

**TÉCNICAS OTIMIZADAS PARA  
REIDENTIFICAÇÃO DE PESSOAS**



CRISTIANNE RODRIGUES SANTOS DUTRA

**TÉCNICAS OTIMIZADAS PARA  
REIDENTIFICAÇÃO DE PESSOAS**

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

ORIENTADOR: WILLIAM ROBSON SCHWARTZ

Belo Horizonte

Julho de 2016

**Ficha catalográfica elaborada pela Biblioteca do ICEx - UFMG**

Dutra, Cristianne Rodrigues Santos.

D978t Técnicas otimizadas para reidentificação de pessoas. /  
Cristianne Rodrigues Santos. – Belo Horizonte, 2016.  
xxvi, 70 f.: il.; 29 cm.

Dissertação (mestrado) - Universidade Federal de  
Minas Gerais – Departamento de Ciência da Computação.

Orientador: William Robson Schwartz.

1. Computação – Teses. 2. Visão por computador. 3.  
Sistemas de reconhecimento de padrões. 4. Processamento  
de imagens. I. Orientador. II. Título.

CDU 519.6\*82.10 (043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

## FOLHA DE APROVAÇÃO

Técnicas otimizadas para reidentificação de pessoas

**CRISTIANNE RODRIGUES SANTOS DUTRA**

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

*William Robson De Hartz*

PROF. WILLIAM ROBSON SCHWARTZ - Orientador  
Departamento de Ciência da Computação - UFMG

*Jefferson Alex dos Santos*

PROF. JEFFERSON ALEX DOS SANTOS  
Departamento de Ciência da Computação - UFMG

*João Paulo Papa*

PROF. JOÃO PAULO PAPA  
Departamento de Computação - UNESP

Belo Horizonte, 12 de julho de 2016.



*A Jesus, o Cristo, quem nos salvou de nossos pecados e prometeu voltar como Rei dos reis e Senhor dos senhores.*



# Agradecimentos

A cada vitória o reconhecimento devido ao meu Deus, pois só Ele é digno de toda honra, glória e louvor! Jesus, obrigada pelo fim de mais essa etapa.

Ao professor William Robson Schwartz pela orientação notável.

Ao Marcos, pessoa com quem amo partilhar a vida, que soube exercitar o dom da paciência e se fez fonte de incentivos em nome do nosso amor.

Ao meu filho Rafael que a cada dia enche minha vida de amor e alegria.

À minha família que com renúncia, amor e dedicação me deram força e toda ajuda necessária para o alcance de meus objetivos.

Aos membros da equipe SSIG que sempre me ajudaram no que foi possível.

Aos amigos queridos, de perto e de longe, a minha eterna gratidão.

À Fundação de Pesquisa de Minas Gerais - FAPEMIG (Projetos PPM-00025-15 e APQ-00567-14), à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - CAPES (Projeto DeepEyes) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq (Projeto 477457/2013-4).

Sem todos vocês nada disso seria possível.



*“Quanto mais vivo, mais profundamente me convenço de que o que faz a diferença entre o homem e outro homem é a energia, uma determinação invencível, uma decisão tomada e mantida até a vitória final.”*

*(Fowell Buxton)*



# Resumo

A importância da reidentificação de pessoas tem aumentado nos últimos anos, principalmente devido às grandes redes de câmeras utilizadas em sistemas de vigilância. O objetivo é o de reidentificar os indivíduos que tenham sido previamente identificados em uma câmera distinta em um tempo anterior. Muitas abordagens têm sido propostas para reduzir o custo computacional e aumentar a acurácia durante a busca destes indivíduos na galeria. Porém, ainda existem desafios a serem enfrentados e solucionados, tais como mudanças de iluminação, variação de pose, oclusões, baixa qualidade no processo de aquisição da imagem, modelagem da aparência e a gestão de um crescente número de indivíduos que estão sendo monitorados pelo sistema de vigilância. O presente trabalho aborda o último problema por meio da criação de uma estrutura de indexação baseada em listas invertidas, trazendo para as primeiras posições os candidatos com maior probabilidade de serem o indivíduo alvo. Duas abordagens foram propostas. A primeira faz uso de dicionários visuais que são aprendidos para armazenar as características das amostras e permitir a indexação dessas características nas listas invertidas. A segunda utiliza o método Partial Least Squares para reduzir a ambiguidade entre as diversas aparências. Os métodos são avaliados nas bases de dados VIPeR e CUHK01. Os resultados mostram redução significativa da lista de candidatos, por exemplo, utilizando apenas 25% dos indivíduos da base de dados VIPeR, obtém-se resultados melhores que aqueles alcançados no estado da arte para o reconhecimento rank-1.



# Abstract

The importance of peoples re-identification has increased recently, mainly due to the large camera networks used in surveillance systems. The goal is to re-identify individuals who have been previously identified in a different camera at an earlier time. Many approaches have been proposed to reduce the computational cost and increase the accuracy during the search of these individuals in the gallery. However there still challenges to be faced and solved, such as lighting changes, pose variation, occlusions, low quality of the image acquisition process, appearance modeling and management of a increasing number of individuals who are being monitored by the surveillance system. This thesis addresses the latter problem by creating an index structure based on inverted index to bring to the first positions the candidates most likely to be the target individual. Two approaches have been proposed: the first one is based on visual dictionaries which are learned to store the features of the samples and allow indexing of these features in the inverted index. The second approach uses the Partial Least Squares method for reducing the ambiguity among different appearances. The methods are evaluated by VIPeR CUHK01 databases. The results show a significant reduction in the list of candidates, for example, our method obtains better results than those achieved by the state-of-the-art methods rank-1 recognition rate using only 25 % of the individuals VIPeR database.



# Lista de Figuras

1.1	Rede de câmeras composta por quatro câmeras com campos de visão não sobrepostos [Gilbert & Bowden, 2006]. . . . .	1
1.2	Ambiente monitorado por duas câmeras com campos de visão não sobrepostos. O objetivo da reidentificação de pessoas é determinar se uma pessoa atualmente visível na câmera A já foi observada pela câmera B [Fonte: Google imagens]. . . . .	2
1.3	Observação de Pessoas e aplicações [Schwartz, 2012a]. . . . .	4
1.4	Exemplo de imagens da base VIPeR [Gray et al., 2007], onde ocorrem diferenças da aparência do mesmo indivíduo ao longo do tempo. Note que as imagens da mesma coluna pertencem ao mesmo indivíduo. . . . .	6
2.1	Espaço de cor RGB e efeito do filtro de predominância com $T = 0.1, 0.3$ e $0.5$ , respectivamente. . . . .	13
3.1	Etapas do processo de reidentificação de pessoas. . . . .	21
3.2	Etapas do processo de reidentificação de pessoas. . . . .	22
3.3	Aprendizado do dicionário. Primeiramente as amostras dos indivíduos são divididas em $m$ regiões horizontais não sobrepostas (1), em seguida é realizada a extração de características (2). Por último, é criado o dicionário selecionando aleatoriamente $k$ vetores de características como palavras visuais (3). . . . .	25
3.4	Abordagens para a estrutura de indexação baseado em dicionários: dicionário global único, dicionário local para cada região e múltiplos dicionários globais. . . . .	26

3.5	Na etapa de treinamento, os vetores de características das amostras da galeria são comparados ao dicionário visual (1), o índice da palavra visual mais próxima é utilizado para indexar a lista invertida, onde será armazenado o identificador da amostra analisada (2). No teste, a amostra de teste é particionada em regiões não sobrepostas (3), em seguida é feito a extração de características (4). Os vetores de características da amostra de teste são comparados ao dicionário visual (5), o índice da palavra visual que mais se assemelha ao vetor de característica é utilizado para indexar a lista invertida (6) e recuperar os identificadores das amostras da galeria que possuem maior similaridade com a amostra de teste (7).	27
3.6	Aprendizado dos dicionários utilizando o PLS-OAA com dicionários visuais.	31
3.7	Etapas de treinamento e teste do processo de reidentificação usando PLS-OAA com dicionários visuais.	32
3.8	Estratégia para gerar uma única lista global a partir dos candidatos locais, considerando a quantidade de indivíduos igual a 10.	32
3.9	Abordagem PLS-OAA usando permutação de indivíduos.	33
4.1	Amostras dos conjuntos de dados: (a) VIPeR e (b) CUHK01. As amostras que estão na mesma linha pertencem ao mesmo indivíduo.	36
4.2	Taxa de reconhecimento alcançada pela abordagem baseada em dicionários visuais como uma função do número de palavras visuais do dicionário.	38
4.3	Avaliação do número de dicionários globais da abordagem múltiplos dicionários.	39
4.4	Gráfico de barras MARR e curva CMC mostrando os resultados obtidos pela abordagem baseada em dicionários visuais utilizando vários descritores de características.	40
4.5	Gráfico de barras MARR e curva CMC mostrando os resultados alcançados pelos métodos de ponderação.	41
4.6	Avaliação do número de dimensões do PLS-OAA usando dicionário visual.	42
4.7	Avaliação do número de palavras visuais do dicionário utilizando PLS-OAA usando dicionário visual.	43
4.8	Avaliação do número necessário de dicionários utilizando o PLS-OAA usando dicionário visual.	43
4.9	Gráfico de barras MARR e curva CMC mostrando os resultados alcançados pelo PLS-OAA usando dicionários variando os descritores de características utilizados.	44

4.10	Curvas MARR e CMC mostrando os resultados alcançados pelo PLS-OAA usando dicionários para diferentes tamanhos da lista de candidatos recuperada. . . . .	45
4.11	Avaliação do número de fatores utilizando o PLS-OAA usando permutação de indivíduos. . . . .	46
4.12	Avaliação do número de modelos utilizando o PLS-OAA usando permutação de indivíduos. . . . .	47
4.13	Avaliação do número de subconjuntos utilizando o PLS-OAA usando permutação de indivíduos. . . . .	47
4.14	Curvas MARR e CMC obtidas utilizando o PLS-OAA usando permutação de indivíduos considerando diferentes porcentagens da lista de candidatos. . . . .	48
4.15	Curvas MARR e CMC mostrando os resultados alcançados pelo PLS-OAA usando permutação de indivíduos na base de dados CUHK01 para diferentes tamanhos da lista de candidatos recuperada. . . . .	50
5.1	Média MARR e desvio padrão considerando 10 execuções das abordagens: (a) Indexação baseado dicionário visual, (b) Indexação usando PLS-OAA com dicionário visual e (c) Indexação usando PLS-OAA com permutação de indivíduos. . . . .	53



# Lista de Tabelas

5.1	Custo computacional de cada abordagem de acordo com o número de operações relevantes necessárias. . . . .	51
5.2	Resultados obtidos nos principais ranks utilizando todas as abordagens e diferentes tamanhos da lista de candidatos. . . . .	52
5.3	Tabela CMC com os principais resultados do estado da arte para a base de dados VIPeR (%). A tabela foi ordenada considerando a taxa de reconhecimento obtida no rank=1. Tabela disponível em <a href="http://www.ssig.dcc.ufmg.br/reid-results/">http://www.ssig.dcc.ufmg.br/reid-results/</a> . . . . .	54
5.4	Resultado CUHK01 . . . . .	56
5.5	Tabela CMC com os principais resultados do estado da arte para a base de dados CUHK01 (%). Tabela disponível em <a href="http://www.ssig.dcc.ufmg.br/reid-results/">http://www.ssig.dcc.ufmg.br/reid-results/</a> . . . . .	56



# Lista de Abreviações

- CMC** Cumulative Match Characteristic
- HOG** Histograms of Oriented Gradient
- KissMe** Keep It Simple and Straightforwar Metric
- SCNCD** Salient Color Names based Color Descriptor
- NIPALS** Nonlinear Iterative PLS
- PLS** Partial Least Squares
- PLS-OAA** Partial Least Squares um-contra-todos
- PLSH** Partial Least Squares for Face Hashing
- ePLSH** extended PLS for face hashing
- MARR** Maximum Achievable Recognition Rate
- SDALF** Symmetry-Driven Accumulation of Local Features
- MSCR** Maximally Stable Color Regions
- RHSP** Recurrent High-Structured Patches
- CPS** Custom Pictorial Structure
- AUC** Area Under the Curve
- FPNN** Filter Pairing Neural Network
- k-NN** k-Nearest Neighbors
- PS** Pictorial Structures
- SVM** Support Vector Machines

**BRO** Block-Based Ratio-Occurrence

**CRRRO** Center Rectangular Ring Ratio-Occurrence

**CPDL** Cross-view Projective Dictionary Learning

**SIFT** Scale-Invariant Feature Transform

**WISE** Visual Search Engine

**rgb** RGB normalizado

**HS** Hue-Saturation

**WHOS** Weighted Histograms of Overlapping Stripes

**LOMO** Local Maximal Occurrence

**PRDC** Probabilistic Relative Distance Comparison

**LFDA** Local Fisher Discriminant Analysis

**LMNN** Large Margin Nearest Neighbor

**XQDA** Cross-view Quadratic Discriminant Analysis

**X-CRC** Cross-view Collaborative Representation based Classification

**CRC** Collaborative Representation based Classification

**PCA** Principal Component Analysis

# Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xxi
Lista de Abreviações	xxiii
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	4
1.2 Problemática . . . . .	5
1.3 Objetivos e Contribuições . . . . .	6
1.4 Organização . . . . .	7
<b>2 Trabalhos Relacionados</b>	<b>9</b>
2.1 Reidentificação de Pessoas . . . . .	9
2.1.1 Modelagem de Aparência . . . . .	10
2.1.2 Descritores de características . . . . .	12
2.1.3 Técnicas baseadas em Aprendizagem . . . . .	14
2.2 Indexação de Dados . . . . .	17
<b>3 Metodologia</b>	<b>21</b>
3.1 Descritores de características . . . . .	22
3.2 Indexação de Dados . . . . .	23
3.2.1 Indexação baseada em Dicionários Visuais . . . . .	24
3.2.1.1 Aprendizado do Dicionário . . . . .	24

3.2.1.2	Treinamento . . . . .	25
3.2.1.3	Teste . . . . .	27
3.2.2	Indexação baseada em Partial Least Squares . . . . .	28
3.2.2.1	Regressão PLS . . . . .	29
3.2.2.2	PLS-OAA usando dicionário visual . . . . .	30
3.2.2.3	PLS-OAA usando permutação de indivíduos . . . . .	31
3.3	Método de Reidentificação . . . . .	34
<b>4</b>	<b>Resultados Experimentais</b>	<b>35</b>
4.1	Configuração Experimental . . . . .	35
4.1.1	Base de dados . . . . .	36
4.1.2	Medidas de avaliação . . . . .	37
4.2	Indexação baseada em Dicionários Visuais . . . . .	37
4.3	Indexação baseada em Partial Least Squares . . . . .	42
4.3.1	PLS-OAA usando dicionário visual . . . . .	42
4.3.2	PLS-OAA usando permutação de indivíduos . . . . .	46
4.4	Resultados utilizando a base de dados CUHK01 . . . . .	49
<b>5</b>	<b>Análise dos Resultados</b>	<b>51</b>
5.1	Análise do custo computacional . . . . .	51
5.2	Estabilidade dos resultados . . . . .	52
5.3	Comparação dos métodos com o estado da arte . . . . .	52
5.4	Resultados sobre a base de dados CUHK01 . . . . .	55
<b>6</b>	<b>Conclusão e Trabalhos Futuros</b>	<b>57</b>
	<b>Referências Bibliográficas</b>	<b>59</b>
	<b>Apêndice A Multi-Task Cross-View Collaborative Representation based Classification</b>	<b>69</b>

# Capítulo 1

## Introdução

A crescente preocupação com segurança tem tornado as câmeras de vigilância onipresentes em locais públicos e privados, proporcionando uma grande quantidade de dados visuais. A análise destes dados, em geral, é realizada por meio de um operador humano após a ocorrência de incidentes. Porém, esta análise está sujeita a erros, devidos principalmente à vulnerabilidade a falhas de atenção, pois em geral um único operador é responsável pelo monitoramento simultâneo de várias câmeras. Além disso, o desempenho do monitoramento está intimamente relacionado à experiência adquirida pelo operador ao longo de sua carreira profissional, o que torna o monitoramento caro e impreciso [Keval, 2006].

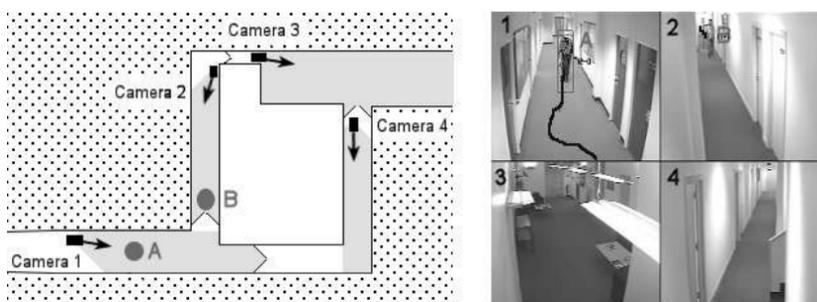


Figura 1.1: Rede de câmeras composta por quatro câmeras com campos de visão não sobrepostos [Gilbert & Bowden, 2006].

Em vigilância, as redes de câmeras (figura 1.1) são geralmente esparsas, ou seja, uma grande área é monitorada pelo menor número possível de câmeras e a maioria das câmeras possuem campos de visão não sobrepostos, o que torna a vigilância uma tarefa difícil pela falta de conhecimento sobre a topologia da rede e a dificuldade de manter a identidade dos indivíduos rastreados devido a mudanças da aparência relacionadas às variações de iluminação, mudanças de pose e parâmetros das câme-

ras [Song et al., 2013].

A automação dos sistemas de vigilância tem como objetivo permitir com que um único operador monitore as informações coletadas por várias câmeras proporcionando assim um apontamento correto de eventos críticos em tempo real, além de extrair informações a respeito de atividades desempenhadas pelos agentes de modo a detectar interações e identificar padrões de comportamentos que sejam suspeitos antes da ocorrência de possíveis incidentes.

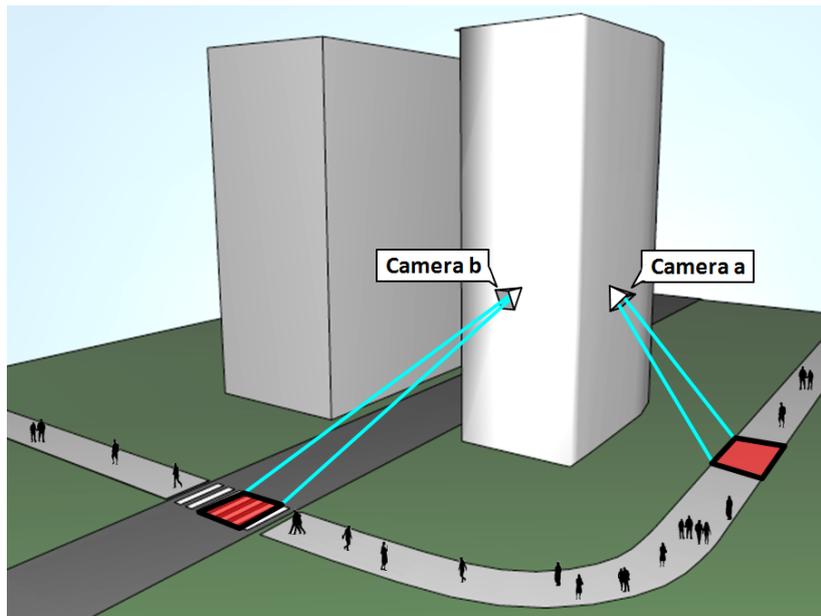


Figura 1.2: Ambiente monitorado por duas câmeras com campos de visão não sobrepostos. O objetivo da reidentificação de pessoas é determinar se uma pessoa atualmente visível na câmera A já foi observada pela câmera B [Fonte: Google imagens].

Dentre outros, um dos problemas a ser resolvido em um sistema de vigilância é determinar se uma pessoa atualmente visível já foi anteriormente observada em outra região monitorada pela rede de câmeras (figura 1.2). Este problema é conhecido como reidentificação de pessoas. Porém, para que isto seja possível, abordagens escaláveis e eficazes devem ser empregadas por meio do uso de algoritmos capazes de se adaptar a mudanças de iluminação e de ambiente, a oclusão entre pessoas e objetos, a movimentação de pessoas (mudanças de pose), a mudanças de aparência e a baixa resolução das imagens adquiridas.

A reidentificação é responsável pela manutenção da identidade de um grande número de indivíduos em uma rede de câmeras, na qual as câmeras não necessariamente possuem campos de visão sobrepostos, ou seja, é responsável por identificar ocorrências e associar a localização de agentes entre inúmeros indivíduos visualmente semelhantes que trafegam no ambiente monitorado mesmo após a ocorrência de la-

cunas significativas ao longo do tempo ou espaço. Este problema tem sido considerado para diversas aplicações, tais como vigilância [Tu et al., 2007], eventos esportivos [Ben Shitrit et al., 2011], monitoramento de tráfego [Sun et al., 2004], ajudando na prevenção, controle e investigação criminal.

Existem duas preocupações principais na elaboração de abordagens para lidar com reidentificação de pessoas. Em primeiro lugar, nem sempre é possível utilizar a informação facial, devido à baixa resolução das imagens extraídas dos vídeos adquiridos, mas apesar desta limitação, a aparência dos indivíduos deve ser modelada com precisão. Em segundo lugar, a possibilidade de haver um grande número de indivíduos em um cenário e a necessidade de salvar constantemente informações destes indivíduos requer o uso de abordagens escaláveis e eficientes para as amostras correspondentes.

Várias abordagens têm sido propostas para resolver o problema de reidentificação utilizando diferentes tipos de características, principalmente focadas na modelagem da aparência dos indivíduos [Csurka et al., 2004; Park et al., 2006; Schwartz & Davis, 2009; Bak et al., 2010; Cai & Pietikäinen, 2011; Bazzani et al., 2013], determinados principalmente por cor, textura e forma. No entanto, o aspecto visual é intrinsecamente fraco para realizar a correspondência entre imagens de indivíduos. No inverno, por exemplo, existe uma tendência em se utilizar roupas escuras, o que torna apenas a informação de cor ineficiente para discriminar um único indivíduo. Além do mais, a aparência do indivíduo pode mudar em diferentes pontos de vista da rede de câmeras devido a mudanças de iluminação e oclusão entre outros. Desta maneira, a variabilidade intra-classe pode ser significativamente maior do que a variabilidade inter-classes quando mudanças de pontos de vista da câmera estão envolvidas [Gong et al., 2014].

Devido à utilização de descritores de características baseados na aparência do indivíduo, a reidentificação pode ser tratada simultaneamente como uma indexação de descritores e como uma modelagem de aparência com o objetivo de encontrar características capazes de discriminar os indivíduos que trafegam no ambiente monitorado.

A contribuição deste trabalho é desenvolver um método de filtragem escalável, acurado e com baixo custo computacional, composto de uma estrutura de indexação que será utilizada para melhorar o desempenho na recuperação de registros. A filtragem é utilizada para encontrar a lista de candidatos mais provável para uma amostra de teste através do uso de técnicas baseadas em listas invertidas [Knuth, 1997] associadas com *bag-of-words* [Sivic & Zisserman, 2003] em conjunto com o Partial Least Squares um-contra-todos (PLS-OAA) [Schwartz & Davis, 2009].

A lista invertida é uma estratégia de indexação que permite o mapeamento de atributos para objetos ao invés de objetos para atributos, o que torna possível a realização de buscas precisas e rápidas através do uso de um identificador. Neste trabalho,

utilizou-se como identificador para a lista invertida um dicionário de *bag-of-words* que armazena as características das amostras e permite a indexação dessas características nas listas invertidas. O PLS tem sido intensamente explorado em nosso grupo de pesquisa e utilizado em vários trabalhos anteriores sobre reidentificação de pessoas [Prado et al., 2013; Schwartz, 2012b; Schwartz & Davis, 2009], reconhecimento de faces [de Paulo Carlos et al., 2015; Jr. et al., 2015] e outras numerosas tarefas.

