M.; Stoppa, M.; Bazzani, L.; Murino, V. "Custom pictorial structures for re-identification." In: *BMVC, 2011*, pp. 6.

- [CG99] Cai, J.; Goshtasby, A. "Detecting human faces in color images", *Image and Vision Computing*, vol. 18–1, 1999, pp. 63–75.
- [Cho14] Choudhury, A. K. R. "Principles of Colour and Appearance Measurement: Volume 2: Visual Measurement of Colour, Colour Comparison and Management". Woodhead Publishing, 2014.
- [CJSW01] Cheng, H.-D.; Jiang, X.; Sun, Y.; Wang, J. "Color image segmentation: advances and prospects", *Pattern recognition*, vol. 34–12, 2001, pp. 2259–2281.
- [CV95] Cortes, C.; Vapnik, V. "Support-vector networks", *Machine learning*, vol. 20–3, 1995, pp. 273–297.
- [DAL12] Du, Y.; Ai, H.; Lao, S. "Evaluation of color spaces for person re-identification". In: *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 1371–1374.
- [DD01] Dayhoff, J. E.; DeLeo, J. M. "Artificial neural networks", *Cancer*, vol. 91–S8, 2001, pp. 1615–1635.
- [DJLW08] Datta, R.; Joshi, D.; Li, J.; Wang, J. Z. "Image retrieval: Ideas, influences, and trends of the new age", *ACM Computing Surveys (CSUR)*, vol. 40–2, 2008, pp. 5.
- [DSH+09] Dreuw, P.; Steingrube, P.; Hanselmann, H.; Ney, H.; Aachen, G. "Surf-face: Face recognition under viewpoint consistency constraints." In: *BMVC*, 2009, pp. 1–11.
- [DSS+13] Dutra, C. R.; Schwartz, W. R.; Souza, T.; Alves, R.; Oliveira, L. "Re-identifying people based on indexing structure and manifold appearance modeling". In: *Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI-Conference on*, 2013, pp. 218–225.
- [DT05] Dalal, N.; Triggs, B. "Histograms of oriented gradients for human detection". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 886–893.
- [DTS06] Dalal, N.; Triggs, B.; Schmid, C. "Human detection using oriented histograms of flow and appearance". In: *Computer Vision–ECCV 2006*, Springer, 2006, pp. 428–441.
- [EG09] Enzweiler, M.; Gavrilu, D. M. "Monocular pedestrian detection: Survey and experiments", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31–12, 2009, pp. 2179–2195.



- [ELS<sup>+</sup>08] Ess, A.; Leibe, B.; Schindler, K.; ; van Gool, L. “A mobile vision system for robust multi-person tracking”. In: IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1–8.
- [ELVG07] Ess, A.; Leibe, B.; Van Gool, L. “Depth and appearance for mobile scene analysis”. In: 11th IEEE International Conference on Computer Vision, 2007, pp. 1–8.
- [FBP<sup>+</sup>10] Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V.; Cristani, M. “Person re-identification by symmetry-driven accumulation of local features”. In: Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, 2010, pp. 2360–2367.
- [FC54] Farley, B.; Clark, W. “Simulation of self-organizing systems by digital computer”, *Information Theory, IRE Professional Group on*, vol. 4–4, 1954, pp. 76–84.
- [Fis99] Fisher, R. “Change detection in color images”. In: Proceedings of 7th IEEE Conference on Computer Vision and Pattern, 1999.
- [FMI83] Fukushima, K.; Miyake, S.; Ito, T. “Neocognitron: A neural network model for a mechanism of visual pattern recognition”, *Systems, Man and Cybernetics, IEEE Transactions on*, –5, 1983, pp. 826–834.
- [FS95] Freund, Y.; Schapire, R. E. “A decision-theoretic generalization of on-line learning and an application to boosting”. In: Computational learning theory, 1995, pp. 23–37.
- [GBT07] Gray, D.; Brennan, S.; Tao, H. “Evaluating appearance models for recognition, reacquisition, and tracking”. In: IEEE International Workshop on Performance Evaluation for Tracking and Surveillance, 2007, pp. 1 – 7.
- [GCLH14] Gong, S.; Cristani, M.; Loy, C. C.; Hospedales, T. M. “The re-identification challenge”. In: *Person Re-Identification*, Springer, 2014, pp. 1–20.
- [GH92] Gauch, J. M.; Hsia, C. W. “Comparison of three-color image segmentation algorithms in four color spaces”. In: Applications in optical science and engineering, 1992, pp. 1168–1181.
- [Goo12] Goodman, T. M. “International standards for colour”. In: *Colour Design - Theories and Applications*, Best, J. (Editor), Woodhead Publishing, 2012, pp. 177 – 218.
- [HAMJ02] Hsu, R.-L.; Abdel-Mottaleb, M.; Jain, A. K. “Face detection in color images”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24–5, 2002, pp. 696–706.