## 1.1 Motivação

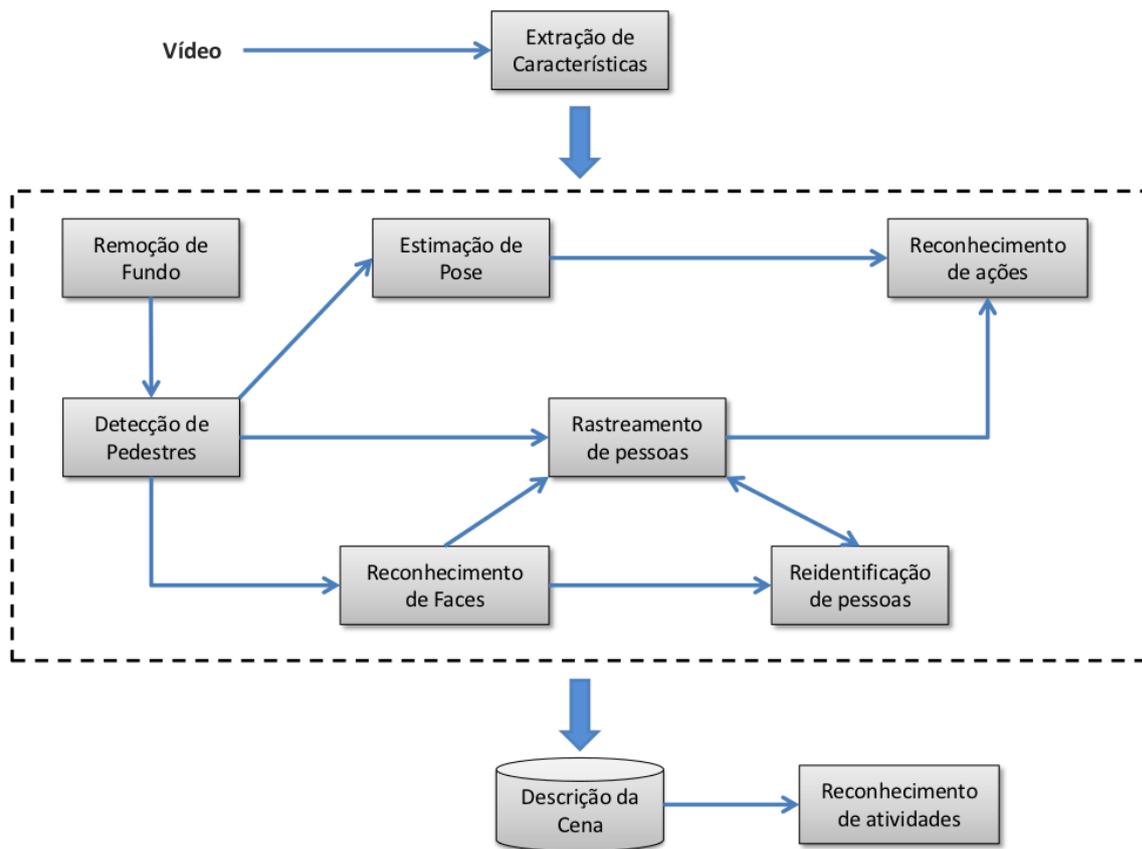


Figura 1.3: Observação de Pessoas e aplicações [Schwartz, 2012a].

Os benefícios da vigilância automática para sociedade incluem prevenção de delitos, contenção de incivildades e auxílio na investigação de fatos tidos como ilícitos, entre outros. A vigilância é uma das aplicações diretas da reidentificação, pois objetiva criar uma representação da cena para que atividades suspeitas sejam detectadas e alguma atitude por parte de agentes de segurança seja tomada preventivamente. A computação forense também pode ser beneficiada principalmente através da redução da busca manual de indivíduos em sequências de vídeos coletadas na cena de um crime.

Em algumas aplicações, a reidentificação é apenas uma etapa de um sistema mais complexo. A observação de pessoas, por exemplo, envolve etapas como reconhecimento de atividades e ações que exigem métodos rápidos e acurados [Schwartz, 2012a]. Como pode ser observado na figura 1.3, a reidentificação é uma parte de extrema importância no contexto de observação de pessoas, pois é um passo primordial para que seja possível inferir sobre as ações e atividades de um indivíduo em um sistema de tempo real.

Métodos de reidentificação rápidos e que mantenham uma acurácia alta desempenharão um importante papel nestas atividades apresentadas. Portanto, é necessário encontrar características que possibilitem a modelagem de aparência dos indivíduos e técnicas que reduzam o custo computacional durante a busca destes indivíduos na galeria<sup>1</sup>. Embora avanços tenham sido obtidos no estado da arte, a quase totalidade dos métodos de reidentificação dependem de força bruta para alcançar bons resultados [Zhao et al., 2013a,b]. Esta linha de métodos sofre de baixa eficiência computacional, o que limita o seu uso em aplicações de grande escala.

## 1.2 Problemática

Apesar dos esforços de pesquisadores da área de visão computacional, o problema de reidentificação permanece em grande parte sem solução, constituindo um problema desafiador devido principalmente a limitações das características extraídas a partir de informações visuais, muitas vezes, insuficientemente discriminativas.

A partir da observação das imagens contidas nas bases de dados disponíveis na literatura (ver figura 1.4 para um exemplo), é possível perceber que a aparência de uma pessoa sofre grandes variações entre quadros do vídeo, que podem ser ocasionadas por diferenças de iluminação, pose, oclusão, o que pode fazer com que um indivíduo se pareça mais com outro indivíduo pertencente à galeria do que com ele mesmo.

Outra dificuldade a ser considerada é a variação do tempo de transição entre câmeras, uma vez que não é possível prever quanto tempo um indivíduo permanecerá em determinado ponto de vista. O tempo varia de indivíduo para indivíduo, tornando impraticável a imposição de restrições temporais e espaciais. Por conseguinte, a reidentificação se torna mais difícil pois ela deverá lidar apenas com informações visuais da aparência dos indivíduos presentes na cena [Zheng et al., 2011, 2013].

Finalmente, o custo computacional requerido para processar enormes quantidades de dados visuais capturados a partir da rede de câmeras e adição de forma incremental

---

<sup>1</sup>No problema de reidentificação, galeria se refere ao conjunto de aparências dos indivíduos considerados, ou seja, pessoas previamente conhecidas que possuem ou não alguma correspondência com a identidade de uma amostra de teste.



Figura 1.4: Exemplo de imagens da base VIPeR [Gray et al., 2007], onde ocorrem diferenças da aparência do mesmo indivíduo ao longo do tempo. Note que as imagens da mesma coluna pertencem ao mesmo indivíduo.

de novos indivíduos à galeria requerendo a reconstrução de modelos cada vez que um novo indivíduo é visto pela primeira vez [Schwartz, 2012a].

### 1.3 Objetivos e Contribuições

O principal objetivo deste trabalho é criar um método de filtragem para encontrar a lista de candidatos mais prováveis para uma amostra, apresentando pouca dependência do número de indivíduos considerados. A lista de prováveis candidatos gerada será utilizada como entrada para um método de reidentificação do estado da arte.

A filtragem é composta de uma estrutura de indexação rápida e escalável, que tem como finalidade permitir a busca de um indivíduo detectado em um banco de dados de forma a ser independente do número de indivíduos distintos sendo considerado. A estrutura de indexação é baseada em listas invertidas [Knuth, 1997], que permitem o mapeamento de atributos para objetos, ao invés de objetos para atributos, o que mantém a escalabilidade mesmo quando há um aumento significativo no número de indivíduos. Dicionários de *bag-of-words* [Sivic & Zisserman, 2003] foram criados para armazenar e permitir a indexação de características nas listas invertidas. Com isso, é possível a obtenção da lista com as identidades mais prováveis para um candidato a um custo computacional baixo.

As seguintes contribuições são apresentadas neste trabalho:

1. Um método de filtragem rápido e acurado capaz de reduzir o tamanho da lista de

candidatos que será apresentada ao método de reidentificação em até 10 vezes, mantendo o resultado da reidentificação no rank-1 entre os melhores obtidos no estado da arte;

2. Utilização de uma estrutura de indexação com base em listas invertidas contendo um número reduzido de identidades mais prováveis para um candidato com intuito de se obter uma indexação rápida;
3. Utilização da técnica *Partial Least Squares* em conjunto com a abordagem um-contratodos [Schwartz et al., 2009] para reduzir a ambiguidade entre aparências distintas;

Como resultado dos esforços aplicados a pesquisa científica, três artigos científicos foram submetidos para publicação em conferências de aplicações em visão computacional. A lista a seguir fornece as referências a esses documentos.

- Dutra, C.; Schwartz, W.; Souza, T.; Alves, R. and Oliveira, L. (2013). Re-identifying people based on indexing structure and manifold appearance modeling. Em Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI - Conference on, pp. 218-225. ISSN 1530-1834.
- Dutra, C. R. S.; Rocha, M. C. & Schwartz, W. R. (2014). Person re-identification based on weighted indexing structures. Em Bayro-Corrochano, E. & Hancock, E., editores, Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, CIARP 2014, Puerto Vallarta, Mexico, November 2-5, 2014. Proceedings, pp. 359–366, Cham. Springer International Publishing.
- Prates, R.; Dutra, C. R. S. & Schwartz, W. R. (2016). Predominant color name indexing structure for person re-identification. Em 2016 IEEE International Conference on Image Processing (ICIP), pp. 779-783.

Esses artigos contextualizam o problema de reidentificação de pessoas, apresentam as principais ferramentas desenvolvidas e experimentos que demonstram a acurácia e ganho computacional da abordagem proposta.

## 1.4 Organização

O restante deste trabalho está organizado da seguinte forma. No Capítulo 2 são analisados os trabalhos relacionados à reidentificação de pessoas e indexação de dados.

No Capítulo 3, são detalhados as abordagens propostas. Resultados experimentais são apresentados no Capítulo 4, as discussões e análises dos resultados são relatadas no Capítulo 5. Finalmente, a conclusão deste trabalho com as observações finais e as sugestões do autor para futuras direções podem ser vista no Capítulo 6.

# Capítulo 2

## Trabalhos Relacionados

Este capítulo analisa trabalhos relacionados à reidentificação de pessoas (seção 2.1) e indexação de dados (seção 2.2). O objetivo é fornecer uma base para compreender o desenvolvimento da abordagem proposta. Uma descrição estendida e uma discussão de métodos de reidentificação de pessoas pode ser encontrado em [D’Orazio & Cicirelli, 2012; Mazzon et al., 2012; Vezzani et al., 2013; Song et al., 2013; Gong et al., 2014; Bedagkar-Gala et al., 2014].

### 2.1 Reidentificação de Pessoas

Em um sistema de vigilância, uma das tarefas fundamentais, é associar pessoas entre câmeras com diferentes campos de visão em diferentes posições e tempos [Gong et al., 2014]. Essa tarefa é conhecida como reidentificação de pessoas.

Diversas abordagens para resolver o problema de reidentificação têm sido desenvolvidas e aprimoradas ao longo dos últimos anos. Neste trabalho, essas abordagens foram divididas em três grupos. O primeiro é baseado nas características extraídas das amostras, ou seja, objetiva a modelagem de aparência por meio da criação de métodos que sejam eficientes e robustos para distinguir indivíduos parecidos (seção 2.1.1). O segundo descreve os descritores de características utilizados neste trabalho (seção 2.1.2). Por fim, o terceiro grupo se baseia na criação de técnicas de aprendizado fazendo uso de um conjunto de características para treinar modelos e realizar a correspondência das amostras (seção 2.1.3).

### 2.1.1 Modelagem de Aparência

Na reidentificação de pessoas baseada em modelagem de aparência, uma assinatura visual de um determinado indivíduo consiste em um conjunto de características extraídas a partir de uma única ou múltiplas imagens desse indivíduo. A escolha destas características e a maneira pela qual elas são combinadas para obter uma assinatura visual discriminativa é o principal desafio em reidentificação de pessoas. Essa assinatura, tem que ser capaz de maximizar a distância inter-classes entre diferentes assinaturas e, ao mesmo tempo, minimizar a distância intra-classe entre amostras do mesmo indivíduo observado em condições distintas e em diferentes câmeras. Muitas técnicas com diferentes níveis de complexidade têm sido propostas para enfrentar este desafio.

Algumas abordagens são baseadas em cálculos estatísticos sobre as amostras do indivíduo, usando informações de baixo nível e uma representação através de histogramas [Schwartz & Davis, 2009; Gray et al., 2007; Kostinger et al., 2012]. Histogramas de cores são descritores simples que acumulam os valores dos *pixels* em um histograma. Outras técnicas utilizam recursos de informação espacial, como coordenadas de pontos de interesse [Csurka et al., 2004], subdivisão das amostras em blocos [Cong et al., 2010], *grid* [Bak et al., 2012] ou delimitação de uma área específica, como partes do corpo humano [Bak et al., 2010; Farenzena et al., 2010].

Em Park et al. [2006] foi proposta uma abordagem chamada Visual Search Engine (VISE) que se baseia, em primeiro lugar, em uma segmentação do corpo do indivíduo em três partes: cabeça, tronco e pernas. Esta segmentação visa manter a correlação espacial da distribuição das características e baseia-se em dimensões aproximadas de cada parte do corpo. A parte da cabeça é ignorada devido ao baixo poder discriminativo das suas características. Cada região restante, tronco e pernas, é caracterizada por um histograma contendo 10 *bins* no espaço HSV. Cada *bin* corresponde às cores mais distinguíveis por humanos e são: vermelho, marrom, amarelo, verde, azul, violeta, rosa, branco, preto e cinza. Assim, cada região recebe como cor final aquela que foi mais frequente.

Uma abordagem discriminativa baseada em aprendizado de aparência foi proposta por Schwartz & Davis [2009], onde a assinatura de cada indivíduo é composta por diversos descritores de características. A ideia básica é construir um novo preditor de variáveis como uma combinação linear das variáveis originais (vetor de características extraído das amostras) e um vetor latente de variáveis respostas (rótulos de classes) que será utilizado na classificação<sup>1</sup>. O vetor de características é formado

---

<sup>1</sup>Um classificador é um mapeamento entre os atributos de um exemplo e sua classe. É gerado a partir de um conjunto de dados de treinamento.

através da concatenação de descritores de textura (matrizes de concorrência), borda (HOG) e cor (histograma RGB). Dessa forma, é obtido uma assinatura, ou seja, um vetor de alta dimensionalidade. Essa assinatura é transformada em um espaço latente de baixa dimensionalidade utilizando o método estatístico Partial Least Squares (PLS) em conjunto com um esquema de classificação um-contra-todos (a aparência de um indivíduo é aprendida utilizando informação sobre a aparência de outros indivíduos). A correspondência entre as amostras é realizada considerando a menor distância euclidiana encontrada entre as aparências. A desvantagem desse método é a necessidade de reconstruir todos os modelos caso um novo indivíduo seja visto na rede de câmeras. Assim, para tornar essa abordagem mais escalável, Schwartz [2012b] propôs a abordagem um-contra-alguns, selecionando apenas alguns indivíduos como amostra positiva para uma determinada amostra de teste.

Pedagadi et al. [2013] extraem informações dos espaços de cor HSV e YUV para criar um vetor de características de alta dimensionalidade. Sua dimensionalidade é então reduzida utilizando o Principal Component Analysis (PCA) de forma não supervisionada, seguido pela aplicação do Local Fisher Discriminant Analysis (LFDA) [Sugiyama, 2006] para otimizar a função de distância durante a comparação.

Em [Kenk et al., 2015], a imagem de uma pessoa é dividida em 25 segmentos retangulares sobrepostos. Para cada um, são construídos três histogramas RGB, uma para cada canal. As imagens são comparadas utilizando a métrica de distância Hellinger, em que 0 representa que duas imagens são semelhantes. Uma estrutura de distribuição hierárquica é utilizada para fazer a correspondência de forma rápida e eficiente. Também são aplicadas regras para esquecer os modelos mais antigos, evitando assim, o aumento descontrolado do número de comparações entre pessoas que há muito tempo deixaram a cena e não são mais susceptíveis a voltar.

Leng et al. [2015] se basearam na hipótese de que se duas imagens pertencem ao mesmo indivíduo, elas devem ser similares em conteúdo (similaridade visual) e em contexto (compartilhar os mesmos vizinhos mais próximos). Para calcular a medida de similaridade, além da amostra de teste e a galeria, um novo conjunto de dados formado pela amostra de teste e por outras imagens da galeria original é criado. Por fim, um *ranking* bidirecional é proposto para calcular o resultado final a partir das similaridades de conteúdo e contexto. A vantagem deste método é não utilizar técnicas de aprendizagem.

O método proposto por Liu et al. [2014a] cria uma ponderação, dando peso para características que discriminam mais determinada amostra através de um aprendizado *top-down* não supervisionado. Segundo os autores, algumas características de aparência podem ser mais importantes do que outras para descrever e distinguir um indivíduo.

Baseados nesta hipótese, vários descritores de características (cor e textura) são extraídos das amostras, um conjunto de protótipos é selecionado a partir de um conjunto de amostras de treinamento não etiquetados usando *clustering random forest* [Breiman, 2001; Liu et al., 2000] e um classificador *random forest* [Breiman, 2001] é treinado para discriminar diferentes protótipos. Dessa maneira, para cada amostra de teste, a importância de cada característica é calculada utilizando o classificador treinado.

Em Cheng et al. [2011] foi utilizado o conceito de Pictorial Structures (PS) para segmentar o corpo humano em partes a fim de estimar a pose e criar uma assinatura única. A assinatura é formada pela união dos histogramas de cor (modelo HSV) de cada parte do corpo em conjunto com o uso do algoritmo Maximally Stable Color Regions (MSCR) aplicado em regiões da imagem. No caso do uso de *multiple-shots*, primeiramente um modelo Custom Pictorial Structure (CPS) que captura a aparência local é aprendido. Dadas duas assinaturas, o casamento (*matching*) entre elas é realizado utilizando uma função de distância baseada em Bhattacharyya [Farenzena et al., 2010]. A principal desvantagem desse método é que ele não funciona corretamente quando há oclusão nas imagens. Cheng & Cristani [2014] mostraram os resultados dessa abordagem aplicada a vários tipos de problemas como estimação de poses e detecção de pedestres.

Cai & Pietikäinen [2011] modelaram a aparência criando um dicionário baseado em contexto de cor global. O método agrupa características de cor visualmente similares para obter um dicionário de cor, através do k-means. A distribuição de ocorrência espacial de alto-similaridades das palavras de cor são aprendidas e combinadas para caracterizar a aparência dos indivíduos. O método Symmetry-Driven Accumulation of Local Features (SDALF), proposto por [Bazzani et al., 2013], modela a aparência de indivíduos baseado na simetria do corpo humano. A ideia é que características que se encontram mais próximas ao eixo de simetria recebam maior peso que as características que se encontram em partes mais periféricas do corpo. Primeiramente, o método faz a separação dos *pixels* de primeiro plano e do plano de fundo para obter a silhueta e, em seguida, a silhueta é dividida e três atributos são analisados: a cromaticidade global da imagem através da ponderação de histogramas de cores (HSV, RGB), a distribuição espacial das cores em regiões estáveis -MSCR e a presença de temas recorrentes com alta entropia -Recurrent High-Structured Patches (RHSP).

### 2.1.2 Descritores de características

Um aspecto importante em reidentificação de pessoas é a utilização de diferentes descritores de características para validar o desempenho dos métodos. Isto se deve ao fato de

que a importância de um descritor de características depende da aplicação (ambiente interno ou externo) e até mesmo das características das imagens da galeria e de teste utilizadas [Liu et al., 2012]. Todos os descritores de características listados a seguir foram utilizados neste trabalho.

O filtro de predominância RGB [Dutra et al., 2013]<sup>2</sup> tem como objetivo destacar o canal de cor que é predominante em cada pixel. Dada uma imagem RGB,  $I = (C_R, C_G, C_B)$ , onde  $C_{R,G,B}$  representa um valor entre 0 e 1 para cada canal, o espaço de cor é dividido em 8 regiões bem definidas por um parâmetro  $T$ , como se segue

$$P_T(C_R, C_G, C_B) = \begin{cases} (\mu_R, 0, 0) & \text{if } C_R - \max(C_G, C_B) > T \\ (0, \mu_G, 0) & \text{if } C_G - \max(C_R, C_B) > T \\ (0, 0, \mu_B) & \text{if } C_B - \max(C_G, C_R) > T \\ (0, \mu_G, \mu_B) & \text{if } C_G \approx C_B > C_R + T \\ (\mu_R, 0, \mu_B) & \text{if } C_R \approx C_B > C_G + T \\ (\mu_R, \mu_G, 0) & \text{if } C_R \approx C_G > C_B + T \\ (\mu_R, \mu_G, \mu_B) & \text{if } C_R \approx C_G \approx C_B > 0.5 \\ (0, 0, 0) & \text{if } C_R \approx C_G \approx C_B \leq 0.5 \end{cases}$$

onde  $\mu_{\{R,G,B\}}$  indica a média do canal que satisfaz as condições da equação para uma região da imagem, e  $T$  está no intervalo  $[0; 1]$ ; " $\approx$ " significa  $|C_i - C_j| < T$ , com  $i, j \in \{R, G, B\}$ . O filtro de predominância tem o efeito físico de separar as cores de acordo com a figura 2.1.

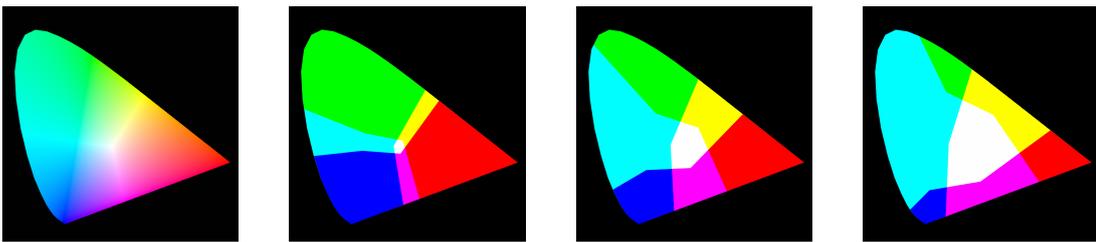


Figura 2.1: Espaço de cor RGB e efeito do filtro de predominância com  $T = 0.1, 0.3$  e  $0.5$ , respectivamente.

O descritor Salient Color Names based Color Descriptor (SCNCD), proposto por Yang et al. [2014b], relaciona cores marcantes com 16 nomes<sup>3</sup> de cores no espaço RGB. Essas cores são extraídas de modelos baseados em partes, em diferentes espaços

<sup>2</sup>o filtro de predominância foi proposto por Souza em [Dutra et al., 2013].

<sup>3</sup>fuchsia, blue, aqua, lime, yellow, red, purple, navy, teal, green, olive, maroon, black, gray, silver and white.

de cor (RGB, HSV, RGB normalizado (rgb) e  $l_1l_2l_3$  [Gevers & Smeulders, 1999]). Após tais cores serem obtidas, são fundidas para gerar uma representação das características das amostras. Para melhorar o desempenho, Yang et al. [2014b] utiliza em conjunto com o SCNCD os histogramas de cor obtidos em cada espaço de cor, realizando uma fusão desses dois descritores.

Lisanti et al. [2014] propuseram o descritor Weighted Histograms of Overlapping Stripes (WHOS). Este descritor faz uso de uma reponderação para redistribuir a energia entre os elementos que contribuíram de forma mais significativa para garantir que os melhores candidatos sejam classificados em cada iteração do método. Pesos são calculados para ponderar cada pixel dos histogramas HSV e RGB computados sobre faixas sobrepostas e o descritor HOG é calculado sobre a imagem original. Em seguida esses descritores são concatenados.

O descritor Probabilistic Relative Distance Comparison (PRDC), proposto por Zheng et al. [2011], visa maximizar a probabilidade de um par correto de amostras de treino e teste terem uma distância menor do que um par de imagens erradas, ou seja, o objetivo é aprender a distância ótima que pode maximizar a acurácia independente da escolha da representação das características. Isso faz com que modelo seja mais tolerante à mudanças de aparência.

O descritor de características Local Maximal Occurrence (LOMO) [Liao et al., 2015] analisa a ocorrência horizontal de características locais, e maximiza a ocorrência de uma representação estável mesmo quando ocorre mudanças de ponto de vista. Para lidar com as variações de iluminação, a transformada Retinex e um operador de textura invariável à escala é aplicado. Em seguida, um método de aprendizado foi criado chamado Cross-view Quadratic Discriminant Analysis (XQDA).

### 2.1.3 Técnicas baseadas em Aprendizagem

Métodos recentes de reidentificação baseados em técnicas de aprendizado consideram o uso de descritores de características para modelagem de aparência. Cor, textura e forma são as características mais utilizadas, onde estas características são combinadas a fim de obter um descritor mais representativo [Cai & Pietikäinen, 2011; D'angelo & Dugelay, 2011; Li et al., 2014].

Uma abordagem recorrente, é o uso de métricas de distância que são aprendidas a partir de vetores de características extraídos das amostras obtidas por diferentes pontos de vista de uma rede de câmeras. Dentro deste contexto, a distância de Mahalanobis tem recebido considerável interesse. Seu objetivo principal é construir uma represen-

tação e aprender uma métrica que reflita a transição entre as câmeras, permitindo que ela seja usada para determinar a similaridade entre uma amostra desconhecida e uma conhecida. Uma discussão sobre as métricas que utilizam a distância de Mahalanobis pode ser encontrada em [Roth et al., 2014].

Hirzer et al. [2012a] e Hirzer et al. [2012b] utilizaram uma métrica de distância chamada Large Margin Nearest Neighbor (LMNN) que tem como objetivo melhorar a classificação utilizando k-Nearest Neighbors (k-NN). A ideia do LMNN é treinar a métrica de Mahalanobis de tal forma que os k-vizinhos mais próximos sempre pertençam à mesma classe, enquanto que os vizinhos mais distantes sejam separados por uma grande margem. Em [Hirzer et al., 2012b], o conjunto de treinamento é dado por pares de vetores que compartilham a mesma etiqueta. Em ambos os trabalhos, o aprendizado das métricas foi realizado a partir de imagens obtidas por duas câmeras com pontos de visão diferentes. O método proposto por [Liu et al., 2014b] é bem semelhante, pois aprende a similaridade com base na hipótese de que as variações de aparência de imagens de uma mesma pessoa deveriam ser similares após uma deformação, enquanto que as imagens de aparências diferentes deveriam ser dissimilares após a mesma deformação.

Kostinger et al. [2012] utilizaram inferência estatística para o aprendizado de uma métrica de distância baseada na distância de Mahalanobis chamada de Keep It Simple and Straightforwar Metric (KissMe). Para calcular a covariância entre as amostras e aprender a distância de Mahalanobis, são utilizados pares de imagens de indivíduos iguais e pares de imagens de indivíduos distintos, gerados aleatoriamente. Zhao et al. [2014b] propuseram um método que aprende filtros *mid-level* automaticamente para capturar padrões visuais de partes específicas do corpo humano. Durante o processo de treinamento, cada amostra é dividida em *patches* regulares e características foram densamente extraídas. Em seguida, é calculada uma correspondência entre a amostra de treinamento da câmera A contra todas as outras amostras de treinamento da câmera B, e *patches* com alto poder de discriminação e generalização são selecionados. O próximo passo consiste na construção de um agrupamento hierárquico de *patches* para agrupar subconjuntos de aparência visual semelhantes. Filtros *mid-level* são então aprendidos de forma supervisionada. Um Support Vector Machines (SVM) linear é criado para cada nó de uma árvore contendo os exemplos de treinamento positivos formados por *patches* que pertencem à um mesmo nó, e exemplos negativos formados pelos nós restantes. A correspondência entre as imagens é realizada utilizando as respostas dos filtros.

O Filter Pairing Neural Network (FPNN) foi apresentado por Li et al. [2014] para tratar e otimizar conjuntamente desalinhamento, transformações geométricas e fotomé-

tricas, oclusão e desordem de fundo por meio do aprendizado de características ideais para a tarefa de reidentificação. Em seguida, a rede neural é treinada utilizando estratégias que incluem *bootstrapping*, *dropout*, *data augmentation* e *data balancing* buscando diminuir erros, *overfitting* e um possível desequilíbrio das amostras de treinamento.

Conjuntos de dados distintos podem apresentar características muito diferentes entre si como variação do ângulo de visão, iluminação e oclusão. Assim, combinar múltiplas funções de distância usando pesos predeterminados é indesejável, pois características altamente discriminativas em um ambiente podem se tornar irrelevantes em outro. Diante disso, Paisitkriangkrai et al. [2015] apresentaram abordagens eficazes para aprender pesos de funções de distância. Duas abordagens foram consideradas: a primeira maximiza a distância relativa entre as imagens de diferentes indivíduos e imagens do mesmo indivíduo, de tal forma que a curva Cumulative Match Characteristic (CMC) aproxima-se com um número pequeno de candidatos retornados. A segunda otimiza a probabilidade dos candidatos retornados estarem corretos utilizando o algoritmo Cutting-plane de aprendizado estruturado.

Baseado em métodos de pesquisa de imagens, Zheng et al. [2015] utilizaram uma representação *bag-of-words* não supervisionada para resolver o problema da reidentificação. Um dicionário é criado com as amostras de treinamento e cada indivíduo é representado por um histograma de palavras visuais. Em seguida, várias técnicas são empregadas, como *burstiness weighting* [Jégou et al., 2009], *negative evidences* [Jégou & Chum, 2012] e *root descriptor* [Arandjelovic & Zisserman, 2012]. Finalmente, um etapa automática de re-ranking é realizada para refinar a lista final de candidatos.

Com o intuito de obter representações discriminativas e robustas, Li et al. [2015] propuseram o *framework* Cross-view Projective Dictionary Learning (CPDL) para aprender um dicionário visual. Utilizando o CPDL, os autores criam duas representações de baixa dimensionalidade para cada indivíduo, uma a nível de *patch* e outra a nível de imagem, com o objetivo de capturar as relações intrínsecas de diferentes coeficientes de configuração utilizando algoritmos de otimização. Por fim, uma estratégia de fusão é aplicada para calcular a similaridade das amostras.

Uma abordagem mais recente para resolver o problema da reidentificação de pessoas é utilizar a informação de grupo ou contexto de grupo. Esta abordagem não leva em conta apenas o indivíduo em questão, mas também as pessoas que se encontram em seu entorno, partindo da informação que geralmente em espaços públicos as pessoas andam em grupos. Utilizando essa ideia, Zheng et al. [2014] propuseram um método que, por meio da associação de grupos de pessoas, tenta compreender e inferir o comportamento de um indivíduo dentro do grupo e provê um contexto vi-

sual capaz de auxiliar a reidentificação, uma vez que a aparência frequentemente sofre alterações causadas por fatores como mudança de iluminação, oclusão ou pose. Os autores mostram que combinar o contexto de grupo gera uma melhoria nos resultados da reidentificação de pessoas. Para isto, eles combinam duas métricas, Block-Based Ratio-Occurrence (BRO) e Center Rectangular Ring Ratio-Occurrence (CRRRO), com o descritor Scale-Invariant Feature Transform (SIFT) no espaço RGB aplicadas sobre grupos e indivíduos. Para fazer a classificação, é utilizada a técnica de aprendizado Ranking SVM sobre as características mais discriminativas. Porém, uma detecção automática de grupos se faz necessária, o que na prática pode não ser muito viável.

Neste trabalho, propomos uma estrutura de indexação baseada em listas invertidas que permite o mapeamento de atributos para objetos. Assim como Zheng et al. [2015] e Li et al. [2015], utilizamos um dicionário visual de *bag-of-word* para armazenar as características das amostras e permitir a indexação dessas características. Uma otimização proposta é a modelagem das relações entre as amostras utilizando o algoritmo PLS em conjunto com o esquema de classificação um-contratodos para reduzir a ambiguidade entre as aparências. Diferentemente de Kostinger et al. [2012], Hirzer et al. [2012a], Hirzer et al. [2012b] e Liu et al. [2014b] que utilizaram a distância de Mahalanobis para calcular a similaridade entre as amostras, a abordagem proposta calcula a similaridade através da distância euclidiana que é muito mais simples, rápida de ser calculada e com resultados comparáveis.

Em todos os trabalhos listados neste capítulo, o resultado final foi obtido considerando todos os indivíduos da galeria. A principal vantagem dos nossos métodos, é a não utilização de toda a galeria. Os resultados experimentais mostram que é possível reduzir em até 10 vezes o tamanho da lista de candidatos e ainda assim ter resultados equiparáveis ao estado da arte.

## 2.2 Indexação de Dados

Estruturas de indexação de dados são estruturas de acesso auxiliar utilizadas para melhorar o desempenho na recuperação de registros. Em aplicações de recuperação de texto, as listas invertidas [Knuth, 1997] são as estruturas de indexação mais comuns [Zobel & Moffat, 2006], elas mapeiam termos às suas ocorrências em um documento, dessa forma, permitem o mapeamento de atributos para objetos ao invés de objeto para atributos como normalmente utilizado. É uma estrutura de indexação que permite a realização de buscas rápidas e precisas, porém com um custo maior para inserção de objetos. Neste trabalho, utilizam-se listas invertidas para criar uma estrutura

de indexação rápida contendo um número reduzido de candidatos para uma amostra de teste. Uma breve descrição de métodos relacionados é apresentada a seguir.

Outra aplicação que requer o uso de estruturas de indexação é a recuperação de imagens em grande escala. Essa aplicação tem como objetivo retornar uma lista ordenada de imagens relevantes da galeria considerando sua similaridade com a amostra de teste apresentada. A maneira mais comum de consultar uma imagem é realizar a consulta por vizinhos mais próximos conhecido por  $k - NN$ , que consiste em calcular as distâncias entre o teste e as amostras da galeria e retornar as primeiras  $k$  amostras mais próximas. A abordagem força bruta  $k - NN$  consiste em calcular todas as distâncias entre as amostras de teste e a galeria e classificá-las, o que é inviável quando a galeria é muito grande. Além disso, a pequena diferença entre os valores de distância calculados diminui o desempenho e dificulta a análise correta das amostras.

Para resolver este problema de forma eficiente, muitos algoritmos têm sido propostos. Um exemplo é a árvore kd-tree [Gan et al., 2007], que divide recursivamente o espaço em sub-regiões e os organiza em uma estrutura hierárquica que fornece complexidade assintótica de  $O(\log(n))$  para uma galeria de tamanho  $n$ . A consulta por  $r$  vizinhos mais próximos,  $r - NN$ , é mais eficiente do que o  $k - NN$ . Ele utiliza uma abordagem de pesquisa binária [Chávez et al., 2001] e consiste em retornar todas as amostras na galeria que estão dentro de um raio  $r$  fixo de distância a partir da amostra teste. Schwartz et al. [2012] utilizaram uma estrutura baseada em árvores para otimizar a busca por correspondências de uma amostra de teste para o problema de identificação de faces. As partições da árvore são feitas de forma aleatória e um modelo de regressão baseado em PLS é usado para guiar a pesquisa.

O Partial Least Squares for Face Hashing (PLSH) proposto por [?] é um método simples e eficiente para identificação de faces baseado em PLS e em funções hash aleatórias e independentes. O primeiro passo do método extrai características das amostras utilizando vários descritores resultando em um grande vetor de características. Em seguida, este vetor é apresentado à abordagem de filtragem, que consiste basicamente em dividir os indivíduos da galeria em dois conjuntos de mesmo tamanho: positivo e negativo. O modelo de regressão PLS é então utilizado para determinar se a amostra de teste pertence ao conjunto positivo ou negativo. Ao fim desta etapa, é gerada uma lista de candidatos, onde o primeiro indivíduo da lista é o que possui a maior probabilidade de ser o indivíduo correspondente à amostra de teste. Os experimentos mostram que o *speedup* médio é de até 233 vezes comparado com a abordagem força bruta. Esta redução do custo computacional se deve ao fato de não ser necessário avaliar toda a lista de candidatos. Também foi utilizada seleção de características para reduzir o custo computacional para avaliar as funções hash resultando no extended PLS for face

hashing (ePLSH).

Assim como nos trabalhos descritos, a abordagem proposta neste trabalho tem como objetivo tornar a busca de uma amostra de teste rápida, escalável e acurada. Ao invés de focar na modelagem de aparência, faz-se uso de uma estrutura de indexação que é capaz de tornar a busca independente do número de indivíduos presentes na galeria, gerando como resposta uma lista com os prováveis candidatos. Dessa forma, diferentemente de Schwartz & Davis [2009], é possível criar um modelo um-contratodos com todos os indivíduos e testar apenas aqueles que estão na lista dos prováveis candidatos, o que torna a correspondência mais escalável e eficiente.



# Capítulo 3

## Metodologia

O processo geral de reidentificação pode ser descrito da seguinte forma. Dado um grande número de imagens capturadas pela câmera  $c_1$  (amostras da galeria) e uma imagem de teste capturada pela câmera  $c_2$ , o objetivo da reidentificação de pessoas é encontrar a correspondência entre as imagens de testes e a galeria.

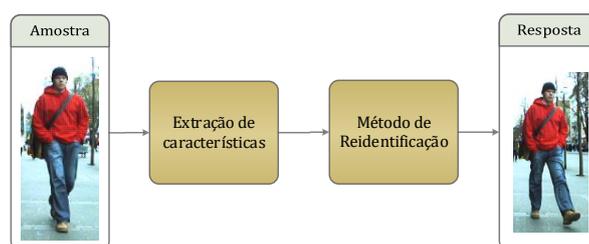


Figura 3.1: Etapas do processo de reidentificação de pessoas.

A figura 3.1 mostra as etapas do método de reidentificação de pessoas tradicional. Em primeiro lugar, é encontrada uma representação das características das amostras de teste<sup>1</sup> e da galeria. Em seguida, um método de reidentificação é utilizado para medir a similaridade entre as amostras de teste e a galeria, e assim gerar uma resposta final.

A contribuição deste trabalho é criar um método de filtragem para encontrar a lista de candidatos mais prováveis para uma amostra, que será utilizada com entrada para o método de reidentificação. Esse processo pode ser observado na figura 3.2.

A filtragem é composta de uma estrutura de indexação baseada listas invertidas [Knuth, 1997], que tem como objetivo melhorar o desempenho na recuperação de registros. A similaridade entre cada par de potenciais correspondências é medida, e a maior encontrada indica que aquele par de amostras é correspondente, sendo então obtida uma lista com os prováveis candidatos. Parte desta lista é utilizada como galeria e

<sup>1</sup>Amostra de teste se refere a um conjunto de elementos extraídos de uma imagem, no contexto da reidentificação de pessoas é a parte da imagem que contém apenas a pessoa que se deseja reidentificar.



Figura 3.2: Etapas do processo de reidentificação de pessoas.

um método de reidentificação da literatura é aplicado para encontrar a correspondência entre a amostra de teste e a nova galeria, obtendo assim a resposta final.

Neste capítulo, dois métodos que utilizam uma estrutura de indexação baseada em listas invertidas são apresentados para abordar o problema da reidentificação de pessoas. O primeiro utiliza dicionários de *bag-of-words* [Sivic & Zisserman, 2003] como índices das listas invertidas, onde cada dicionário contém as palavras visuais que representam da melhor forma possível um indivíduo da galeria. As palavras visuais utilizadas são os descritores de características extraídos das amostras. O segundo faz uso do algoritmo *Partial Least Squares* (PLS) com o objetivo de reduzir a ambiguidade entre as aparências dos indivíduos.

Na seção 3.1 são apresentados os descritores de características utilizados. O processo de indexação e os métodos desenvolvidos para abordar o problema de reidentificação de pessoas, denominados *indexação baseada em Dicionários Visuais* e *indexação baseada em Partial Least Squares* são apresentados na seção 3.2. Por fim, o método de reidentificação utilizado é descrito na seção 3.3.

### 3.1 Descritores de características

Devido à presença de ruído nos processos de aquisição de dados, a obtenção de correspondências de qualidade entre indivíduos em imagens distintas é um grande desafio em Visão Computacional. Dessa forma, a criação de descritores que identifiquem corretamente as partes de maior importância nos pares de imagens correspondentes é um dos fatores mais importantes para a construção de um sistema de visão ou reconhecimento de padrões. O desempenho alcançado por estes sistemas está diretamente relacionado ao poder de discriminação do conjunto de características utilizado. Os descritores de características devem minimizar possíveis diferenças dentro da mesma classe de objetos (intra-classe) e destacar as diferenças entre classes distintas (inter-classes).

De forma geral, os descritores de características são classificados em globais,

quando são obtidos com base em todos os pontos da imagem e, locais, normalmente calculados a partir de pequenas regiões na imagem. Uma descrição estendida sobre vários descritores de características pode ser encontrada em [Mikolajczyk & Schmid, 2005; Li & Allinson, 2008].

Difícilmente, um único descritor de característica é robusto às variações de aparência causadas por mudanças de pose, iluminação ou ponto de vista. Por isso, neste trabalho, foram considerados vários descritores de características baseados em cor, textura e forma que obtiveram bom desempenho em trabalhos do estado da arte em reidentificação de pessoas. São eles: o descritores WHOS [Lisanti et al., 2014], LOMO [Liao et al., 2015], PRDC [Zheng et al., 2011], SCNCD [Yang et al., 2014b], os histogramas de cores nos espaços RGB e HSV, o filtro de predominância RGB [Dutra et al., 2013] e as características utilizadas pelos autores do KissMe [Kostinger et al., 2012], intitulado neste trabalho de KISSME. Detalhes sobre estes descritores podem ser encontrados na seção 2.1.1.

Os descritores WHOS, LOMO, PRDC e KISSME são obtidos a partir da amostra inteira, enquanto que os outros utilizam o conceito de modelos baseados em partes [Satta, 2013], onde cada amostra é particionada em seis regiões horizontais não sobrepostas de igual tamanho: cabeça, tronco superior e inferior, pernas superiores e inferiores e pés.

## 3.2 Indexação de Dados

Estruturas de indexação de dados são estruturas de acesso auxiliar utilizadas para melhorar o desempenho na recuperação de registros, onde a pesquisa por um registro é restrita a um subconjunto, ou seja, não é necessário analisar o conjunto de dados completo.

Neste trabalho, utilizou-se listas invertidas que são estruturas de indexação de dados proposta por Knuth [1997] que permitem o mapeamento de atributos para objetos, ao invés de objetos para atributos, como normalmente utilizado. Aplicado ao nosso problema, ela permite mapear características extraídas de um indivíduo à seus identificadores que, diferentemente de Schwartz [2012b] e Schwartz & Davis [2009], nos permite criar uma estrutura de indexação cujo tamanho independe do número de indivíduos na galeria, tornando-a rápida e escalável para efetuar a correspondência entre a imagem de teste e a galeria.

O processo de indexação pode ser dividido em três etapas, organização das características, criação das listas e teste. A etapa de organização ocorre após a obtenção

dos vetores de características de cada amostra da galeria. Nesta fase, é necessário organizar e armazenar, da melhor forma possível, em um índice as características obtidas de um subconjunto das amostras, pois são elas que identificam as partes de maior importância de uma amostra. Na segunda etapa, é calculada a similaridade entre as amostras utilizadas para aprendizado e as amostras do subconjunto, o índice é então utilizado para criar e armazenar nas listas invertidas os identificadores das amostras com maior similaridade. Adicionalmente, um esquema de ponderação pode ser utilizado para identificar as regiões mais significativas. Na etapa de teste, a similaridade entre uma amostra de teste e o subconjunto de características é medida e o índice é utilizado para recuperar os identificadores das amostras associados a cada posição da lista invertida. É possível que uma mesma amostra pertença à mais de uma lista invertida, desta maneira, um esquema de votação se faz necessário para finalmente obter a lista de candidatos.

### 3.2.1 Indexação baseada em Dicionários Visuais

Com objetivo de obter uma lista de indivíduos candidatos para coincidir com a identidade de uma amostra de teste, uma estrutura de indexação [Dutra et al., 2013] é criada. Durante o processo de construção da estrutura de indexação, um dicionário visual baseado em *bag-of-words* é aprendido para armazenar as características das amostras e permitir a indexação dessas características. O dicionário é criado selecionando características de um indivíduo e atribuindo cada indivíduo a uma palavra visual de acordo com a similaridade entre os vetores de características das amostras de teste e galeria. O processo é dividido em três etapas: *aprendizado do dicionário*, *treinamento* e *teste*.

#### 3.2.1.1 Aprendizado do Dicionário

Como a lista invertida é indexada por atributos do objeto, dicionários baseados em *bag-of-words* foram construídos para extrair tais atributos das amostras. O processo é exemplificado na figura 3.3. Primeiramente, as amostras dos  $n$  indivíduos com identidade conhecida são divididas utilizando o modelo baseado em partes que divide cada imagem em  $m$  regiões horizontais não sobrepostas de igual tamanho: cabeça, tronco superior e inferior, pernas superiores e inferiores, e pés (passo 1 da figura 3.3). Em seguida é feito a extração das características utilizando descritores de características (passo 2 da figura 3.3). O descritores são armazenados em vetores de características.

O dicionário é então criado pela seleção aleatória de  $k$  vetores de características como palavras visuais (passo 3 da figura 3.3). Experimentos foram realizados utili-

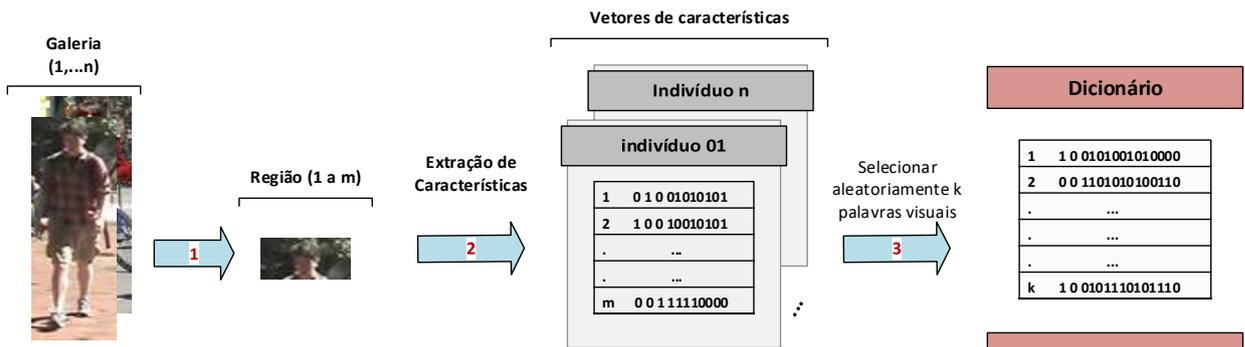


Figura 3.3: Aprendizado do dicionário. Primeiramente as amostras dos indivíduos são divididas em  $m$  regiões horizontais não sobrepostas (1), em seguida é realizada a extração de características (2). Por último, é criado o dicionário selecionando aleatoriamente  $k$  vetores de características como palavras visuais (3).

zando o algoritmo de clusterização k-means para selecionar as palavras visuais, mas os resultados são muito semelhantes aos obtidos pela seleção aleatória.

A razão para se utilizar um dicionário é encontrar as palavras visuais que representam da melhor forma possível as regiões da amostra de modo que suas posições no dicionário possam ser usadas como índices nas listas invertidas. Por exemplo, se a  $i$ -ésima palavra visual é a mais próxima de um conjunto de indivíduos, identificadores do indivíduo serão adicionados na lista invertida atrelada ao  $i$ -ésimo índice correspondente aquela região. Desta forma, é considerado que, se uma amostra de teste tem um vetor de características semelhante à  $i$ -ésima palavra visual do dicionário, é provável que a sua identidade pertença a um dos indivíduos que estão na lista invertida ligada ao  $i$ -ésimo índice.