- [HBRB11] Hirzer, M.; Beleznai, C.; Roth, P. M.; Bischof, H. "Person re-identification by descriptive and discriminative classification". In: *Image Analysis*, Springer, 2011, pp. 91–102.
- [Heb49] Hebb, D. "The organization of behavior: A neuropsychological theory", 1949.
- [HS88] Harris, C.; Stephens, M. "A combined corner and edge detector." In: Alvey vision conference, 1988, pp. 50.
- [ITM01] Imai, F. H.; Tsumura, N.; Miyake, Y. "Perceptual color difference metric for complex images based on mahalanobis distance", *Journal of Electronic Imaging*, vol. 10–2, 2001, pp. 385–393.
- [JDN04] Jain, A. K.; Dass, S. C.; Nandakumar, K. "Can soft biometric traits assist user recognition?" In: Defense and Security, 2004, pp. 561–572.
- [JG09] Juan, L.; Gwon, O. "A comparison of sift, pca-sift and surf", *International Journal of Image Processing (IJIP)*, vol. 3–4, 2009, pp. 143–152.
- [JJDJ+10] Jacques Junior, J. C. S.; Dihl, L.; Jung, C.; Thielo, M.; Keshet, R.; Musse, S. "Human upper body identification from images". In: 17th IEEE International Conference on Image Processing, 2010, pp. 1717–1720.
- [JJJRM14] Jacques Junior, J. C. S.; Jung, C.; R.; Musse, S. "Head-shoulders human contour estimation in still images". In: 21th IEEE International Conference on Image Processing, 2014, pp. 278–282.
- [JJMJ10] Jacques Junior, J. C. S.; Musse, S. R.; Jung, C. R. "Crowd analysis using computer vision techniques", *Signal Processing Magazine, IEEE*, vol. 27–5, 2010, pp. 66–77.
- [Jol05] Jolliffe, I. "Principal component analysis". Wiley Online Library, 2005.
- [JZ08] Jia, H.-X.; Zhang, Y.-J. "Human detection in static images", *Pattern Recog. Tech. App.: Recent Advances*, vol. 1, 2008, pp. 227–243.
- [KGT05] Kong, D.; Gray, D.; Tao, H. "Counting pedestrians in crowds using viewpoint invariant training." In: BMVC, 2005.
- [KJK02] Kim, K. I.; Jung, K.; Kim, H. J. "Face recognition using kernel principal component analysis", *Signal Processing Letters, IEEE*, vol. 9–2, 2002, pp. 40–42.
- [KS04] Ke, Y.; Sukthankar, R. "Pca-sift: A more distinctive representation for local image descriptors". In: Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, 2004, pp. II–506.

- [LGL14] Liu, C.; Gong, S.; Loy, C. C. “On-the-fly feature importance mining for person re-identification”, *Pattern Recognition*, vol. 47–4, 2014, pp. 1602–1615.
- [LHG12] Layne, R.; Hospedales, T.; Gong, S. “Person re-identification by attributes”. In: *Proceedings of the British Machine Vision Conference*, 2012, pp. 24.1–24.11.
- [Lin98] Lindeberg, T. “Feature detection with automatic scale selection”, *International journal of computer vision*, vol. 30–2, 1998, pp. 79–116.
- [LM01] Lucchese, L.; Mitray, S. “Color image segmentation: A state-of-the-art survey”, *Proceedings of the Indian National Science Academy (INSA-A). Delhi, Indian: Natl Sci Acad*, vol. 67–2, 2001, pp. 207–221.
- [LM02] Lienhart, R.; Maydt, J. “An extended set of haar-like features for rapid object detection”. In: *Image Processing. 2002. Proceedings. 2002 International Conference on*, 2002, pp. 1–900.
- [Low99] Lowe, D. G. “Object recognition from local scale-invariant features”. In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*, 1999, pp. 1150–1157.
- [LS05] Le, D.-D.; Sato, S. “An efficient feature selection method for object detection”. In: *Pattern Recognition and Data Mining*, Springer, 2005, pp. 461–468.
- [LW13] Li, W.; Wang, X. “Locally aligned feature transforms across views”. In: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 3594–3601.
- [LZA11] Lukashevich, P.; Zalesky, B.; Ablameyko, S. “Medical image registration based on surf detector”, *Pattern Recognition and Image Analysis*, vol. 21–3, 2011, pp. 519–521.
- [LZLM07] Liu, Y.; Zhang, D.; Lu, G.; Ma, W.-Y. “A survey of content-based image retrieval with high-level semantics”, *Pattern Recognition*, vol. 40–1, 2007, pp. 262–282.
- [LZW13] Li, W.; Zhao, R.; Wang, X. “Human reidentification with transferred metric learning”. In: *Computer Vision–ACCV 2012*, Springer, 2013, pp. 31–44.
- [Mah36] Mahalanobis, P. C. “On the generalized distance in statistics”, *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, 1936, pp. 49–55.
- [MG06] Munder, S.; Gavrilu, D. M. “An experimental study on pedestrian classification”, *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28–11, 2006, pp. 1863–1868.

- [MHB<sup>+</sup>10] Mair, E.; Hager, G. D.; Burschka, D.; Suppa, M.; Hirzinger, G. "Adaptive and generic corner detection based on the accelerated segment test". In: *Computer Vision—ECCV 2010*, Springer, 2010, pp. 183–196.
- [MP43] McCulloch, W. S.; Pitts, W. "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, vol. 5–4, 1943, pp. 115–133.
- [MPP01] Mohan, A.; Papageorgiou, C.; Poggio, T. "Example-based object detection in images by components", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 23–4, 2001, pp. 349–361.
- [MS02] Mikolajczyk, K.; Schmid, C. "An affine invariant interest point detector". In: *Computer Vision - ECCV 2002*, Springer, 2002, pp. 128–142.
- [MS04] Mikolajczyk, K.; Schmid, C. "Scale and affine invariant interest point detectors", *International journal of computer vision*, vol. 60–1, 2004, pp. 63–86.
- [MS05] Mikolajczyk, K.; Schmid, C. "A performance evaluation of local descriptors", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27–10, 2005, pp. 1615–1630.
- [MTC12] Mazzon, R.; Tahir, S. F.; Cavallaro, A. "Person re-identification in crowd", *Pattern Recognition Letters*, vol. 33–14, Oct 2012, pp. 1828–1837.
- [NJ01] Ng, A. Y.; Jordan, M. I. "On discriminative vs. generative classifiers: A comparison of logistic regression and naive bayes", 2001.
- [OPH96] Ojala, T.; Pietikäinen, M.; Harwood, D. "A comparative study of texture measures with classification based on featured distributions", *Pattern recognition*, vol. 29–1, 1996, pp. 51–59.
- [OPM02] Ojala, T.; Pietikainen, M.; Maenpaa, T. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24–7, 2002, pp. 971–987.
- [OPS<sup>+</sup>97] Oren, M.; Papageorgiou, C.; Sinha, P.; Osuna, E.; Poggio, T. "Pedestrian detection using wavelet templates". In: *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, 1997, pp. 193–199.
- [PFA06] Pennec, X.; Fillard, P.; Ayache, N. "A riemannian framework for tensor computing", *International Journal of Computer Vision*, vol. 66–1, 2006, pp. 41–66.