Buscando melhorar a precisão da estrutura de indexação, foi realizada a seguinte otimização: variou-se o número de dicionários visuais utilizados e assim extraiu-se uma quantidade maior de características relevantes [Dutra et al., 2014], tornando o dicionário mais discriminativo. A figura 3.4 apresenta as diferentes formas utilizadas para aprender o dicionário. A primeira consiste em aprender um dicionário para cada região da amostra, dessa forma é possível capturar informações locais das partes do corpo. A segunda aprende vários dicionários globais, o que permite que vários votos sejam considerados ao calcular a similaridade entre uma amostra de teste e a galeria.

### 3.2.1.2 Treinamento

Uma vez que o dicionário foi aprendido de forma não-supervisionada, o processo de treinamento é executado para preencher as listas invertidas, e  $m$  listas são criadas – uma por região da amostra. A ideia de criar várias listas invertidas é permitir a votação

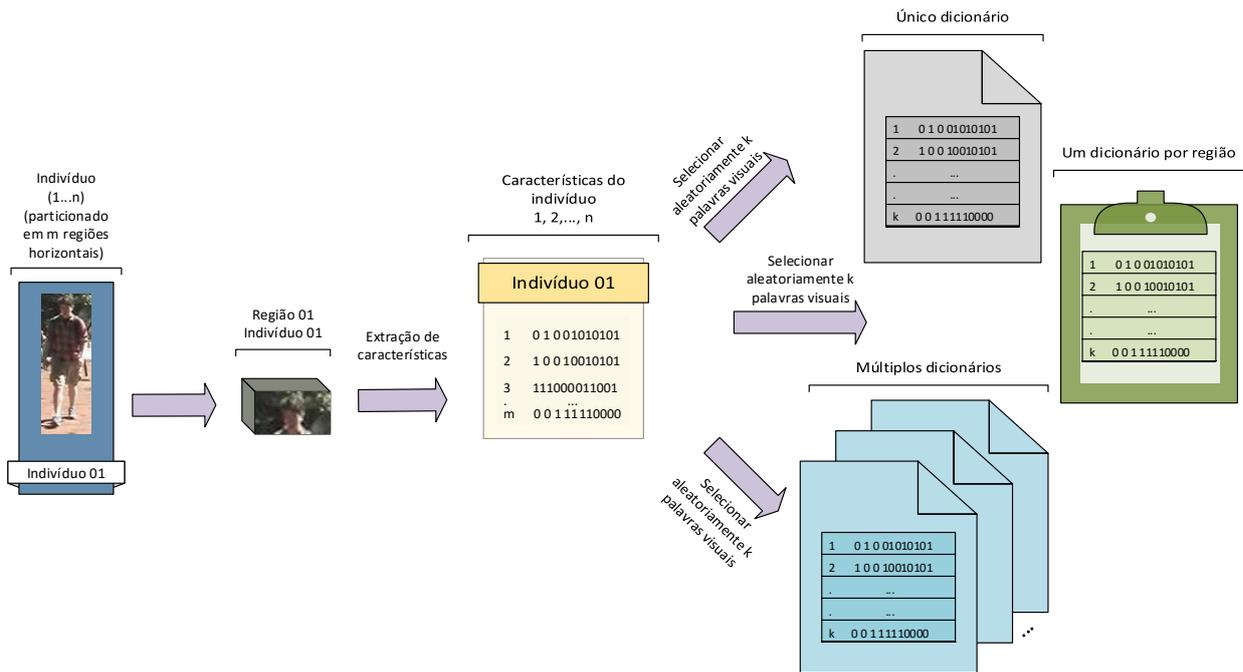


Figura 3.4: Abordagens para a estrutura de indexação baseado em dicionários: dicionário global único, dicionário local para cada região e múltiplos dicionários globais.

para selecionar os indivíduos com maior probabilidade de identificar uma amostra de teste sem ser afetado por ruído incorrido em algumas regiões, o que pode interferir nos descritores de características. A figura 3.5 exhibe as etapas de treinamento e teste da estrutura de indexação.

Na fase de treinamento, as  $n$  amostras da galeria de teste são particionadas em  $m$  regiões não sobrepostas. Os descritores de características são extraídos das  $m$  regiões de cada amostra e armazenados em vetores de características. Em seguida, utilizando o dicionário criado na etapa anterior, o vetor de características de cada região é comparado (passo 1 da figura 3.5) a todas as palavras visuais do dicionário e o índice mais próximo – a  $i$ -ésima palavra visual – é utilizado para indexar a  $i$ -ésima posição da lista invertida (passo 2 da figura 3.5). As comparações foram realizadas utilizando o cálculo da distância euclidiana entre as características do dicionário e as características das amostras de treinamento.

Finalmente, se o vetor de características da  $j$ -ésima região da amostra do indivíduo  $l$  for mais próximo da  $i$ -ésima palavra visual no dicionário, o identificador do indivíduo  $l$  será adicionado na  $i$ -ésima posição da lista invertida. Desta forma, durante o teste, se a amostra de teste apresenta um vetor de características semelhante a  $i$ -ésima palavra visual no dicionário, ela será indexada à identidade do indivíduo  $l$ . Note que a  $i$ -ésima posição da  $j$ -ésima lista invertida pode conter mais que um indivíduo.

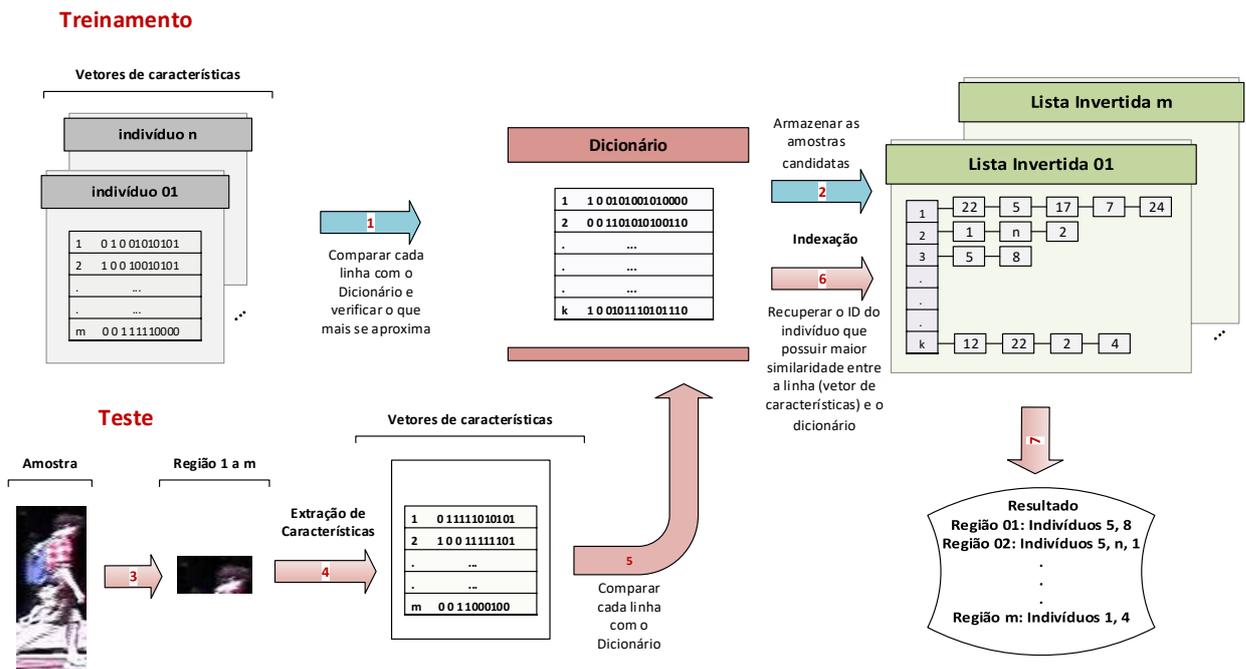


Figura 3.5: Na etapa de treinamento, os vetores de características das amostras da galeria são comparados ao dicionário visual (1), o índice da palavra visual mais próxima é utilizado para indexar a lista invertida, onde será armazenado o identificador da amostra analisada (2). No teste, a amostra de teste é particionada em regiões não sobrepostas (3), em seguida é feita a extração de características (4). Os vetores de características da amostra de teste são comparados ao dicionário visual (5), o índice da palavra visual que mais se assemelha ao vetor de característica é utilizado para indexar a lista invertida (6) e recuperar os identificadores das amostras da galeria que possuem maior similaridade com a amostra de teste (7).

Buscando atenuar a influência de regiões da imagem que pertencem ao plano de fundo e aumentar o significado de regiões pertencentes ao indivíduo a ser identificado, foi realizada uma otimização do método utilizando uma função Gaussiana [Cai & Pietikäinen, 2011] para ponderar as partições mais relevantes. A função Gaussiana atribui maior importância às regiões mais próximas do centro da amostra e menor importância às regiões mais distantes. Isso é possível uma vez que os conjuntos de dados assumem que os indivíduos foram detectados e se encontram no centro da imagem.

### 3.2.1.3 Teste

O objetivo deste passo final é selecionar, a partir de uma consulta na estrutura de indexação, um subconjunto de indivíduos que é mais provável de conter a identidade correta da amostra de teste apresentada.

Dada uma amostra de teste, o primeiro passo é dividi-la em  $m$  regiões da mesma maneira que foi feito nas etapas de aprendizado do dicionário e treinamento (passo 3 da figura 3.5) e, em seguida, obter as características para cada uma das regiões (passo 4 da figura 3.5). O vetor de características da  $j$ -ésima região é então comparado com todas as palavras visuais contidas no dicionário calculando-se a distância euclidiana entre elas (passo 5 da figura 3.5). Por fim, o índice da palavra visual mais próxima, o  $i$ -ésimo, é utilizado para indexar o subconjunto  $s_j$  contendo os identificadores das amostras da galeria mais similares a  $j$ -ésima região da amostra de teste, estes identificadores estão na  $i$ -ésima posição da  $j$ -ésima lista invertida (passo 6 da figura 3.5).

Após o conjunto  $S$  ter sido obtido (passo 7 da figura 3.5) com o processo descrito (o conjunto  $S$  é a união dos subconjuntos  $\{s_1, s_2, \dots, s_m\}$  que contém os identificadores mais semelhantes as regiões  $1, 2, \dots, m$ , uma vez que, foi criado uma lista invertida para cada região da amostra de teste), a estratégia adotada para encontrar os  $p$  indivíduos mais prováveis que identificam uma amostra de teste é a seguinte: o indivíduo mais provável é o que aparece o maior número de vezes em  $S$ , o segundo, é aquele que aparece o segundo maior número de vezes, e assim por diante, até a obtenção dos  $p$  indivíduos que aparecem mais em  $S$ . A motivação para esta abordagem é que um indivíduo que aparece várias vezes, possivelmente apresenta vetores de características semelhantes ao extraído da amostra de teste. Dessa forma, o resultado final é uma lista ordenada pela maior similaridade entre a amostra de teste e a galeria, contendo os prováveis candidatos.

### 3.2.2 Indexação baseada em Partial Least Squares

Com o objetivo de obter uma lista de candidatos mais acurada, em que o candidato correto esteja nas primeiras posições da lista, uma segunda abordagem é proposta. Esta abordagem difere da anterior por utilizar o algoritmo *Partial Least Squares* (PLS) nos vetores de características buscando diminuir a ambiguidade entre aparências distintas e obter uma correspondência melhor.

PLS é um método de regressão por mínimos quadrados parciais utilizado para modelar as relações entre um conjunto de variáveis observadas trocando as variáveis originais por um subconjunto de variáveis latentes. Foi criado por Herman Wold na década de 1970 e tem sido explorado em diversas áreas, tais como Bioinformática, Quimiometria e Neurociências. Mais recentemente, PLS tem sido empregado com sucesso a problemas de Visão Computacional para realizar classificação de dados, redução de dimensionalidade e regressão [Schwartz & Davis, 2009; Schwartz et al., 2009, 2010, 2012; Sharma & Jacobs, 2011; Junior & Schwartz, 2014]. Enquanto detalhes adicionais

sobre o método PLS podem ser encontrado em [Rosipal & Krämer, 2006], uma breve descrição matemática é fornecida na seção 3.2.2.1.

O uso do PLS é motivado pela sua robustez em combinar vários descritores de características e sua capacidade de lidar com alta dimensionalidade. Além disso, ele apresenta baixo custo computacional para análise das amostras, uma vez que, apenas um único produto escalar entre coeficientes de regressão e o vetor de características é necessário para estimar uma resposta.

Para aprender um modelo PLS discriminativo para cada indivíduo foi utilizado o esquema *one-against-all* (um-contra-todos) [Schwartz & Davis, 2009]. Dessa forma, quando um indivíduo  $l$  é considerado, todos os indivíduos restantes são utilizado como contraexemplos para o indivíduo  $l$ . A utilização do PLS em conjunto com o esquema um-contra-todos permite fazer um processo de seleção das características mais importantes, pois o PLS consegue dar pesos mais elevados para as características localizadas em regiões mais discriminativas. Uma vez estimado o modelo PLS para cada aparência, os vetores de características que descrevem esta aparência são projetados sobre os vetores de peso, resultando em características de baixa dimensionalidade que serão utilizadas durante a fase de teste para corresponder à uma amostra de consulta.

Neste trabalho, o *Partial Least Squares one-against-all* (PLS-OAA) foi utilizado em duas abordagens distintas. Na primeira, ele é aplicado sobre o dicionário visual e a escolha das características foi feita de forma aleatória (seção 3.2.2.2); na segunda, a galeria foi dividida em subconjuntos formados pela permutação dos vetores de características (seção 3.2.2.3) sem a utilização de um dicionário visual.

### 3.2.2.1 Regressão PLS

Seja  $\chi \subset \mathbb{R}^m$  um espaço de características de dimensão  $m$  e similarmente seja  $\gamma \subset \mathbb{R}$  um espaço unidimensional representando os rótulos das classes. A quantidade de amostras é dada por  $n$ . O PLS decompõe a matriz  $X(n \times m)$  e o vetor  $Y(n \times 1)$  em

$$\begin{aligned} X_{n \times m} &= T_{n \times p} P_{m \times p}^T + E_{n \times m} \\ Y_{n \times 1} &= U_{n \times p} Q_{p \times 1}^T + F_{n \times 1}, \end{aligned}$$

onde  $T_{n \times p}$  e  $U_{n \times p}$  são as variáveis latentes para os vetores de características e as respostas, respectivamente,  $P_{m \times p}$  e  $Q_p$  representam os *loadings* e a matriz  $E$  e o vetor  $F$  são os respectivos resíduos. O algoritmo PLS computa  $P$  e  $Q$  de tal maneira que a covariância entre  $U$  e  $T$  seja máxima.

O método PLS utiliza o algoritmo Nonlinear Iterative PLS

(NIPALS) [Wold, 1985], para construir o subespaço latente composto de um conjunto de vetores de projeção (vetores de peso)  $W = \{w_1, w_2 \cdots w_p\}$  tal que

$$[\text{cov}(t_i, u_i)]^2 = \max_{|w_i|=1} [\text{cov}(Xw_i, y)]^2,$$

onde  $t_i$  é a  $i$ -ésima coluna da matriz  $T$ ,  $u_i$  a  $i$ -ésima coluna da matriz  $U$  e  $\text{cov}(t_i, u_i)$  é a covariância entre os vetores latentes  $t_i$  e  $u_i$ . O vetor de regressão  $\beta$  entre  $T$  e  $U$  é calculado usando a matriz  $W$  de acordo com

$$\beta = W(P^T W)^{-1}(T^T T)^{-1} T^T Y.$$

A resposta  $\hat{y}$  da regressão PLS para uma amostra do vetor de características  $x_{1 \times d}$  é dada por  $\hat{y} = \bar{y} + \beta^T (x - \bar{x})$  onde  $\bar{y}$  e  $\bar{x}$  são a média de  $Y$  e dos elementos de  $X$ , respectivamente [Schwartz, 2012b].

### 3.2.2.2 PLS-OAA usando dicionário visual

Esta abordagem é uma otimização do método descrito na seção 3.2.1. Ao invés de utilizar diretamente as palavras visuais do dicionário e a distância euclidiana para encontrar a palavra visual mais próxima de uma amostra de teste, o aprendizado é realizado utilizando o modelo PLS calculado para cada palavra visual. O modelo é obtido utilizando o esquema um-contratodos, onde a  $i$ -ésima palavra visual é utilizada como amostra positiva e as palavras visuais restantes com amostras negativas. Após a aplicação do PLS, cada dicionário visual contém os modelos PLS obtidos que serão utilizados nas etapas de treinamento e teste (figura 3.6). Nesta abordagem vários dicionários são criados, o que torna o método mais discriminativo e aumenta a taxa de reconhecimento obtida.

Nas etapas de treinamento e teste é realizada a projeção dos vetores de características de cada indivíduo sobre o subespaço latente estimado para cada palavra visual, resultando em um valor de regressão para cada palavra visual. Na etapa de treinamento, as palavras visuais que obtiveram melhores respostas na projeção (maiores valores de regressão), são utilizadas para indexar a lista invertida, que são preenchidas com os identificadores dos indivíduos. Na fase de teste, as melhores respostas da projeção são utilizadas para indexar a lista invertida e recuperar a lista de candidatos. Esse processo pode ser visto na figura 3.7.

Ao final da etapa de teste, são obtidas as listas invertidas de cada dicionário com base na similaridade entre o descritor de características e as palavras visuais do dicionário, estas listas contém os candidatos locais. Finalmente, várias listas são

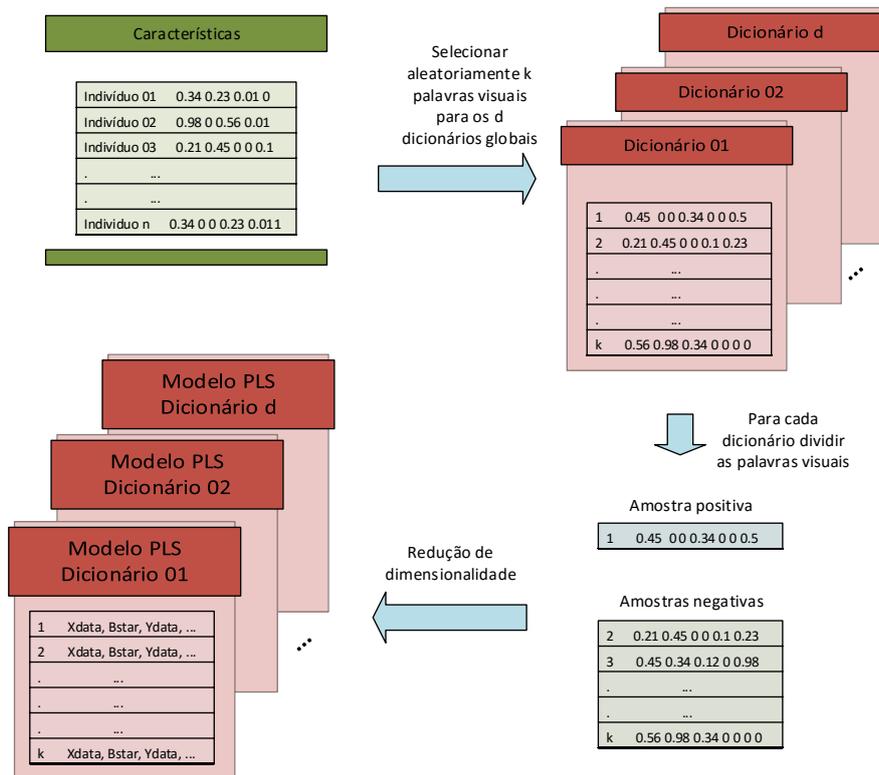


Figura 3.6: Aprendizado dos dicionários utilizando o PLS-OAA com dicionários visuais.

combinadas em uma lista global única contendo os potenciais candidatos para coincidir com uma amostra de teste. A lista global possui os identificadores da galeria ordenados com o uso de um histograma que realiza a votação. O histograma tem tamanho igual ao número de indivíduos na galeria. A resposta da regressão é utilizada para ponderar cada posição do histograma que equivale ao identificador de cada indivíduo presente na lista de candidatos, dessa forma, os candidatos com melhor resposta de regressão recebem peso maior. Uma ilustração desse processo é mostrada na figura 3.8.

### 3.2.2.3 PLS-OAA usando permutação de indivíduos

Nas abordagens que fazem uso do dicionário visual, apenas uma pequena parte das amostras da galeria são utilizadas (poucas palavras visuais são escolhidas para o aprendizado do dicionário), o que pode tornar o método pouco discriminativo. Por outro lado, a utilização de todas as amostras seria impraticável, uma vez que, para cada amostra seria necessário aprender um modelo PLS. Buscando tornar a abordagem mais discriminativa e ao mesmo tempo gerar poucos modelos PLS, visando a obtenção de um modelo escalável, optou-se por dividir a galeria em subconjuntos, onde cada subconjunto é formado permutando os vetores de características de todos os indivíduos da galeria.

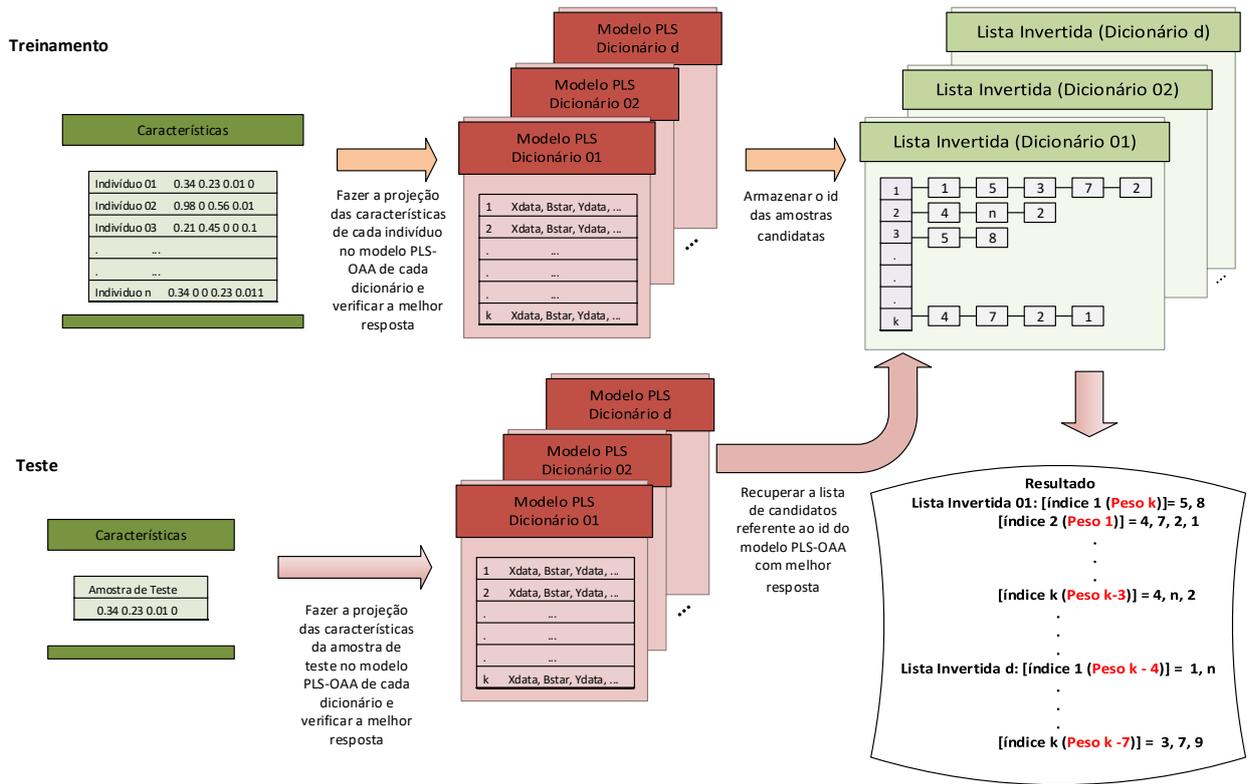


Figura 3.7: Etapas de treinamento e teste do processo de reidentificação usando PLS-OAA com dicionários visuais.

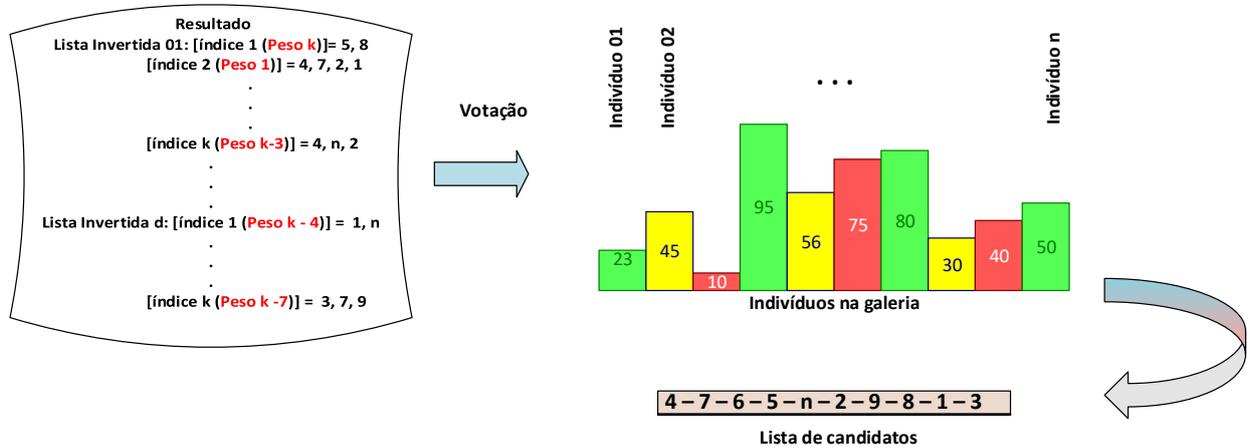


Figura 3.8: Estratégia para gerar uma única lista global a partir dos candidatos locais, considerando a quantidade de indivíduos igual a 10.

Esta abordagem não faz uso de um dicionário visual e possui apenas duas etapas: treinamento e teste. No treinamento, os vetores de características das amostras da galeria são divididos aleatoriamente em  $s$  subconjuntos. Em seguida, um modelo PLS é aprendido para cada subconjunto utilizando o esquema um-contra-todos (quando o  $i$ -ésimo subconjunto é considerado como classe positiva, todos os subconjuntos restan-

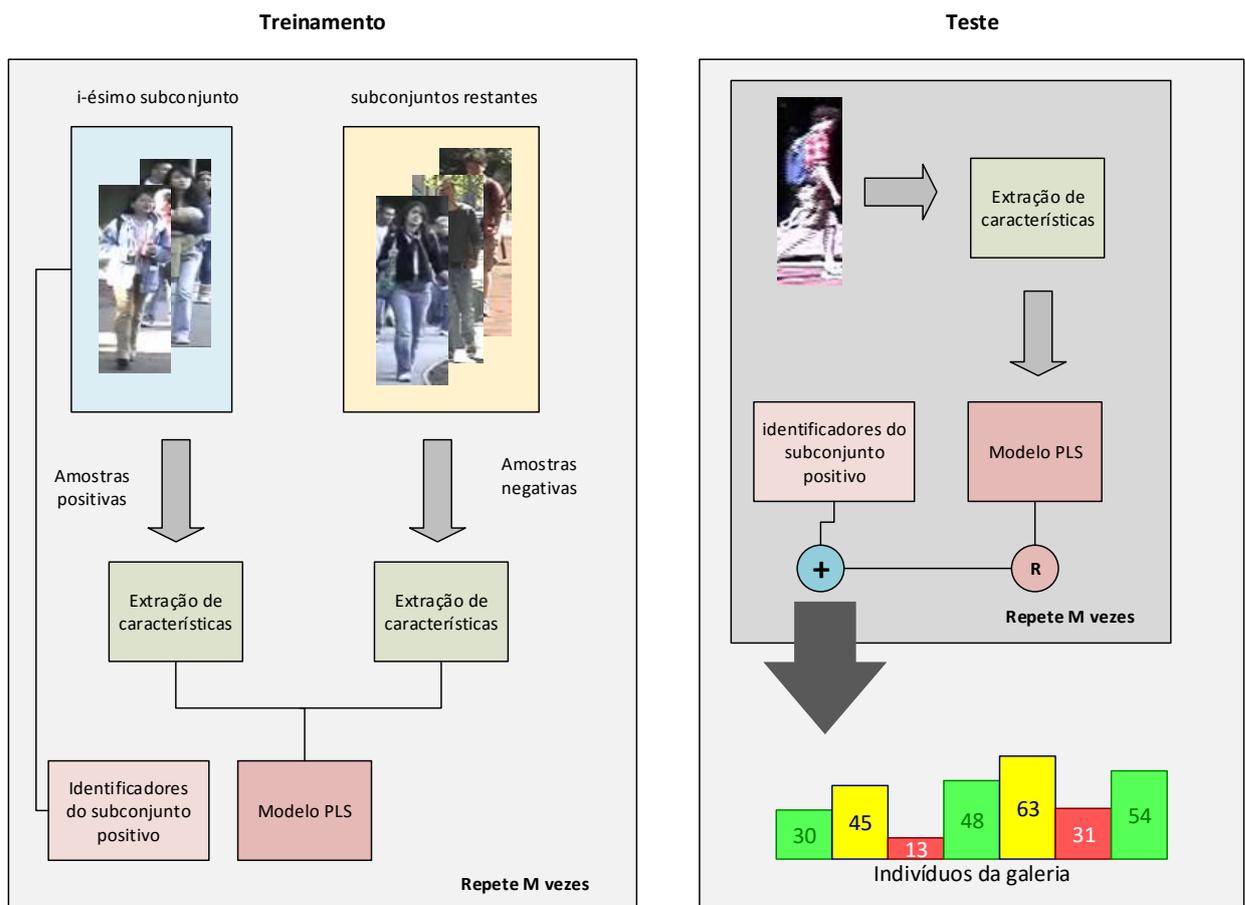


Figura 3.9: Abordagem PLS-OAA usando permutação de indivíduos.

tes são utilizados com contraexemplo). Uma vez estimado o modelo PLS para cada subconjunto, os identificadores dos indivíduos pertencentes ao subconjunto rotulado como amostras positivas são armazenados. Esse processo é repetido *M* vezes e vários modelos são criados.

Na fase de teste, são utilizados os mesmos descritores de características e eles são apresentados aos modelos criados para obter um valor de regressão *r*. Em seguida, os identificadores armazenados para o subconjunto de amostras positivas é recuperado. Um histograma de votação de tamanho igual a galeria é criado, e cada posição do histograma referente aos identificadores é incrementada em *r*. Este histograma é ordenado obtendo assim a lista de candidatos final. A ilustração deste processo pode ser visto na figura 3.9.

### 3.3 Método de Reidentificação

O método de reidentificação utilizado neste trabalho foi o algoritmo Cross-view Collaborative Representation based Classification (X-CRC) criado por de Carvalho Prates & Schwartz [2015]. O X-CRC foi inspirado no esquema de codificação conhecido como Collaborative Representation based Classification (CRC) [Zhang et al., 2012] que representa cada amostra de teste usando todas as imagens da galeria, em vez de comparar diretamente as distâncias entre uma amostra de teste e uma amostra da galeria. Um dos problemas do CRC é ignorar a transição entre as câmeras, que é um dos principais desafios em reidentificação.

O X-CRC representa cada par de imagens de teste  $y$  e a galeria  $x$  usando de forma colaborativa dicionários de campos de visão específicos nomeados de  $D1$  e  $D2$  respectivamente, de forma a evitar a comparação direta entre imagens de diferentes campos de visão. Para incorporar informações de campos de visão distintos, o autor atribui a cada coluna dos dicionários  $D1$  e  $D2$ , os descritores de características que pertencem a uma mesma amostra de treinamento capturada pelas câmeras utilizadas para teste e galeria, respectivamente. Portanto, coeficientes de codificação semelhantes devem ser obtidos quando  $y$  e  $x$  correspondem ao mesmo indivíduo obtidos de câmeras com campos de visão diferentes. A correspondência entre os pares de amostras de teste e galeria ocorre de forma indireta utilizando esses coeficientes ( $\alpha_1$  e  $\alpha_2$ ). Uma breve descrição do algoritmo X-CRC é apresentada no apêndice A.

O principal pressuposto feito pelos autores é que imagem de teste  $y$  será representada com um pequeno erro de reconstrução somente se  $x$  corresponde à sua respectiva imagem da galeria. Devido ao termo de similaridade, qualquer imagem distinta irá forçar um grande erro de reconstrução para a imagem de teste.

Neste trabalho, a base de dados foi particionada em dois grupos: amostras de treinamento e amostra de teste. As amostras de treinamento foram utilizada para realizar o aprendizado do algoritmo de reidentificação X-CRC, e todas as amostras foram utilizadas. Após a aplicação dos métodos de indexação, uma porcentagem da lista de candidatos obtida é apresentada ao algoritmo X-CRC como amostras de teste para assim obtermos a lista de candidatos final.

# Capítulo 4

## Resultados Experimentais

Neste capítulo são realizados os experimentos para avaliar as abordagens propostas em duas bases de dados. A seção 4.1 descreve a configuração experimental comum, incluindo base de dados e medidas de avaliação. Os resultados da reidentificação baseada na indexação efetuada por dicionários visuais para a base de dados VIPeR são apresentados na seção 4.2. Os resultados da reidentificação baseada na indexação utilizando Partial Least Squares para a base de dados VIPeR são apresentados na seção 4.3. Finalmente, a seção 4.4 mostra os resultados obtidos utilizando a base de dados CUHK01.

### 4.1 Configuração Experimental

Em ambientes reais de reidentificação de pessoas, a imagem de cada pessoa que aparece pela primeira vez em uma das câmeras é armazenada na base de dados ou galeria. Em um segundo momento, quando essa mesma pessoa reaparece em outra câmera, uma nova imagem capturada é utilizada no sistema como imagem de consulta para procurar a imagem correspondente na galeria. Procurando simular o ambiente real, as imagens da galeria e de consulta são selecionadas de forma aleatória e cada teste é realizado 10 vezes, um procedimento comum para alcançar resultados mais estáveis.

Os conjuntos de dados para reidentificação são geralmente formados por pares de imagens capturadas em câmeras distintas e são divididos em conjuntos de treinamento e teste. Os pares de imagens no conjunto de treinamento são utilizados para aprender alguns modelos de aprendizado de máquina, e os pares restantes são empregados para avaliação utilizando estes modelos. Em Song et al. [2013] e Vezzani et al. [2013] podemos encontrar uma extensa discussão sobre este tema.

### 4.1.1 Base de dados

Os experimentos foram realizados em duas bases de dados públicas e muito utilizadas no estado da arte: VIPeR [Gray et al., 2007], e CUHK01 [Li et al., 2013a].

O conjunto de dados VIPeR (figura 4.1a) contém 632 indivíduos, tendo apenas duas amostras por indivíduo, capturadas a partir de diferentes pontos de vista de duas câmeras. Essa base de dados é muito desafiadora devido à existência de apenas uma imagem por pessoa de cada uma das vistas das câmeras. Além disso, as imagens apresentam variações de aparência, iluminação e pose. Todas as imagens estão normalizadas no tamanho  $128 \times 48$  *pixels*.

A base de dados CUHK01 (figura 4.1b) possui um número maior de indivíduos (971) extraídos a partir de duas câmeras, onde cada indivíduo possui duas amostras para cada ponto de vista. A câmera A captura a vista frontal ou a vista traseira de pedestres, enquanto a câmera B captura a vista lateral. A base de dados CUHK01 difere da base de dados VIPeR devido à menor variação de ponto de vista e pose. As imagens estão normalizadas no tamanho  $160 \times 60$  *pixels*.



Figura 4.1: Amostras dos conjuntos de dados: (a) VIPeR e (b) CUHK01. As amostras que estão na mesma linha pertencem ao mesmo indivíduo.

Como na maioria dos trabalhos que utilizam a base de dados VIPeR, o conjunto de dados é dividido aleatoriamente em dois subconjuntos de treinamento e teste. Cada subconjunto recebe metade dos pares de imagens. Diferentemente, a base de dados CUHK01 considera 485 imagens para treinamento e 486 imagens para teste. Neste trabalho, utiliza-se apenas uma imagem de cada ponto de vista da câmera, assim, foi selecionado aleatoriamente uma imagem entre as duas disponíveis. No subconjunto

de teste, imagens de uma câmera são consideradas como galeria e imagens da outra câmera são consideradas como amostra de teste.

A base de dados VIPeR foi utilizada para validar os parâmetros das abordagens propostas, pois é a base de dados mais difícil da literatura em reidentificação de pessoas.

### 4.1.2 Medidas de avaliação

Uma maneira típica para se avaliar o problema de reidentificação de pessoas é a utilização da curva *Cumulative Match Characteristic* (CMC) [Gray et al., 2007], a qual representa a expectativa de encontrar uma correspondência correta nos primeiros  $n$  *ranks*, ou seja, a taxa de reconhecimento obtida nos primeiros  $n$  candidatos presentes nas listas. Na curva CMC, a taxa de reconhecimento aumenta ou permanece constante a medida que o número de prováveis candidatos são recuperados.

Com base nas propriedades da curva CMC, utilizamos a medida da área sob a curva, *Area Under the Curve* (AUC), que apesar de incomum, em reidentificação de pessoas é uma boa maneira de comparar resultados diferentes usando uma medida única, assim é possível quantificar o desempenho de forma mais compacta. Os valores da AUC são computados no intervalo [1 : 30] (a partir do rank 1 até o rank 30) e são normalizados dentro do intervalo [0, 1].

Também reportamos os resultados utilizando a métrica de avaliação *Maximum Achievable Recognition Rate* (MARR) proposta por ?. Essa métrica é calculada considerando que um método de reidentificação perfeito é empregado para diferentes tamanhos da lista de candidatos, proporcionando um limite superior para a taxa de reconhecimento alcançada pelos métodos de reidentificação. Dessa forma, a melhor curva apresenta MARR igual a um para qualquer porcentagem de indivíduos na lista de candidatos. Para a avaliação dos parâmetros necessários dos métodos de indexação, o gráfico MARR foi plotado utilizando 50% da lista de candidatos original.

## 4.2 Indexação baseada em Dicionários Visuais

Esta seção avalia a abordagem baseada em dicionários visuais e seus parâmetros utilizando a base de dados VIPeR. O descritor de características utilizado para a validação de todos os parâmetros foi o SCNCD concatenado com os histogramas de cor como sugerido por Yang et al. [2014b]. Optou-se por utilizar um descritor baseado em partes pois a abordagem baseada em dicionários foi proposta considerando que as amostras são particionadas em regiões não sobrepostas.

São avaliados o número de palavras visuais para o aprendizado do dicionário, o número de dicionários globais, a avaliação dos descritores de características utilizados e, por último, a avaliação dos métodos de otimização.

**Número de palavras visuais** Este teste avalia o número de palavras visuais utilizadas para o aprendizado do dicionário. Foi realizado um experimento variando o número entre 5 e 30 palavras com incremento de 5.

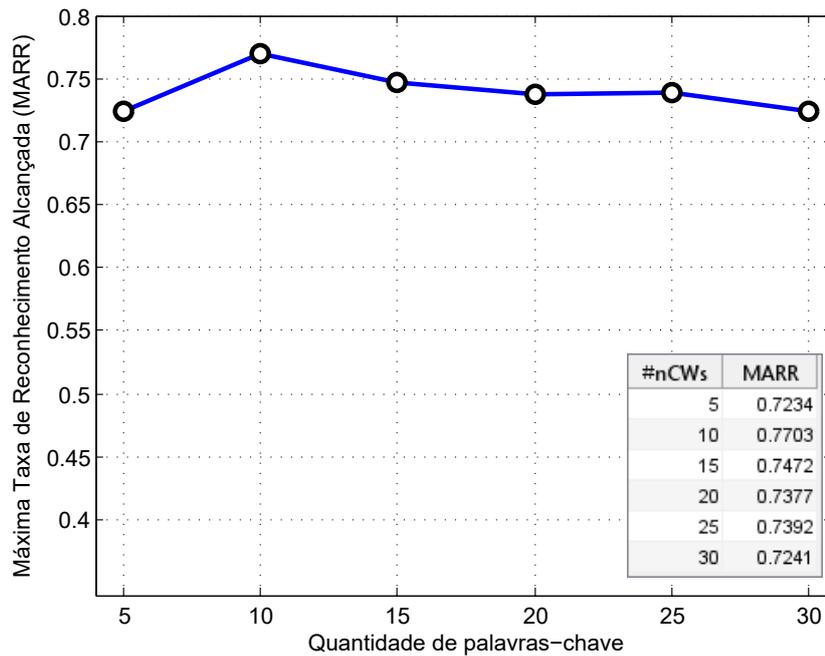


Figura 4.2: Taxa de reconhecimento alcançada pela abordagem baseada em dicionários visuais como uma função do número de palavras visuais do dicionário.

A figura 4.2 mostra o gráfico MARR obtido. O melhor resultado é alcançado com a utilização de 10 palavras visuais, com um MARR igual a 0.7703, o que significa que, ao selecionar 50% da lista de candidatos, 77.03% dos candidatos corretos estão inseridos nesta lista. Assim, o número de palavras visuais foi fixado em 10 nos experimentos restantes.

**Número de dicionários globais** Este teste avalia o número de dicionários globais necessários quando é utilizada a abordagem de otimização que permite a criação de múltiplos dicionários. Foi realizado um experimento variando o número de dicionários globais entre 2 e 20 com incremento de 2. De acordo com a figura 4.3 o melhor resultado é alcançado utilizando 4 dicionários globais.

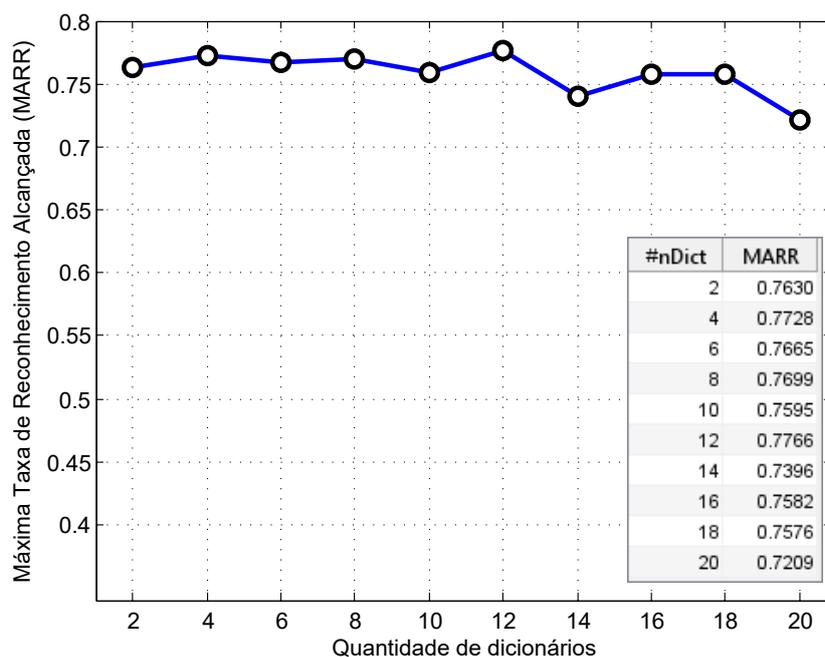


Figura 4.3: Avaliação do número de dicionários globais da abordagem múltiplos dicionários.

**Descritores de características** Este teste avalia a acurácia obtida utilizando vários descritores de características do estado da arte, são eles: Filtro de predominância, histograma RGB, histograma HSV, SCNCD, LOMO, WHOS, PRDC e as características utilizadas no KISSME. Para detalhes ver seção 2.1.1.

A figura 4.4 mostra a curva CMC e o gráfico de barras MARR obtido após a aplicação do algoritmo X-CRC sobre 50% da lista de candidatos. O melhor resultado foi alcançado utilizando o descritor SCNCD.

**Avaliação dos métodos** Avaliou-se a ponderação de blocos por meio da função gaussiana e a variação do número de dicionários visuais utilizados. A figura 4.5 mostra o gráfico de barras MARR e a curva CMC obtidos.

De acordo com os resultados, pode-se observar que a utilização de vários dicionários visuais não melhora a taxa de reconhecimento. Um possível motivo é que ao se utilizar vários dicionários, diversos identificadores de indivíduos podem receber o mesmo número de votos, impedindo que o método se torne mais discriminativo. A utilização de um único dicionário com ponderação gaussiana trouxe uma pequena melhoria na taxa de reconhecimento.

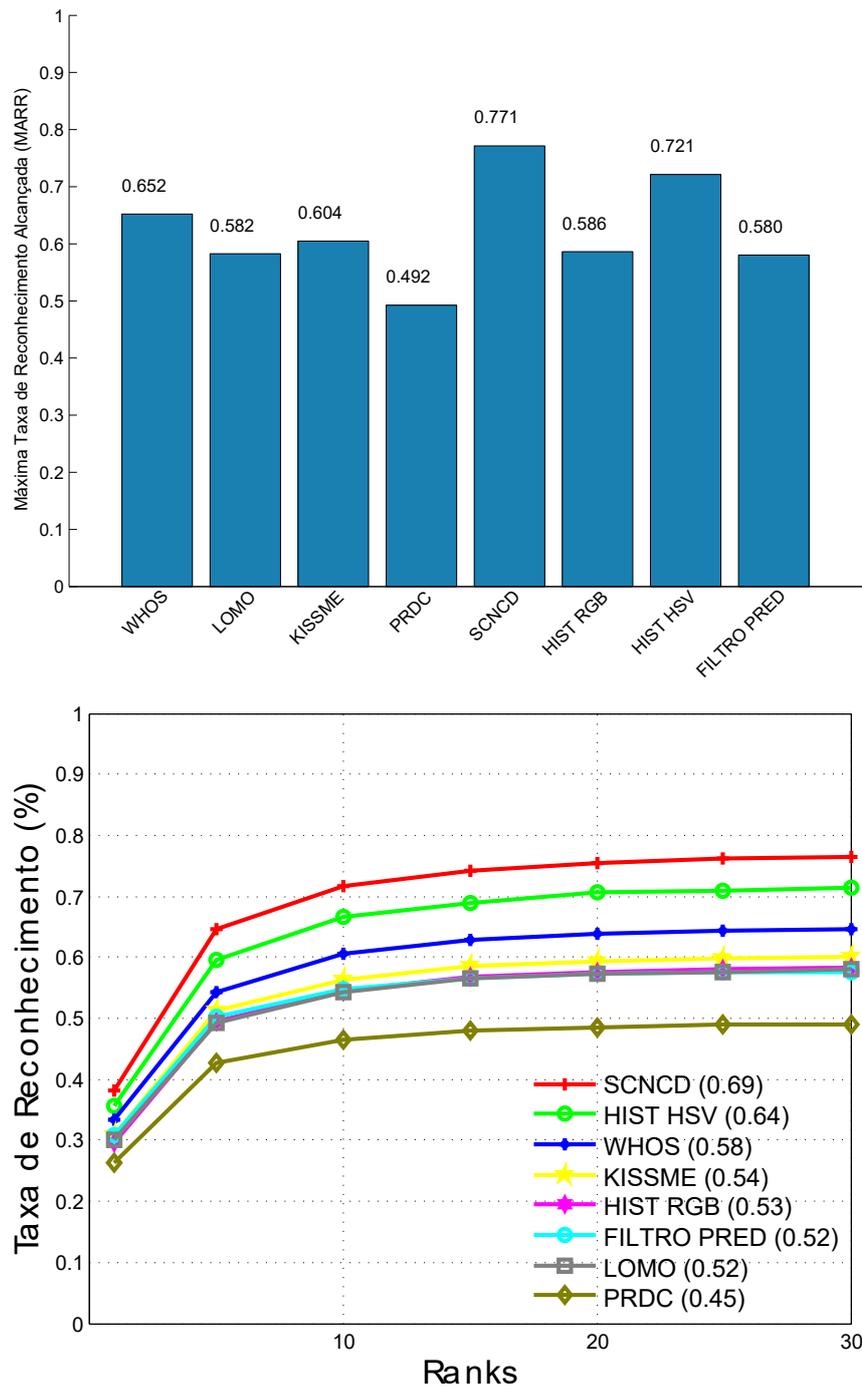


Figura 4.4: Gráfico de barras MARR e curva CMC mostrando os resultados obtidos pela abordagem baseada em dicionários visuais utilizando vários descritores de características.