- [POP98] Papageorgiou, C. P.; Oren, M.; Poggio, T. "A general framework for object detection". In: *Computer Vision, 1998. Sixth International Conference on, 1998*, pp. 555–562.
- [Por05] Porikli, F. "Integral histogram: A fast way to extract histograms in cartesian spaces". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, 2005*, pp. 829–836.
- [PP00] Papageorgiou, C.; Poggio, T. "A trainable system for object detection", *International Journal of Computer Vision*, vol. 38–1, 2000, pp. 15–33.
- [PZG<sup>+</sup>10] Prosser, B.; Zheng, W.-S.; Gong, S.; Xiang, T.; Mary, Q. "Person re-identification by support vector ranking." In: *BMVC, 2010*, pp. 5.
- [Qui86] Quinlan, J. R. "Induction of decision trees", *Machine learning*, vol. 1–1, 1986, pp. 81–106.
- [RBK98] Rowley, H. A.; Baluja, S.; Kanade, T. "Neural network-based face detection", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20–1, 1998, pp. 23–38.
- [RD06] Rosten, E.; Drummond, T. "Machine learning for high-speed corner detection". In: *Computer Vision—ECCV 2006*, Springer, 2006, pp. 430–443.
- [RPD10] Rosten, E.; Porter, R.; Drummond, T. "Faster and better: A machine learning approach to corner detection", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 32–1, 2010, pp. 105–119.
- [Sch90] Schapire, R. E. "The strength of weak learnability", *Machine learning*, vol. 5–2, 1990, pp. 197–227.
- [Sch12] Schwartz, W. R. "Scalable people re-identification based on a one-against-some classification scheme". In: *Image Processing (ICIP), 2012 19th IEEE International Conference on, 2012*, pp. 1613–1616.
- [SD09] Schwartz, W. R.; Davis, L. S. "Learning discriminative appearance-based models using partial least squares". In: *Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on, 2009*, pp. 322–329.
- [SGM09] Shan, C.; Gong, S.; McOwan, P. W. "Facial expression recognition based on local binary patterns: A comprehensive study", *Image and Vision Computing*, vol. 27–6, 2009, pp. 803–816.
- [SK87] Sirovich, L.; Kirby, M. "Low-dimensional procedure for the characterization of human faces", *JOSA A*, vol. 4–3, 1987, pp. 519–524.

- [SKHD09] Schwartz, W. R.; Kembhavi, A.; Harwood, D.; Davis, L. S. "Human detection using partial least squares analysis". In: *Computer vision, 2009 IEEE 12th international conference on*, 2009, pp. 24–31.
- [SKV<sup>+</sup>94] Skarbek, W.; Koschan, A.; Veroffentlichung, Z.; et al.. "Colour image segmentation-a survey", 1994.
- [SM97] Schmid, C.; Mohr, R. "Local grayvalue invariants for image retrieval", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19–5, 1997, pp. 530–535.
- [SP98] Sung, K.-K.; Poggio, T. "Example-based learning for view-based human face detection", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20–1, 1998, pp. 39–51.
- [SRBB06] Suard, F.; Rakotomamonjy, A.; Bensrhair, A.; Broggi, A. "Pedestrian detection using infrared images and histograms of oriented gradients". In: *Intelligent Vehicles Symposium, 2006 IEEE*, 2006, pp. 206–212.
- [SV14] Sobral, A.; Vacavant, A. "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos", *Computer Vision and Image Understanding*, vol. 122–0, 2014, pp. 4 – 21.
- [TCL13] Trzcinski, T.; Christoudias, C. M.; Lepetit, V. "Learning Image Descriptors with Boosting", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.
- [TP91] Turk, M.; Pentland, A. "Eigenfaces for recognition", *Journal of cognitive neuroscience*, vol. 3–1, 1991, pp. 71–86.
- [TT03] Tkalcic, M.; Tasic, J. F. "Colour spaces: perceptual, historical and applicational background". *IEEE*, 2003, vol. 1.
- [VBC13] Vezzani, R.; Baltieri, D.; Cucchiara, R. "People reidentification in surveillance and forensics: A survey", *ACM Computing Surveys*, vol. 46–2, Dez 2013, pp. 29:1–29:37.
- [VJ01] Viola, P.; Jones, M. "Rapid object detection using a boosted cascade of simple features". In: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, pp. 1–511.
- [VJ04] Viola, P.; Jones, M. J. "Robust real-time face detection", *International journal of computer vision*, vol. 57–2, 2004, pp. 137–154.

- [VS12] Vantaram, S. R.; Saber, E. "Survey of contemporary trends in color image segmentation", *Journal of Electronic Imaging*, vol. 21–4, 2012, pp. 040901–1.
- [WCC05] Wang, Y.-F.; Chang, E. Y.; Cheng, K. P. "A video analysis framework for soft biometry security surveillance". In: Proceedings of the third ACM international workshop on Video surveillance & sensor networks, 2005, pp. 71–78.
- [WGDD12] Wang, R.; Guo, H.; Davis, L. S.; Dai, Q. "Covariance discriminative learning: A natural and efficient approach to image set classification". In: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, 2012, pp. 2496–2503.
- [WHY09] Wang, X.; Han, T. X.; Yan, S. "An hog-lbp human detector with partial occlusion handling". In: Computer Vision, 2009 IEEE 12th International Conference on, 2009, pp. 32–39.
- [WMZ<sup>+</sup>14] Wei, W.; Ma, H.; Zhang, H.; Gao, Y.; Wang, Z. "Person re-identification based on human body parts signature". In: Proceedings of the International Conference on Distributed Smart Cameras, 2014, pp. 9.
- [WZM13] Wang, S.; Zhang, J.; Miao, Z. "A new edge feature for head-shoulder detection". In: 20th IEEE International Conference on Image Processing, 2013, pp. 2822–2826.
- [XACT11] Xin, H.; Ai, H.; Chao, H.; Tretter, D. "Human head-shoulder segmentation". In: IEEE International Conference on Automatic Face Gesture Recognition and Workshops, 2011, pp. 227–232.
- [YC12] Yang, X.; Cheng, K.-T. T. "Accelerating surf detector on mobile devices". In: Proceedings of the 20th ACM international conference on Multimedia, 2012, pp. 569–578.
- [YZFY04] Yang, J.; Zhang, D.; Frangi, A. F.; Yang, J.-y. "Two-dimensional pca: a new approach to appearance-based face representation and recognition", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 26–1, 2004, pp. 131–137.
- [ZBMM06] Zhang, H.; Berg, A. C.; Maire, M.; Malik, J. "Svm-knn: Discriminative nearest neighbor classification for visual category recognition". In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2006, pp. 2126–2136.
- [ZCPR03] Zhao, W.; Chellappa, R.; Phillips, P. J.; Rosenfeld, A. "Face recognition: A literature survey", *Acm Computing Surveys (CSUR)*, vol. 35–4, 2003, pp. 399–458.