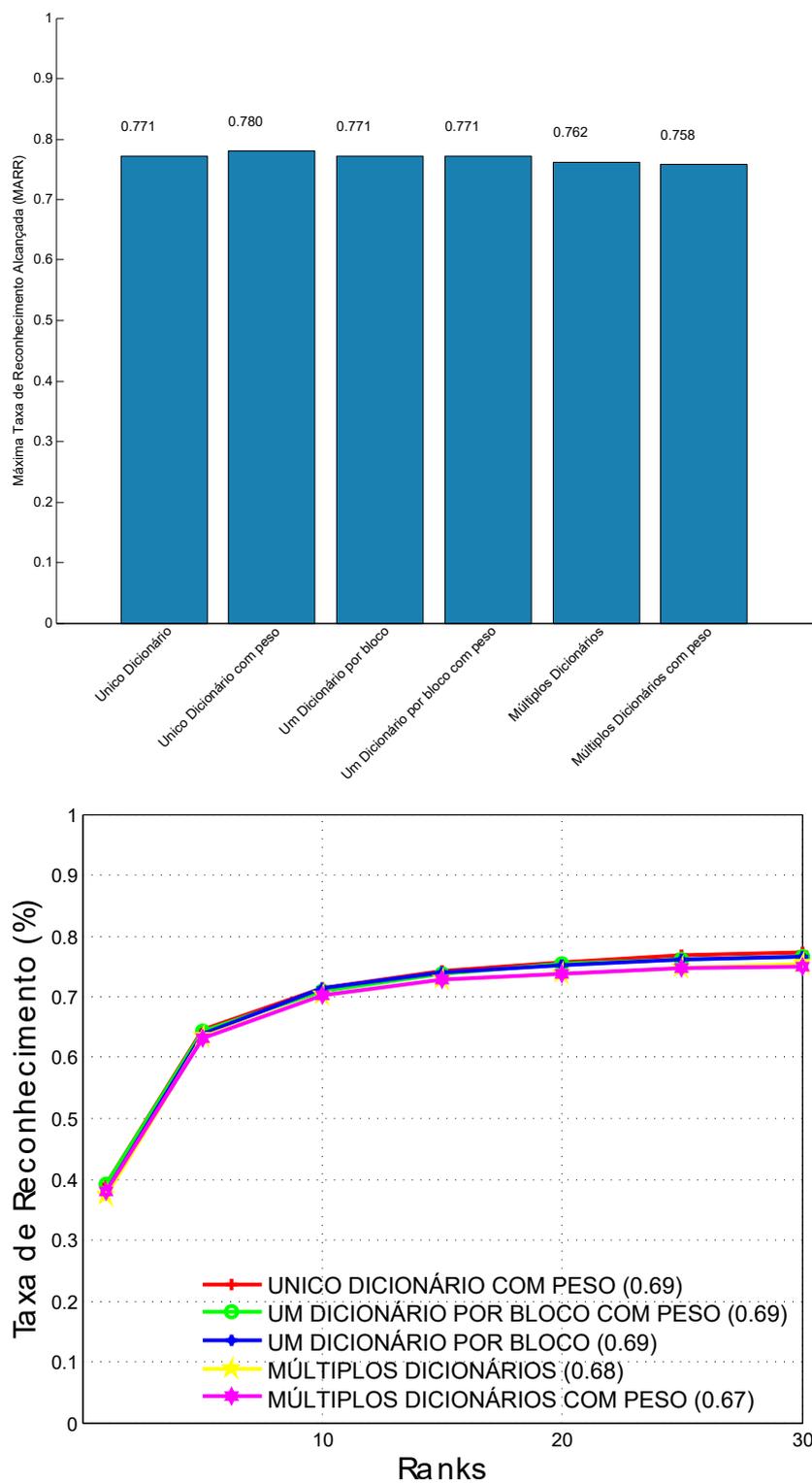


Figura 4.5: Gráfico de barras MARR e curva CMC mostrando os resultados alcançados pelos métodos de ponderação.

## 4.3 Indexação baseada em Partial Least Squares

Nesta seção são avaliados os resultados obtidos utilizando a regressão PLS. A seção 4.3.1 avalia a abordagem PLS-OAA usando dicionário visual. Os resultados da abordagem PLS-OAA usando permutação de indivíduos são exibidos na seção 4.3.2. A validação dos parâmetros das duas abordagens foi realizada utilizando WHOS como descritor de características, pois apresentou para este conjunto de testes, melhor resultado quando comparado ao descritor SCNCD.

### 4.3.1 PLS-OAA usando dicionário visual

Esta seção avalia o número de dimensões do PLS, o número de palavras visuais do dicionário visual, o número de dicionários, os descritores de características utilizados e, a taxa de reconhecimento final obtida.

**Número de dimensões do PLS** O PLS requer apenas um parâmetro, que é o número de dimensões no espaço latente. Realizou-se um experimento com o número de dimensões variando entre 1 e 9, com passo de 2. Os resultados são apresentados na figura 4.6, onde é possível notar que a melhor taxa de reconhecimento foi obtida utilizando 5 dimensões para o PLS-OAA.

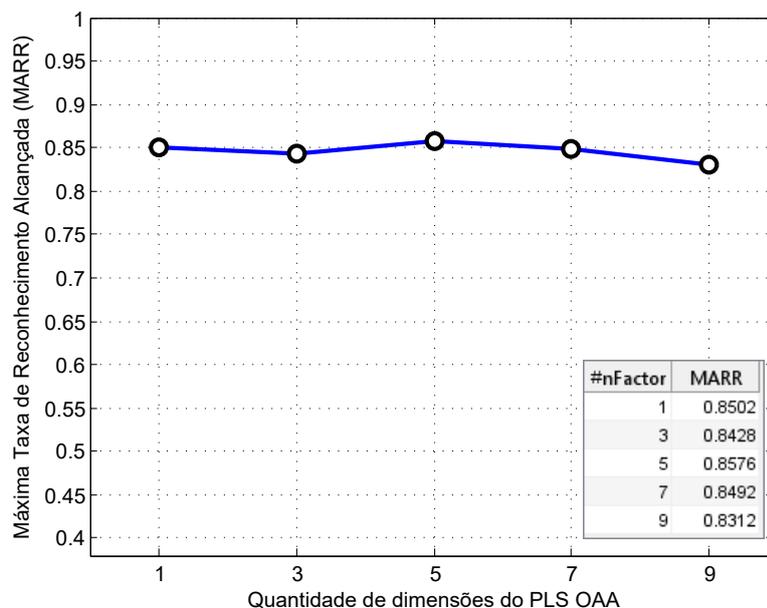


Figura 4.6: Avaliação do número de dimensões do PLS-OAA usando dicionário visual.

**Número de palavras visuais do dicionário** Para este experimento variou-se o número de palavras visuais entre 10 e 30 com incremento de 5. A figura 4.7 mostra que o melhor resultado foi alcançado utilizando 15 palavras visuais. Dessa forma, nos próximos experimentos o número de palavras visuais foi fixada em 15.

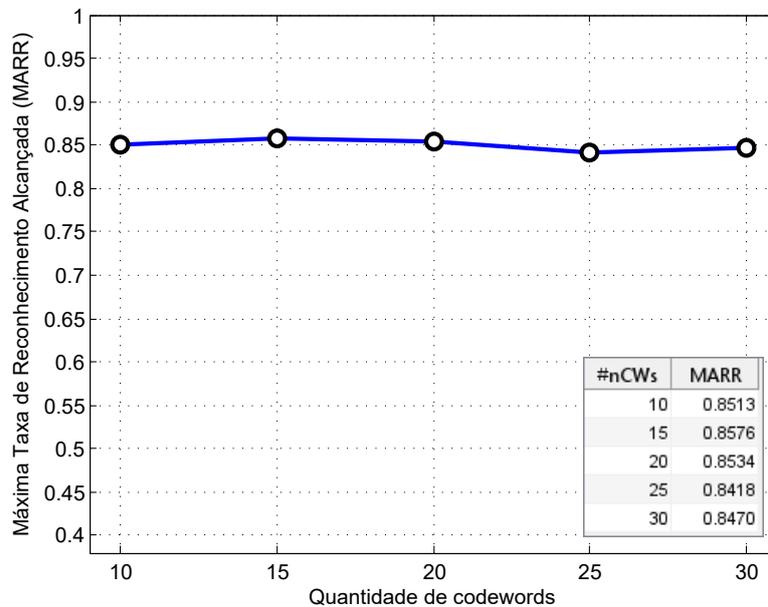


Figura 4.7: Avaliação do número de palavras visuais do dicionário utilizando PLS-OAA usando dicionário visual.

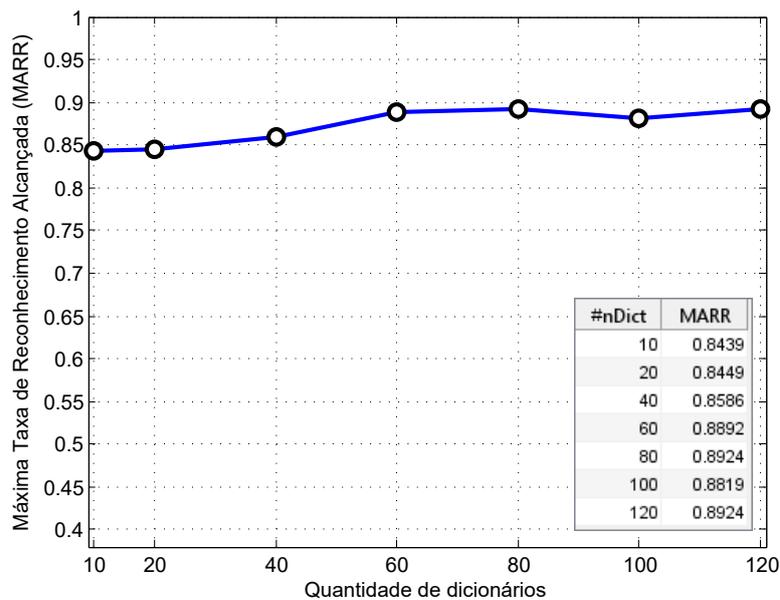


Figura 4.8: Avaliação do número necessário de dicionários utilizando o PLS-OAA usando dicionário visual.

**Número de dicionários** Para avaliar quantos dicionários são necessários para alcançar uma boa acurácia, variou-se o número de dicionários entre 10 e 120, obtendo melhor resultado ao utilizar 80 dicionários (figura 4.8).

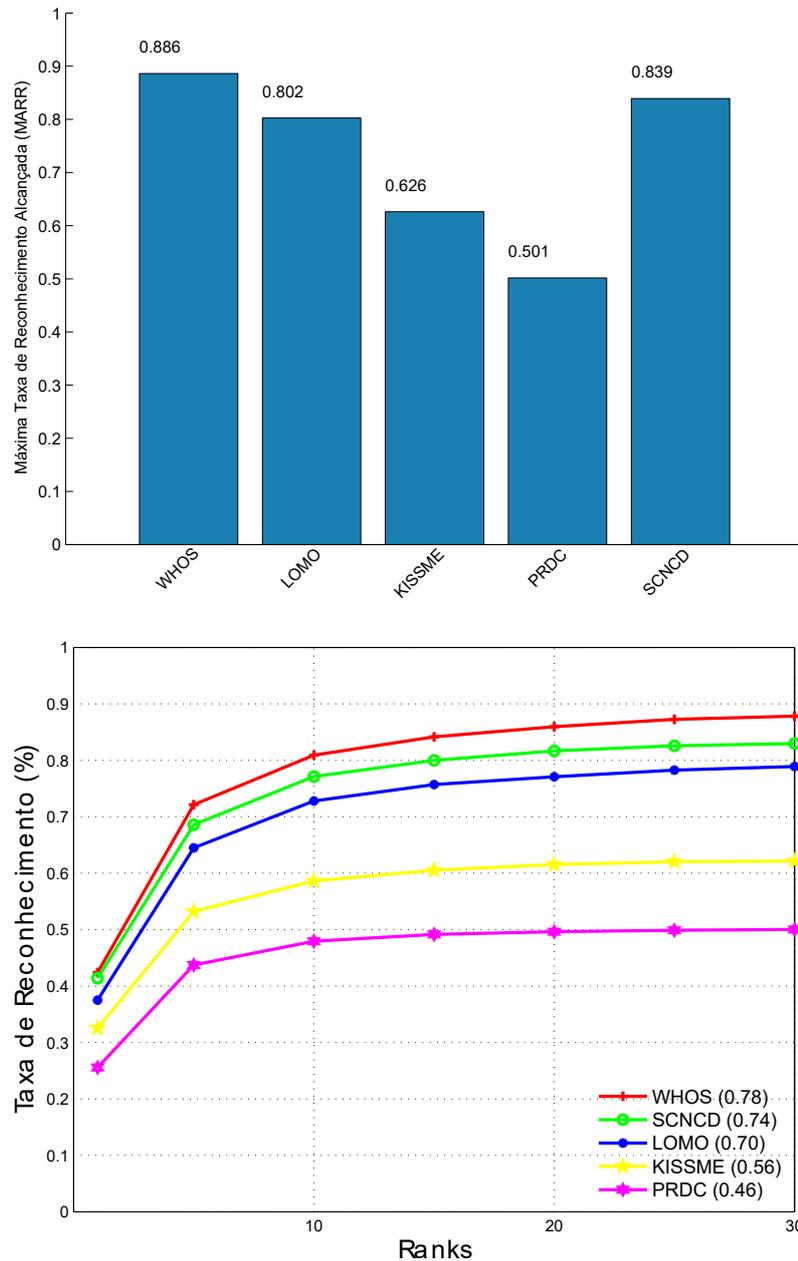


Figura 4.9: Gráfico de barras MARR e curva CMC mostrando os resultados alcançados pelo PLS-OAA usando dicionários variando os descritores de características utilizados.

**Descritores de características** A figura 4.9 apresenta o gráfico de barras MARR e a curva CMC para diferentes tipos de descritores de características quando 50% da lista

de candidatos é recuperada. Os descritores utilizados foram: WHOS, LOMO, KISSME, PRDC e SCNCD. Os outros descritores utilizados na seção 4.2 não foram usados, pois apresentaram resultados ruins nesta abordagem. O descritor WHOS apresentou o melhor resultado com MARR igual a 0.886.

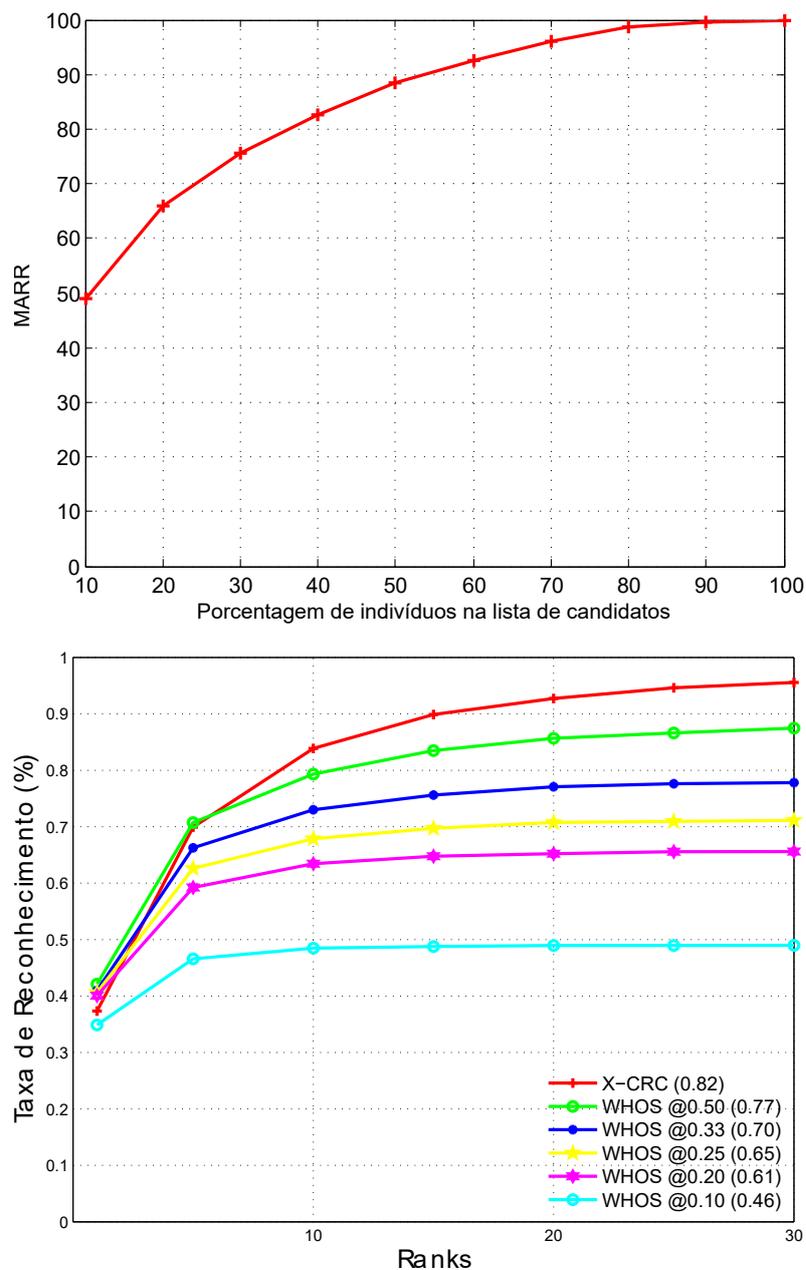


Figura 4.10: Curvas MARR e CMC mostrando os resultados alcançados pelo PLS-OAA usando dicionários para diferentes tamanhos da lista de candidatos recuperada.

**Resultado Final** Este experimento avalia a acurácia da abordagem ao variar o tamanho da lista de candidatos recuperada. Foi avaliado a taxa de reconhecimento ao se recuperar 10% (WHOS @10), 20% (WHOS @20), 25% (WHOS @25), 33% (WHOS @33) e 50% (WHOS @50) da lista original de candidatos. A figura 4.10 mostra a curva MARR e a taxa de reconhecimento alcançada. Os resultados mostram que o método, utilizando 20% da lista de candidatos, consegue superar a taxa de reconhecimento no rank-01 obtida pelo algoritmo X-CRC e que ao recuperar 40% da lista de candidatos, mais de 80% dos candidatos corretos se encontram nesta lista.

### 4.3.2 PLS-OAA usando permutação de indivíduos

Esta seção avalia os parâmetros de configuração do PLS-OAA usando permutação de indivíduos. Foi avaliado, o número de dimensões do PLS, o número de modelos utilizados, o número de subconjuntos em que a base de dados foi particionada e, a taxa de reconhecimento final alcançada.

**Número de dimensões do PLS** A figura 4.11 mostra a variação da taxa de reconhecimento em função do número de dimensões do PLS, novamente a variação foi de 1 a 9 com passo de 2. O melhor resultado foi obtido utilizando apenas uma dimensão.

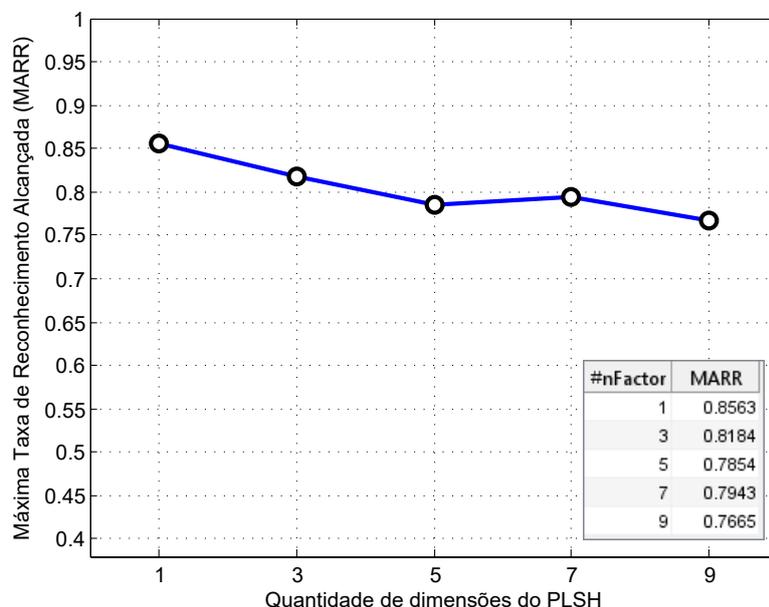


Figura 4.11: Avaliação do número de fatores utilizando o PLS-OAA usando permutação de indivíduos.

**Número de modelos** A figura 4.12 mostra a variação do número de modelos utilizados. Variou-se o número entre 10 e 150 com incremento de 20. O melhor resultado foi obtido utilizando 90 modelos.

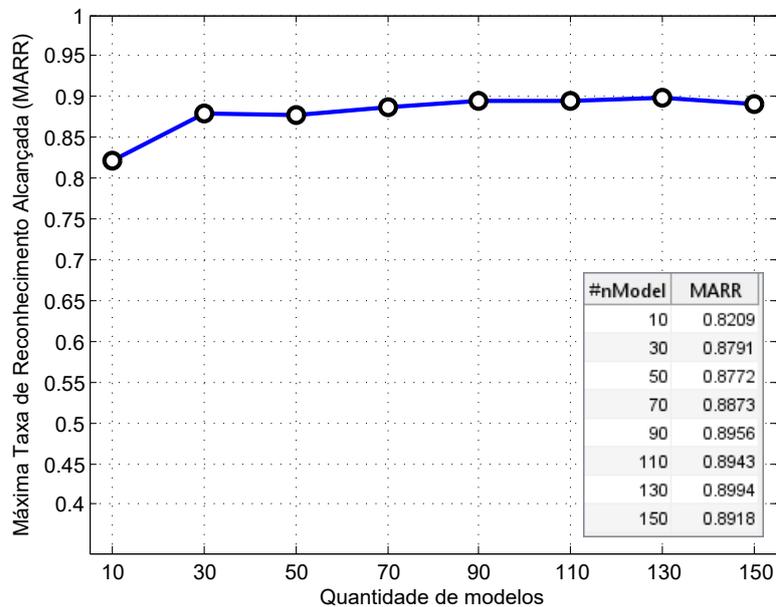


Figura 4.12: Avaliação do número de modelos utilizando o PLS-OAA usando permutação de indivíduos.