- [ZDFL95] Zhang, Z.; Deriche, R.; Faugeras, O.; Luong, Q.-T. "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", *Artificial intelligence*, vol. 78–1, 1995, pp. 87–119.
- [ZE07] Zickler, S.; Efros, A. "Detection of multiple deformable objects using pca-sift." In: Proceedings of the National Conference on Artificial Intelligence, 2007, pp. 1127.
- [ZGX09] Zheng, W.-S.; Gong, S.; Xiang, T. "Associating groups of people". In: Proceedings of the British Machine Vision Conference, 2009, pp. 23.1–23.11.
- [ZGX13] Zheng, W.-S.; Gong, S.; Xiang, T. "Reidentification by relative distance comparison", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35–3, 2013, pp. 653–668.
- [ZGX14] Zheng, W.-S.; Gong, S.; Xiang, T. "Group association: Assisting re-identification by visual context". In: *Person Re-Identification*, Springer, 2014, pp. 183–201.
- [ZMR+08] Zhan, B.; Monekosso, D. N.; Remagnino, P.; Velastin, S. A.; Xu, L.-Q. "Crowd analysis: a survey", *Machine Vision and Applications*, vol. 19–5-6, 2008, pp. 345–357.
- [ZOW13a] Zhao, R.; Ouyang, W.; Wang, X. "Person re-identification by salience matching". In: IEEE International Conference on Computer Vision, 2013, pp. 2528–2535.
- [ZOW13b] Zhao, R.; Ouyang, W.; Wang, X. "Unsupervised salience learning for person re-identification". In: IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp. 3586–3593.
- [ZQJ+14] Zhou, T.; Qi, M.; Jiang, J.; Wang, X.; Hao, S.; Jin, Y. "Person re-identification based on nonlinear ranking with difference vectors", *Information Sciences*, vol. 279, 2014, pp. 604–614.
- [ZYCA06] Zhu, Q.; Yeh, M.-C.; Cheng, K.-T.; Avidan, S. "Fast human detection using a cascade of histograms of oriented gradients". In: Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on, 2006, pp. 1491–1498.



## APÊNDICE A – PROCESSO DE MONTAGEM DOS BANCOS DE IMAGENS E SELEÇÃO DE ENTRADAS

Este apêndice orienta o leitor na aquisição dos bancos de imagens utilizados durante o trabalho. Onde foi necessário um processo adicional para geração dos subconjuntos avaliados, o passo a passo é aqui explicado; nos demais casos, o apêndice apresenta os *links* diretos para os bancos e listagens utilizadas.

### A.1 Definição das imagens utilizadas banco VIPeR

O banco de imagens VIPeR, apresentado em 2007 no trabalho de Gray e equipe [GBT07], contém 632 pares de imagens numerados de pedestres sob diferentes ângulos de visão e condições de iluminação. As imagens são de baixa resolução (128x48 *pixels*), recortadas ao redor do pedestre e capturadas por duas câmeras em diferente posições.

A escolha deste banco foi feita com intuito de comparação com demais trabalhos. O subconjunto do VIPeR utilizado neste trabalho consiste, então, em 50% das imagens selecionadas aleatoriamente, as mesmas para a câmera A e câmera B. Nenhuma modificação adicional no banco foi efetuada.

Página do banco de imagens: <http://vision.soe.ucsc.edu/node/178>

Lista com os identificadores das imagens selecionadas: [http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/index\\_viper\\_316.txt](http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/index_viper_316.txt)

### A.2 Definição do subconjunto de imagens no banco ETHZ

O banco de imagens ETHZ, inicialmente criado para o trabalho de Ess e equipe [ELVG07], foi incrementalmente construído ([ELS<sup>+</sup>08]) com vídeos capturados (a 13 ou 14 quadros por segundo) por duas câmeras pareadas e em movimento, em diferentes ambientes externos. As imagens são de satisfatória resolução (640x480 *pixels*) e, na maioria dos casos, possuem informações anotadas sobre calibração de câmeras e pedestres detectados - todavia, nenhuma destas anotações foi utilizada no presente trabalho, uma vez que o subconjunto foi montado para ter informação contextual de grupos dentre diversas sequências do banco.