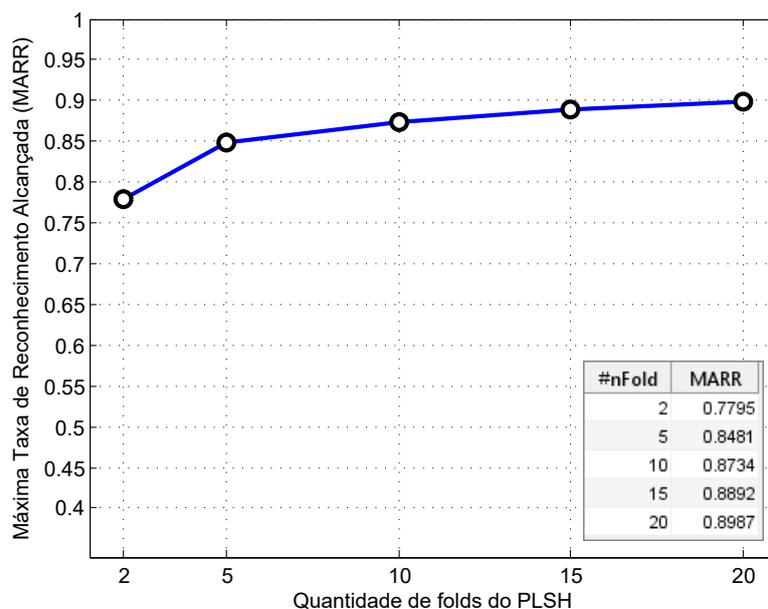


Figura 4.13: Avaliação do número de subconjuntos utilizando o PLS-OAA usando permutação de indivíduos.

**Número de subconjuntos** O resultado da variação do número de subconjuntos (2, 5, 10, 15, 20) é exibida na figura 4.13. Note que dividir a base em 20 subconjuntos gera o melhor resultado.

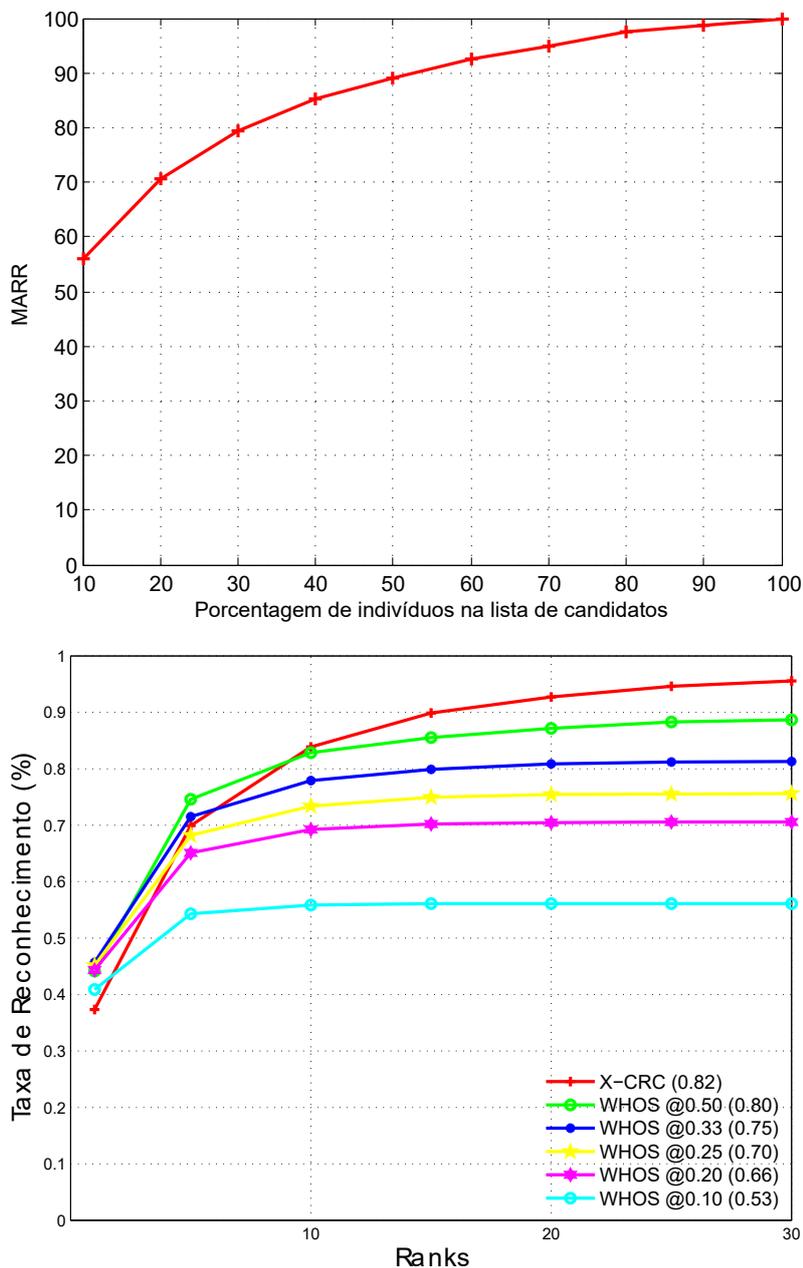


Figura 4.14: Curvas MARR e CMC obtidas utilizando o PLS-OAA usando permutação de indivíduos considerando diferentes porcentagens da lista de candidatos.

**Resultado final** Para avaliar a acurácia desta abordagem, variou-se o tamanho da lista de candidatos que foi apresentada ao algoritmo X-CRC. A figura 4.14 mostra

respectivamente as curvas MARR e CMC obtidas. Os gráficos mostram que o método alcança melhores resultados no rank-01 quando comparado ao algoritmo X-CRC, note que ao recuperar apenas 10% da lista de candidatos, a taxa de reconhecimento no rank-01 é igual a 40.9%, e que ao recuperar 30% da lista de candidatos, praticamente 80% dos candidatos corretos se encontram nesta lista.

## 4.4 Resultados utilizando a base de dados CUHK01

O resultado do método PLS-OAA usando permutação de indivíduos sobre a base de dados CUHK01 pode ser visto na figura 4.15. Nota-se que ao recuperar 20%, 25%, 33% ou 50% da lista de candidatos o resultado praticamente não é alterado. Observa-se também que o melhor resultado (rank=1) foi obtido utilizando apenas 10% da lista de candidatos e que ao recuperar 10% dos candidatos 80% das identidades corretas estão nesta lista.

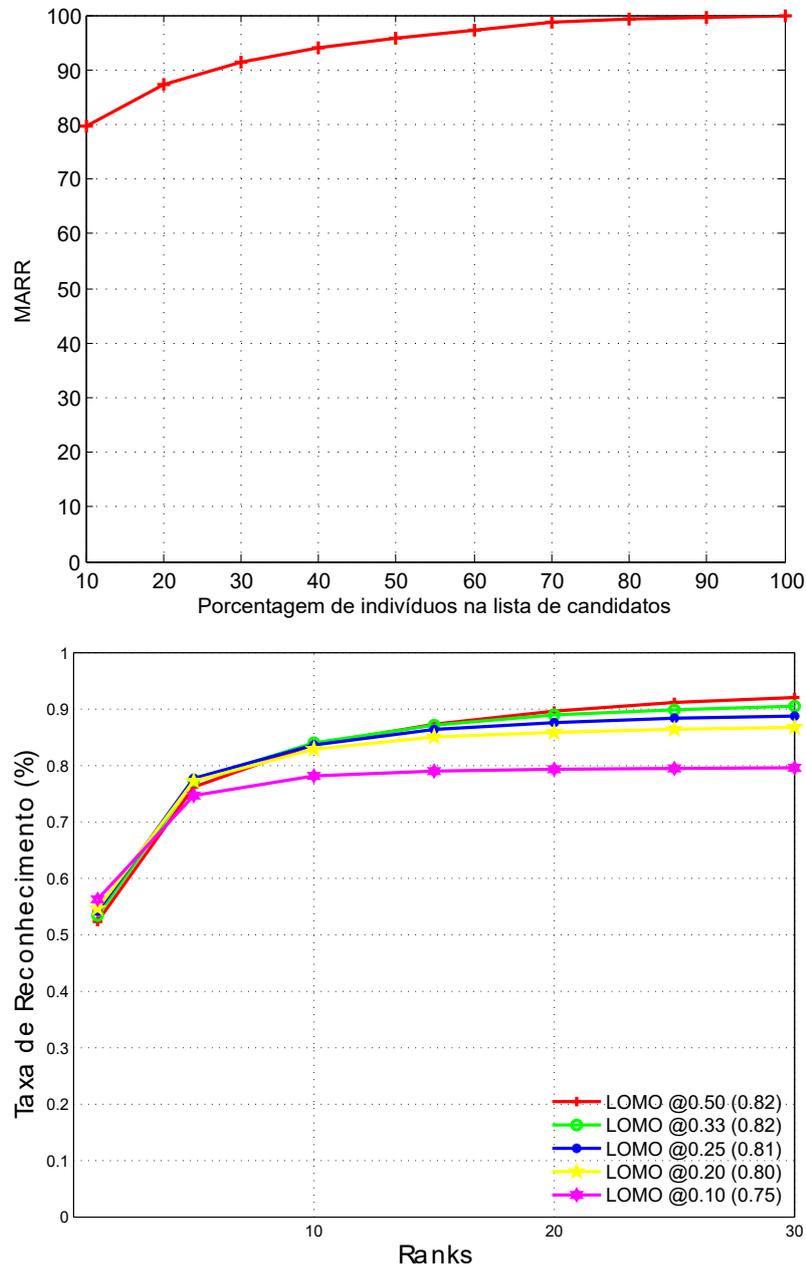


Figura 4.15: Curvas MARR e CMC mostrando os resultados alcançados pelo PLS-OAA usando permutação de indivíduos na base de dados CUHK01 para diferentes tamanhos da lista de candidatos recuperada.

# Capítulo 5

## Análise dos Resultados

Neste capítulo é realizada a análise dos resultados apresentados na capítulo 4. A análise está dividida nas seguintes partes: análise do custo computacional (seção 5.1), estabilidade dos resultados (seção 5.2) e comparação com o estado da arte (seção 5.3). Finalmente, na seção 5.4 são apresentados os resultados obtidos na base de dados CUHK01.

### 5.1 Análise do custo computacional

Para realizar a análise do custo computacional das abordagens de forma simples, calculou-se o número de operações relevantes que cada abordagem realiza para obter seus resultados. A primeira abordagem (Seção 4.2) faz uso da distância euclidiana para medir a similaridade entre as amostras, enquanto que a segunda abordagem (Seção 4.3) utiliza o PLS para estimar essa similaridade.

	Dicionário Visual	PLS-OAA (Dicionário)	PLS-OAA (Permutação)
Variáveis	Características(bins) = 448 Palavras visuais = 10 Número Dicionários = 1	Características = 2.960 Palavras visuais = 15 Número Dicionários = 80	Características = 2.960 Número Modelos = 90 Número Subconjuntos = 20
Quant. Operações relevantes @33% - Rank=01	$448 \times 10 \times 1 = 4.480$ 35.5	$2960 \times 15 \times 80 = 3.552.000$ 41.0	$2960 \times 90 \times 20 = 5.328.000$ 45.7

Tabela 5.1: Custo computacional de cada abordagem de acordo com o número de operações relevantes necessárias.

Sabe-se que o tempo gasto para fazer a comparação entre dois vetores calculando a distância euclidiana é diferente do tempo gasto para projetar um vetor no PLS. Para o cálculo da distância euclidiana são necessários operações de subtrações, somas e quadrado, já a projeção faz um produto interno entre dois vetores e cada produto

interno tem  $n$  multiplicações (onde  $n$  é o tamanho do vetor de características). Porém, para facilitar e padronizar o cálculo do número de operações, considerou-se que o tempo gasto para medir a similaridade nas duas abordagens é igual, ou seja, a multiplicação no produto interno tem o mesmo custo computacional que calcular a distância euclidiana de uma variável, e ambos equivalem a uma operação.

A tabela 5.1 mostra a quantidade de operações relevantes necessárias em cada abordagem, considerando os parâmetros que alcançaram a melhor resposta.

Observa-se que o PLS-OAA usando permutação possui o maior custo computacional, sendo necessário mais operações relevantes. Porém, como pode ser visto na tabela 5.2, o aumento no número de operações é compensado pelos resultados obtidos, comparáveis ao estado da arte. Por exemplo, o PLS-OAA com permutação consegue manter os resultados do estado da arte, e possui custo computacional igual a 5.328.000 operações, se fosse utilizado a abordagem com dicionário visual, o número de operações seria bem menor, 4.480, porém os resultados não são comparáveis aos melhores resultados do estado da arte.

## 5.2 Estabilidade dos resultados

A figura 5.1 exibe a média MARR e o desvio padrão ao executar dez vezes as três abordagens. Observa-se que os métodos baseados no PLS-OAA possuem maior estabilidade, que pode ser consequência da quantidade de modelos criados, o que reduz a variância das somas no histograma de votação, resultando em uma distribuição mais estável.

## 5.3 Comparação dos métodos com o estado da arte

	Dicionário Visual					PLS-OAA Dicionário					PLS-OAA Permutação				
	10%	20%	25%	33%	50%	10%	20%	25%	33%	50%	10%	20%	25%	33%	50%
rank 1	25.2	31.8	32.6	35.5	38.9	34.9	40.0	40.6	41.0	42.1	40.9	44.3	45.1	45.7	44.2
rank 5	36.1	47.4	50.9	56.3	64.4	46.6	59.2	62.6	66.3	70.7	54.3	65.1	68.2	71.5	74.6
rank 10	37.6	50.5	54.9	61.1	71.4	48.5	63.4	67.9	73.0	79.3	55.8	69.2	73.4	77.9	82.8
rank 20	38.0	51.9	56.7	64.0	75.2	49.0	65.2	70.8	77.1	85.7	56.1	70.4	75.4	80.8	87.1
rank 30	38.0	52.0	57.1	64.5	76.5	49.0	65.6	71.1	77.8	87.5	56.1	70.6	75.6	81.3	88.7

Tabela 5.2: Resultados obtidos nos principais ranks utilizando todas as abordagens e diferentes tamanhos da lista de candidatos.

A tabela 5.2 mostra os resultados obtidos pelas três abordagens implementadas sobre a base de dados VIPeR após a aplicação do algoritmo X-CRC utilizando como

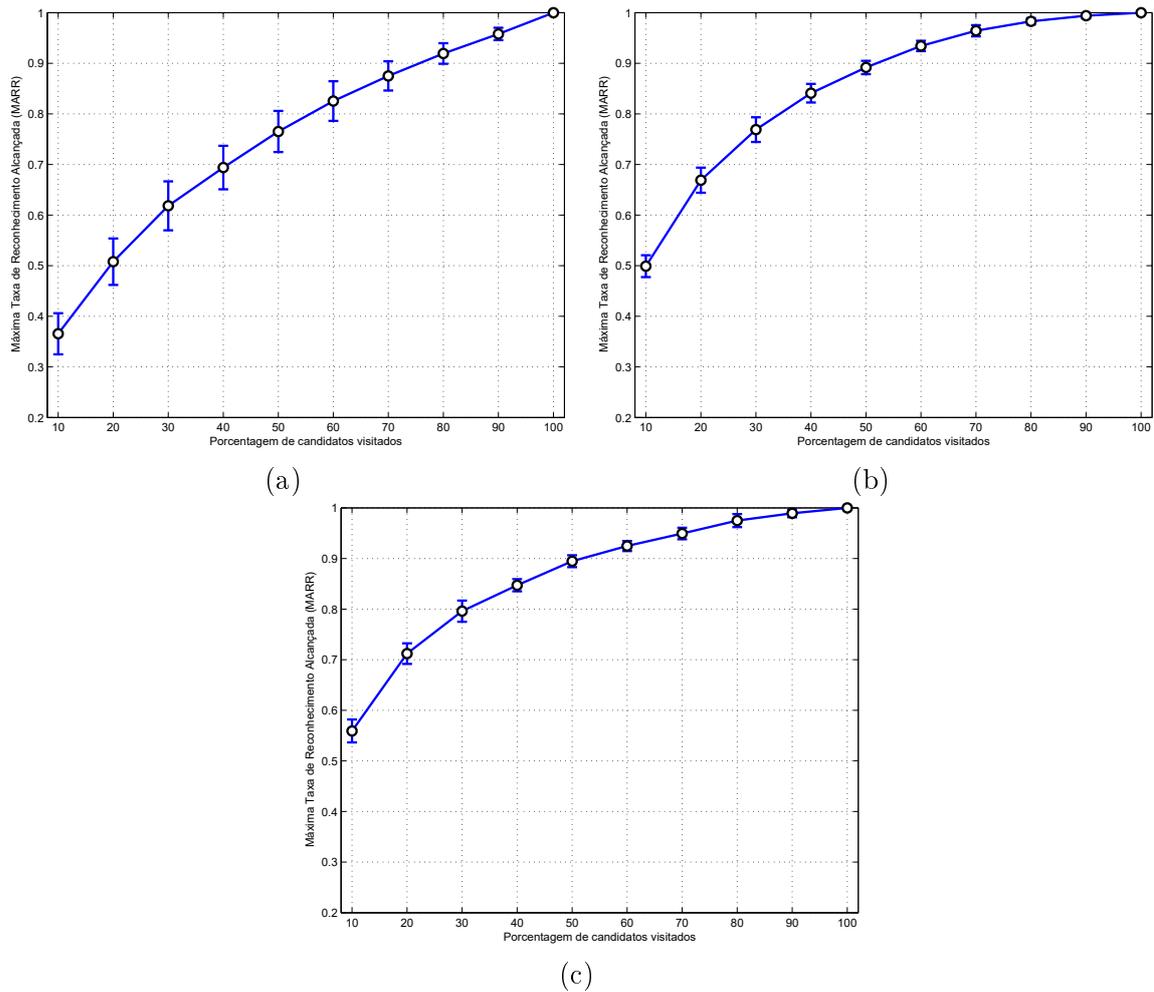


Figura 5.1: Média MARR e desvio padrão considerando 10 execuções das abordagens: (a) Indexação baseado dicionário visual, (b) Indexação usando PLS-OAA com dicionário visual e (c) Indexação usando PLS-OAA com permutação de indivíduos.

entrada para o algoritmo diferentes porcentagens da lista de candidatos original (10%, 20%, 25%, 33% e 50%). Nota-se que a abordagem baseada em dicionário visual, apesar de ser extremamente simples, obteve bons resultados no rank-1 quando comparado ao estado da arte (ver tabela 5.3). Por exemplo, reduzindo a lista de candidatos pela metade, o que significa utilizar 50%, a taxa de reconhecimento no rank-1 é igual a 38.9%, o que coloca este método na sexta posição da tabela 5.3 juntamente com o método ECM [Liu et al., 2015].

Ao observar os resultados obtidos pelos métodos baseados no PLS-OAA, concluímos que o PLS consegue reduzir a ambiguidade entre diferentes aparências e melhorar a taxa de reconhecimento alcançada pela abordagem baseada em dicionário visual. O PLS-OAA usando dicionário visual consegue diminuir a lista de candidatos 5 vezes, ou seja, utilizar apenas 20% da lista de candidatos, e ainda assim obter uma taxa de

Método	rank=1	rank=5	rank=10	rank=20	rank=30	Ano
<b>PLS-OAA Permutação @33%</b>	<b>45.7</b>	<b>71.5</b>	<b>77.9</b>	<b>80.8</b>	<b>81.3</b>	<b>2016</b>
<b>PLS-OAA Permutação @25%</b>	<b>45.1</b>	<b>68.2</b>	<b>73.4</b>	<b>75.4</b>	<b>75.6</b>	<b>2016</b>
[Paisitkriangkrai et al., 2015]	44.9	76.3	88.2	94.9	-	2015
<b>PLS-OAA Permutação @20%</b>	<b>44.3</b>	<b>65.1</b>	<b>69.2</b>	<b>70.4</b>	<b>70.6</b>	<b>2016</b>
<b>PLS-OAA Permutação @50%</b>	<b>44.2</b>	<b>74.6</b>	<b>82.8</b>	<b>87.1</b>	<b>88.7</b>	<b>2016</b>
[Chen et al., 2015b]	43.0	75.2	87.3	94.8	-	2015
[Shi et al., 2015]	41.6	71.9	86.2	95.1	-	2015
<b>PLS-OAA Permutação @10%</b>	<b>40.9</b>	<b>54.3</b>	<b>55.8</b>	<b>56.1</b>	<b>56.1</b>	<b>2016</b>
[Ding et al., 2015]	40.5	60.8	70.4	84.4	90.9	2015
<b>PLS-OAA Dicionário Visual @20%</b>	<b>40.0</b>	<b>59.2</b>	<b>63.4</b>	<b>65.2</b>	<b>65.6</b>	<b>2016</b>
LOMO+XQDA [Liao et al., 2015]	40.0	68.0	80.5	91.1	95.5	2015
<b>Dicionário Visual @50%</b>	<b>38.9</b>	<b>64.4</b>	<b>71.4</b>	<b>75.2</b>	<b>76.5</b>	<b>2016</b>
ECM [Liu et al., 2015]	38.9	67.8	78.4	88.9	-	2015
[Chen et al., 2015a]	38.4	69.2	81.3	90.4	-	2015
SCNCD [Yang et al., 2014b]	37.8	68.5	81.2	90.4	94.2	2014
[Yang et al., 2014a]	37.6	68.1	-	90.2	94.0	2014
X-CRC-[de Carvalho Prates & Schwartz, 2015](em revisão)	36.0	74.2	85.9	94.8	-	2015
RLML [Liong et al., 2015b]	35.3	67.4	81.6	90.8	-	2015
[Ye et al., 2015]	35.0	62.9	72.0	82.5	-	2015
[Shen et al., 2015]	34.8	68.7	82.3	91.8	94.9	2015
[Ahmed et al., 2015]	34.8	63.5	75.0	80.0	-	2015
DRML [Liong et al., 2015a]	34.2	-	78.8	90.2	-	2015
[Li et al., 2015]	34.0	64.2	77.5	88.6	-	2015
[d. C. Prates & Schwartz, 2015]	32.9	62.3	78.7	87.8	91.6	2015
[Xiong et al., 2014]	32.3	65.8	79.7	90.9	-	2014
[Zeng et al., 2015]	31.2	56.0	70.0	82.0	86.5	2015
[Ma et al., 2014]	31.1	58.3	70.7	82.4	89.9	2014
[Zhao et al., 2014c]	30.1	52.3	65.8	-	-	2013
LADF [Li et al., 2013b]	30.0	64.0	80.0	92.0	96.5	2013
[Zhang et al., 2014]	29.8	62.5	73.5	-	-	2014
Mid-Level Filters [Zhao et al., 2014a]	29.1	52.5	65.9	79.9	-	2014
Region-based Saliency [Geng et al., 2015]	28.5	50.6	64.6	76.3	-	2015
[Yi et al., 2014]	28.2	59.3	73.4	86.4	92.3	2014
KISSME [Kostinger et al., 2012]	27.0	55.0	70.0	83.0	89.5	2012
[Zhao et al., 2013b]	26.7	50.7	62.4	76.4	-	2013
LDFV [Ma et al., 2012]	26.5	56.4	70.9	84.6	-	2012
[Leng et al., 2015]	23.8	53.2	70.2	84.0	90.4	2015
EIML [Hirzer et al., 2012b]	22.0	47.5	63.0	78.0	87.0	2014
SDALF [Bazzani et al., 2013]	19.9	40.0	49.4	65.7	75.6	2010
PRDC [Zheng et al., 2011]	15.7	38.4	53.9	70.0	-	2011

Tabela 5.3: Tabela CMC com os principais resultados do estado da arte para a base de dados VIPeR (%). A tabela foi ordenada considerando a taxa de reconhecimento obtida no rank=1. Tabela disponível em <http://www.ssig.dcc.ufmg.br/reid-results/>.

reconhecimento igual a 40.0% no rank-1. Porém, os melhores resultados foram obtidos pelo PLS-OAA usando permutação de indivíduos (ver resultados na tabela 5.3). Ao selecionar 25% (@25%) ou 33% (@33%) da lista de candidatos, a taxa de reconhecimento ultrapassa os valores obtidos no estado da arte, respectivamente 45.1% e 45.7%, contra 44.9% obtido por Paisitkriangkrai et al. [2015]. Note também que, reduzindo a lista de candidatos 10 vezes, o que equivale a selecionar apenas 10% (@10%), é possível obter no rank-1 taxa de reconhecimento igual a 40.9%, valor este inferior à apenas três trabalhos no estado da arte ([Shi et al., 2015], [Chen et al., 2015b] e

[Paisitkriangkrai et al., 2015]). Note que o método PLS-OAA usando permutação de indivíduos também apresenta resultado superior ao alcançado pelo método X-CRC.