A escolha deste banco foi motivada pela presença de cenas completas, onde várias pessoas estão presentes no mesmo quadro podendo formar os grupos de indivíduos

que são avaliados. Uma vez que as sequências de cenas/quadros de cada vídeo são disponibilizadas integralmente no banco de imagens, optou-se por gerar o subconjunto a partir de quadros esparsos selecionados manualmente. As sequências utilizadas foram: Sequência #0 ([ELVG07]); Sequências BAHNHOF, JELMOLI, SUNNY DAY, LINTHESCHER, CROSSING, PEDCROSS e LOEWENPLATZ ([ELS+08]).

O processo de geração do subconjunto utilizado foi focado em grupos e incrementalmente construído da seguinte forma: i) na câmera B (conjunto de busca), 90 quadros distintos e esparsos onde visivelmente ocorria a formação de grupos foram separados; ii) uma lista de pessoas detectadas (utilizando o detector de pessoas baseado em HOG [DT05]) nos quadros antes selecionados foi gerada; iii) falsos positivos e pessoas que não apareciam no mesmo quadro na câmera A foram removidas da lista; iv) grupos foram detectados nos quadros separados da câmera B; v) quadros que não continham detecção de grupos foram removidos e, os restantes, definiram a lista resultante. Note que, na câmera A, não há detecção de pessoas. A restrição estabelecida na geração do subconjunto define que as pessoas detectadas em B deveriam aparecer, sem a necessidade de serem detectadas, na câmera A (conjunto de entrada) para permitir que o usuário selecionasse as cores a serem procuradas no conjunto de imagens de busca.

O resultado final consiste em um subconjunto totalizando 141 grupos detectados, contendo 213 indivíduos (alguns compartilhando mais de um grupo) em 72 cenas para cada câmera. Nesta configuração, pessoas visualmente repetidas e detectadas em cenas esparsas inicialmente receberam identificadores diferentes, vindo a serem relacionados através das listas de equivalências.

Página do banco completo para a Sequência #0 ([ELVG07]): <http://www.vision.ee.ethz.ch/~aess/iccv2007/>

Página do banco completo para a Sequências BAHNHOF, JELMOLI, SUNNY DAY, LINTHESCHER, CROSSING, PEDCROSS e LOEWENPLATZ ([ELS+08]): <http://www.vision.ee.ethz.ch/~aess/dataset/>

Lista com os identificadores das imagens e indivíduos selecionados e regiões das detecções: <http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/TargetPeopleList.txt>

Lista com os identificadores dos grupos detectados: [http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/groups\\_distinct.txt](http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/groups_distinct.txt)

### **A.3 Construção das listas de equivalências**

Para realizar a rotulação manual, a qual permite que a mesma pessoa pode ser visualizada por diferentes câmeras (ou na mesma câmera após significativo espaço de tempo), as imagens resultantes da Seção A.2 foram analisadas em busca de indivíduos

e grupos que aparecessem repetidamente em diferentes cenas, estabelecendo as *listas de equivalências*.

As *listas de equivalências* indicam em quais outros quadros o mesmo indivíduo ou grupo aparece, relacionando seus identificadores (IDs). Dentre os 213 indivíduos no conjunto de busca, 29 indivíduos aparecem pelo menos uma vez mais em outra cena. Destes 29, 2 indivíduos compartilham 6 IDs, 1 indivíduo compartilha 5 IDs, 7 compartilham 3 IDs e 19 indivíduos compartilham 2 IDs. Dentre os 141 grupos, a lista contém 11 equivalentes, onde 9 compartilham 2 IDs, 1 compartilha 3 IDs e 1 compartilha 6 IDs.

A Figura A.1 exemplifica uma equivalência de indivíduos (IDs em verde) e de grupos (IDs em azul) em três cenas do subconjunto utilizado.

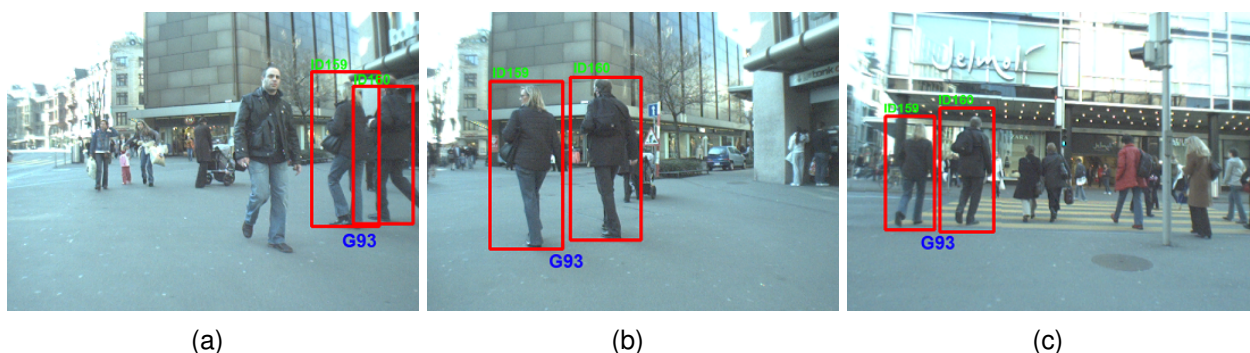


Figura A.1 – As *listas de equivalências*. A mulher (no *bounding-box* esquerdo) e o homem (no *bounding-box* direito) compartilham o mesmo ID (159 e 160, respectivamente) em três cenas/quadros (a-c). O casal também foi detectado como um grupo nestes 3 quadros (a-c), compartilhando do mesmo ID (93).

*Lista de equivalências para pessoas:* [http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/11/equivalence\\_list.txt](http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/11/equivalence_list.txt)

*Lista de equivalências para grupos:* [http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/equivalence\\_list\\_groups.txt](http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/equivalence_list_groups.txt)

#### A.4 Modelos de cores selecionados pelo usuário

Sendo a entrada deste modelo uma etapa manual de interação com o usuário, todas as regiões por ele selecionadas para montagem dos modelos de cores - utilizados nas buscas durante os cenários - foram armazenadas para permitir futuras comparações.

Para cada indivíduo a ser buscado, o usuário pôde escolher até 3 cores para cada atributo (*tronco*, *pernas* e *cabeça*), sendo obrigatória a seleção de ao menos uma cor para o atributo *tronco* e uma para o *pernas*. Através da cor média de cada seleção, os atributos foram semanticamente organizados em um modelo de corpo 2D para realização a re-identificação dos indivíduos.

*Lista seleções para o banco VIPeR:* [http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/Viper\\_Selections.zip](http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/Viper_Selections.zip)

*Lista seleções para o banco ETHZ:* [http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/ETHZ\\_Selections.zip](http://www.cpva.pucrs.br/wp/wp-content/uploads/2014/10/ETHZ_Selections.zip)

## **APÊNDICE B – LISTA DE PUBLICAÇÕES OBTIDAS E SUBMETIDAS**

Este apêndice lista as publicações obtidas e sob revisão resultantes de trabalhos desenvolvido durante o período do mestrado.

### **B.1 Artigos publicados**

Salamon, N. Z.; Jacques Junior, J. C. S.; Musse, S. R.; “Seeing the Movement through Sound: Giving Trajectory Information to Visually Impaired People” Games and Digital Entertainment (SBGAMES), 2014 XIII Brazilian Symposium on. 2014

### **B.2 Artigos submetidos e sob revisão**

Salamon, N. Z.; Jacques Junior, J. C. S.; Musse, S. R.; “People re-identification in still images through color and group belonging features”. Submetido para Expert Systems with Applications, Elsevier.