Uma dúvida que surge, é porque no rank-1 a resposta é tão boa? Vamos analisar a melhor resposta obtida, PLS-OAA usando permutação de indivíduos @33%. No gráfico MARR da figura 4.14, é possível ver que o PLS consegue colocar grande parte das identidades corretas nas primeiras posições da lista de candidatos. Ao utilizar 33% da lista de candidatos, a quantidade de candidatos corretos nesta lista é aproximadamente 80%. O algoritmo X-CRC é então utilizado para reordenar esta lista e arrastar as identidades corretas para a primeira posição da lista, nota-se que este objetivo é alcançado uma vez que a taxa de reconhecimento neste rank é a melhor do estado da arte.

Observando a tabela 5.2 é possível notar que a taxa de reconhecimento a partir do rank-5 aumenta de forma lenta e gradual, apresentando um desempenho inferior quando comparado aos primeiros métodos da tabela. Nota-se também que conforme mais candidatos são recuperados da lista, mais rápido a taxa de reconhecimento se aproxima de 100%. Observe a curva MARR na figura 4.14, quando 10% da lista de candidatos é recuperada pouco mais de 55% dos candidatos corretos estão nesta lista. Ao aplicar o algoritmo X-CRC nesta lista, os candidatos são reordenados de maneira que grande parte dos corretos ficam na primeira posição da lista de resposta. Porém para que a curva CMC seja plotada, é necessário, preencher a lista de candidatos com as identidades restantes. Esta lista é preenchida com as identidades restantes que não foram apresentadas ao X-CRC, e aproximadamente 45% destas identidades são corretas. Como podem estar em qualquer posição da lista, ao plotar a curva CMC a taxa de reconhecimento demora a convergir pois algumas das identidades corretas podem estar, por exemplo, na última posição da lista de respostas. Para resolver este problema, é necessário investigar outras técnicas e abordagens que consigam aumentar a quantidade de candidatos corretos dentro de uma pequena porcentagem da lista de candidatos.

## 5.4 Resultados sobre a base de dados CUHK01

A tabela 5.4 mostra os resultados obtidos sobre a base de dados CUHK01 recuperando diferentes tamanhos da lista de candidatos. O algoritmo utilizado foi o PLS-OAA usando permutação de indivíduos, pois apresentou o melhor desempenho sobre a base de dados VIPeR.

Com pode ser visto, o melhor resultado obtido no rank=1 utilizou apenas 10%

	10%	20%	25%	33%	50%
rank 1	56.3	54.7	54.2	53.4	52.4
rank 5	74.7	77.2	77.7	77.4	76.2
rank 10	78.2	82.9	83.7	84.0	83.7
rank 20	79.3	85.9	87.6	89.3	89.7
rank 30	79.6	86.8	88.8	90.5	92.0

Tabela 5.4: Resultado CUHK01

da lista de candidatos, alcançando uma taxa de reconhecimento igual a 56.3%. Observando o gráfico MARR da figura 4.15, podemos ver que ao recuperar 10% da lista de candidatos, 80% dos candidatos corretos se encontram nesta lista, e que a medida que se recupera mais candidatos, a quantidade de candidatos corretos adicionados é pequena. Dessa forma, podemos concluir que a taxa de reconhecimento diminui nos rank=1 à medida que o tamanho da lista de candidatos aumenta, porque muitos candidatos com identidade incorreta são adicionados nesta lista.

Método	rank=1	rank=5	rank=10	rank=20	rank=30	Ano
[Li et al., 2015]	59.5	81.3	89.7	93.1	-	2015
<b>PLS-OAA Permutação @10%</b>	<b>56.3</b>	<b>74.7</b>	<b>78.2</b>	<b>79.3</b>	<b>79.6</b>	<b>2016</b>
<b>PLS-OAA Permutação @20%</b>	<b>54.7</b>	<b>77.2</b>	<b>82.9</b>	<b>85.9</b>	<b>86.8</b>	<b>2016</b>
<b>PLS-OAA Permutação @25%</b>	<b>54.2</b>	<b>77.7</b>	<b>83.7</b>	<b>87.6</b>	<b>88.8</b>	<b>2016</b>
<b>PLS-OAA Permutação @33%</b>	<b>53.4</b>	<b>77.4</b>	<b>84.0</b>	<b>89.3</b>	<b>90.5</b>	<b>2016</b>
<b>PLS-OAA Permutação @50%</b>	<b>52.4</b>	<b>76.2</b>	<b>83.7</b>	<b>89.7</b>	<b>92.0</b>	<b>2016</b>
[Paisitkriangkrai et al., 2015]	51.9	75.1	83.0	89.4	-	2015
[Chen et al., 2015a]	50.4	75.9	84.0	91.3	-	2015
[Ahmed et al., 2015]	47.5	71.0	80.0	-	-	2015
X-CRC-[de Carvalho Prates & Schwartz, 2015](em revisão)	42.3	69.9	78.3	86.5	-	2015
[Chen et al., 2015b]	40.4	64.6	75.3	84.1	-	2015
Mid-Level Filters [Zhao et al., 2014a]	34.3	55.1	65.0	74.9	80.3	2014
[Zhao et al., 2014c]	28.4	45.8	55.7	67.9	74.9	2013
[Li et al., 2013a]	20.0	43.5	56.0	69.3	77.3	2012
SDALF [Bazzani et al., 2013]	9.9	22.6	30.3	41.0	-	2010

Tabela 5.5: Tabela CMC com os principais resultados do estado da arte para a base de dados CUHK01 (%). Tabela disponível em <http://www.ssig.dcc.ufmg.br/reid-results/>.

A tabela 5.5 compara o resultado alcançado com os resultados do estado da arte, lembrando que foi utilizado para treinamento apenas uma das duas amostras disponíveis por indivíduo. É possível ver que utilizando 10% da lista de candidatos, o resultado é o segundo melhor obtido no estado da arte para esta base de dados, o melhor resultado é alcançado por Li et al. [2015].

## Capítulo 6

# Conclusão e Trabalhos Futuros

Neste trabalho, abordamos o problema de reidentificação de pessoas propondo um método de filtragem que utiliza uma estrutura de indexação de dados que faz uso de listas invertidas e duas abordagens distintas, uma usando dicionário visual e a outra usando *Partial Least Squares* em conjunto com o esquema de classificação um-contra-todos. A estrutura de indexação contém um número reduzido de identidades mais prováveis para um candidato, o que possibilita obter uma indexação mais rápida. O algoritmo PLS um-contra-todos consegue reduzir a ambiguidade entre as aparências e melhorar a acurácia. Os resultados obtidos demonstram que o método de filtragem proposto é acurado e que precisa considerar apenas um pequeno subconjunto de potenciais candidatos quando comparado a outros trabalhos no estado da arte.

Como direções futuras, pretendemos investigar outras técnicas que consigam reduzir o número de potenciais candidatos, de maneira, que todos estejam nas primeiras posições da lista de candidatos e assim aumentar a acurácia do ranking. Além disso, pretende-se analisar a forma como a informação obtida na indexação pode ser empregada nas outras fases para aumentar não apenas a eficiência, mas também a taxa de reconhecimento.



# Referências Bibliográficas

- Ahmed, E.; Jones, M. & Marks, T. K. (2015). An improved deep learning architecture for person re-identification. Em *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Arandjelovic, R. & Zisserman, A. (2012). Three things everyone should know to improve object retrieval. Em *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2911–2918. ISSN 1063-6919.
- Bak, S.; Corvée, E.; Brémond, F. & Thonnat, M. (2012). Boosted human re-identification using riemannian manifolds. *Image and Vision Computing*, 30(6):443 – 452.
- Bak, S.; Corvée, E.; Brémond, F. & Thonnat, M. (2010). Person re-identification using spatial covariance regions of human body parts. Em *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pp. 435–440.
- Bazzani, L.; Cristani, M. & Murino, V. (2013). Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 117(2):130–144.
- Bedagkar-Gala, A.; Shah, S. K.; Bedagkar-Gala, A. & Shah, S. (2014). Editor’s choice article. *Image and Vision Computing*, 32(4):270–286.
- Ben Shitrit, H.; Berclaz, J.; Fleuret, F. & Fua, P. (2011). Tracking multiple people under global appearance constraints. Em *Computer Vision (ICCV), 2011 IEEE International Conference on*, pp. 137–144.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32. ISSN 1573-0565.
- Cai, Y. & Pietikäinen, M. (2011). Person re-identification based on global color context. Em *Proceedings of the 2010 International Conference on Computer Vision - Volume Part I, ACCV’10*, pp. 205–215, Berlin, Heidelberg. Springer-Verlag.

- Chávez, E.; Navarro, G.; Baeza-Yates, R. & Marroquín, J. L. (2001). Searching in metric spaces. *Computing Surveys (CSUR), ACM Transactions on*, 33(3):273–321.
- Chen, S.; Guo, C. & Lai, J. (2015a). Deep ranking for person re-identification via joint representation learning. *CoRR*, abs/1505.06821.
- Chen, Y.-C.; Zheng, W.-S. & Lai, J. (2015b). Mirror representation for modeling view-specific transform in person re-identification. Em Yang, Q. & Wooldridge, M., editores, *IJCAI*, pp. 3402–3408. AAAI Press.
- Cheng, D. & Cristani, M. (2014). Person re-identification by articulated appearance matching. Em Gong, S.; Cristani, M.; Yan, S. & Loy, C. C., editores, *Person Re-Identification, Advances in Computer Vision and Pattern Recognition*, pp. 139–160. Springer London.
- Cheng, D. S.; Cristani, M.; Stoppa, M.; Bazzani, L. & Murino, V. (2011). Custom pictorial structures for re-identification. Em *Proceedings of the British Machine Vision Conference*, pp. 68.1–68.11. BMVA Press. <http://dx.doi.org/10.5244/C.25.68>.
- Cong, D.-N. T.; Khoudour, L.; Achard, C.; Meurie, C. & Lezoray, O. (2010). People re-identification by spectral classification of silhouettes. *Signal Processing*, 90(8):2362 – 2374. ISSN 0165-1684. Special Section on Processing and Analysis of High-Dimensional Masses of Image and Signal Data.
- Csurka, G.; Dance, C. R.; Fan, L.; Willamowski, J. & Bray, C. (2004). Visual categorization with bags of keypoints. Em *In Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1--22.
- d. C. Prates, R. F. & Schwartz, W. R. (2015). Appearance-based person re-identification by intra-camera discriminative models and rank aggregation. Em *Biometrics (ICB), 2015 International Conference on*, pp. 65–72.
- D’angelo, A. & Dugelay, J.-L. (2011). People re-identification in camera networks based on probabilistic color histograms. Em *3DIP 2011, Electronic Imaging Conference on 3D Image Processing and Applications, Vol. 7882, 23-27 January, 2011, San Francisco, CA, USA*, San Francisco, ÉTATS-UNIS.
- de Carvalho Prates, R. F. & Schwartz, W. R. (2015). Extending the collaborative representation based classification to person re-identification. Em *revisão*.

- de Paulo Carlos, G.; Pedrini, H. & Schwartz, W. R. (2015). Classification schemes based on partial least squares for face identification. *Journal of Visual Communication and Image Representation*, 32:170 – 179. ISSN 1047-3203.
- Ding, S.; Lin, L.; Wang, G. & Chao, H. (2015). Deep feature learning with relative distance comparison for person re-identification. *Pattern Recognition*, 48(10):2993 – 3003. ISSN 0031-3203. Discriminative Feature Learning from Big Data for Visual Recognition.
- D’Orazio, T. & Cicirelli, G. (2012). People re-identification and tracking from multiple cameras: A review. Em *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pp. 1601–1604.
- Dutra, C.; Schwartz, W.; Souza, T.; Alves, R. & Oliveira, L. (2013). Re-identifying people based on indexing structure and manifold appearance modeling. Em *Graphics, Patterns and Images (SIBGRAPI), 2013 26th SIBGRAPI - Conference on*, pp. 218–225. ISSN 1530-1834.
- Dutra, C. R. S.; Rocha, M. C. & Schwartz, W. R. (2014). Person re-identification based on weighted indexing structures. Em Bayro-Corrochano, E. & Hancock, E., editores, *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications: 19th Iberoamerican Congress, CIARP 2014, Puerto Vallarta, Mexico, November 2-5, 2014. Proceedings*, pp. 359–366, Cham. Springer International Publishing.
- Farenzena, M.; Bazzani, L.; Perina, A.; Murino, V. & Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. Em *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pp. 2360–2367.
- Gan, G.; Ma, C. & Wu, J. (2007). *Appendix B: The kd-Tree Data Structure*, capítulo 22, pp. 375–376. Siam.
- Geng, Y.; Hu, H.-M.; Zeng, G. & Zheng, J. (2015). A person re-identification algorithm by exploiting region-based feature salience. *Journal of Visual Communication and Image Representation*, 29:89 – 102. ISSN 1047-3203.
- Gevers, T. & Smeulders, A. W. (1999). Color-based object recognition. *Pattern Recognition*, 32(3):453 – 464. ISSN 0031-3203.
- Gilbert, A. & Bowden, R. (2006). Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. Em Leonardis, A.; Bischof, H. & Pinz, A., editores, *Computer Vision ECCV 2006*, volume 3952 of *Lecture Notes in Computer Science*, pp. 125–136. Springer Berlin Heidelberg.

- Gong, S.; Cristani, M.; Loy, C. & Hospedales, T. (2014). The re-identification challenge. Em Gong, S.; Cristani, M.; Yan, S. & Loy, C. C., editores, *Person Re-Identification*, Advances in Computer Vision and Pattern Recognition, pp. 1–20. Springer London.
- Gray, D.; Brennan, S. & Tao, H. (2007). Evaluating appearance models for recognition, reacquisition, and tracking. Em *10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS)*.
- Hirzer, M.; Beleznai, C.; Kostinger, M.; Roth, P. & Bischof, H. (2012a). Dense appearance modeling and efficient learning of camera transitions for person re-identification. Em *Image Processing (ICIP), 2012 19th IEEE International Conference on*, pp. 1617–1620.
- Hirzer, M.; Roth, P. M. & Bischof, H. (2012b). Person re-identification by efficient impostor-based metric learning. Em *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, pp. 203–208.
- Jégou, H. & Chum, O. (2012). Negative evidences and co-occurrences in image retrieval: The benefit of pca and whitening. Em Fitzgibbon, A.; Lazebnik, S.; Perona, P.; Sato, Y. & Schmid, C., editores, *Computer Vision ECCV 2012*, volume 7573 of *Lecture Notes in Computer Science*, pp. 774–787. Springer Berlin Heidelberg.
- Jégou, H.; Douze, M. & Schmid, C. (2009). On the burstiness of visual elements. Em *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pp. 1169–1176. ISSN 1063-6919.
- Jr., C. E. S.; Kijak, E.; Gravier, G. & Schwartz, W. R. (2015). Learning to hash faces using large feature vectors. Em *Content-Based Multimedia Indexing (CBMI), 13th International Workshop on*, pp. 1–6. IEEE.
- Junior, C. E. d. S. & Schwartz, W. R. (2014). Extending face identification to open-set face recognition. Em *Proceedings of the 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, SIBGRAPI '14*, pp. 188–195, Washington, DC, USA. IEEE Computer Society.
- Kenk, V. S.; Mandeljc, R.; Kovačič, S.; Kristan, M.; Hajdinjak, M. & Perš, J. (2015). Visual re-identification across large, distributed camera networks. *Image and Vision Computing*, 34:11 – 26. ISSN 0262-8856.
- Keval, H. (2006). Cctv control room collaboration and communication: Does it work?

- Knuth, D. (1997). Retrieval on secondary keys. Em *The Art of Computer Programming*. Addison-Wesley, Reading, Massachusetts.
- Kostinger, M.; Hirzer, M.; Wohlhart, P.; Roth, P. & Bischof, H. (2012). Large scale metric learning from equivalence constraints. Em *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2288–2295.
- Leng, Q.; Hu, R.; Liang, C.; Wang, Y. & Chen, J. (2015). Person re-identification with content and context re-ranking. *Multimedia Tools and Applications*, 74(17):6989–7014. ISSN 1380-7501.
- Li, J. & Allinson, N. M. (2008). A comprehensive review of current local features for computer vision. *Neurocomputing*, 71(10):1771–1787.
- Li, S.; Shao, M. & Fu, Y. (2015). Cross-view projective dictionary learning for person re-identification. Em *IJCAI - International Joint Conference on Artificial Intelligence*.
- Li, W.; Zhao, R. & Wang, X. (2013a). Human reidentification with transferred metric learning. Em Lee, K.; Matsushita, Y.; Rehg, J. & Hu, Z., editores, *Computer Vision ACCV 2012*, volume 7724 of *Lecture Notes in Computer Science*, pp. 31–44. Springer Berlin Heidelberg.
- Li, W.; Zhao, R.; Xiao, T. & Wang, X. (2014). Deepreid: Deep filter pairing neural network for person re-identification. Em *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 152–159.
- Li, Z.; Chang, S.; Liang, F.; Huang, T. S.; Cao, L. & Smith, J. R. (2013b). Learning locally-adaptive decision functions for person verification. Em *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 3610–3617.
- Liao, S.; Hu, Y.; Zhu, X. & Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. Em *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 2197–2206.
- Liong, V. E.; Ge, Y. & Lu, J. (2015a). Discriminative regularized metric learning for person re-identification. Em *Biometrics (ICB), 2015 International Conference on*, pp. 52–57.
- Liong, V. E.; Lu, J. & Ge, Y. (2015b). *Computer Vision - ECCV 2014 Workshops: Zurich, Switzerland, September 6-7 and 12, 2014, Proceedings, Part III*, capítulo

- Regularized Bayesian Metric Learning for Person Re-identification, pp. 209--224. Springer International Publishing, Cham.
- Lisanti, G.; Masi, I.; Bagdanov, A. & Del Bimbo, A. (2014). Person re-identification by iterative re-weighted sparse ranking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PP(99):1–1. ISSN 0162-8828.
- Liu, B.; Xia, Y. & Yu, P. S. (2000). Clustering through decision tree construction. Em *Proceedings of the Ninth International Conference on Information and Knowledge Management*, CIKM '00, pp. 20--29, New York, NY, USA. ACM.
- Liu, C.; Gong, S. & Loy, C. C. (2014a). On-the-fly feature importance mining for person re-identification. *Pattern Recogn.*, 47(4):1602--1615. ISSN 0031-3203.
- Liu, C.; Gong, S.; Loy, C. C. & Lin, X. (2012). Person re-identification: What features are important? Em *Proceedings of the 12th International Conference on Computer Vision - Volume Part I, ECCV'12*, pp. 391--401, Berlin, Heidelberg. Springer-Verlag.
- Liu, L.; Lu, X.; Yuan, Y. & Li, X. (2014b). Person re-identification by bidirectional projection. Em *Proceedings of International Conference on Internet Multimedia Computing and Service*, ICIMCS '14, pp. 1:1--1:5, New York, NY, USA. ACM.
- Liu, X.; Wang, H.; Wu, Y.; Yang, J. & Yang, M. H. (2015). An ensemble color model for human re-identification. Em *Applications of Computer Vision (WACV), 2015 IEEE Winter Conference on*, pp. 868–875.
- Ma, B.; Su, Y. & Jurie, F. (2012). *Computer Vision – ECCV 2012. Workshops and Demonstrations: Florence, Italy, October 7-13, 2012, Proceedings, Part I*, capítulo Local Descriptors Encoded by Fisher Vectors for Person Re-identification, pp. 413--422. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Ma, B.; Su, Y. & Jurie, F. (2014). Covariance descriptor based on bio-inspired features for person re-identification and face verification. *Image and Vision Computing*, 32(6):379 – 390. ISSN 0262-8856.
- Mazzon, R.; Tahir, S. F. & Cavallaro, A. (2012). Person re-identification in crowd. *Pattern Recognition Letters*, 33(14):1828 – 1837. Novel Pattern Recognition-Based Methods for Re-identification in Biometric Context.
- Mikolajczyk, K. & Schmid, C. (2005). A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630. ISSN 0162-8828.

- Paisitkriangkrai, S.; Shen, C. & van den Hengel, A. (2015). Learning to rank in person re-identification with metric ensembles. *CoRR*, abs/1503.01543.
- Park, U.; Jain, A.; Kitahara, I.; Kogure, K. & Hagita, N. (2006). Vise: Visual search engine using multiple networked cameras. Em *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pp. 1204–1207.
- Pedagadi, S.; Orwell, J.; Velastin, S. & Boghossian, B. (2013). Local fisher discriminant analysis for pedestrian re-identification. Em *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pp. 3318–3325. ISSN 1063-6919.
- Prado, G. L.; Schwartz, W. R. & Pedrini, H. (2013). Person re-identification using partial least squares appearance modeling. Em *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, pp. 382–390.
- Rosipal, R. & Krämer, N. (2006). Overview and recent advances in partial least squares. Em Saunders, C.; Grobelnik, M.; Gunn, S. & Shawe-Taylor, J., editores, in *Subspace, Latent Structure and Feature Selection*, volume 3940 of *Lecture Notes in Computer Science*, pp. 34–51. Springer Berlin Heidelberg.
- Roth, P.; Hirzer, M.; Köstinger, M.; Beleznai, C. & Bischof, H. (2014). Mahalanobis distance learning for person re-identification. Em Gong, S.; Cristani, M.; Yan, S. & Loy, C. C., editores, *Person Re-Identification*, Advances in Computer Vision and Pattern Recognition, pp. 247–267. Springer London, London, United Kingdom.
- Satta, R. (2013). Appearance descriptors for person re-identification: a comprehensive review. *CoRR*, abs/1307.5748.
- Schwartz, W. & Davis, L. (2009). Learning discriminative appearance-based models using partial least squares. Em *XXII Brazilian Symposium on Computer Graphics and Image Processing*, pp. 322–329, Rio de Janeiro, Brazil.
- Schwartz, W.; Guo, H.; Choi, J. & Davis, L. (2012). Face identification using large feature sets. *Image Processing, IEEE Transactions on*, 21(4):2245–2255. ISSN 1057-7149.
- Schwartz, W.; Guo, H. & Davis, L. (2010). A robust and scalable approach to face identification. Em Daniilidis, K.; Maragos, P. & Paragios, N., editores, *Computer Vision ECCV 2010*, volume 6316 of *Lecture Notes in Computer Science*, pp. 476–489. Springer Berlin Heidelberg.

- Schwartz, W.; Kembhavi, A.; Harwood, D. & Davis, L. (2009). Human Detection Using Partial Least Squares Analysis. Em *Proceedings of the International Conference on Computer Vision (CVPR)*.
- Schwartz, W. R. (2012a). Aplicação de observação de pessoas em computação forense. Em *Workshop de Forense Computacional (WFC) in SBSEG 2012 (XII Simpósio Brasileiro em Segurança da Informação e de Sistemas Computacionais)*, pp. 1--14.
- Schwartz, W. R. (2012b). Scalable people re-identification based on a one-against-some classification scheme. Em *IEEE International Conference on Image Processing*.
- Sharma, A. & Jacobs, D. W. (2011). Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. Em *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition, CVPR '11*, pp. 593--600, Washington, DC, USA. IEEE Computer Society.
- Shen, Y.; Lin, W.; Yan, J.; Xu, M.; Wu, J. & Wang, J. (2015). Person re-identification with correspondence structure learning. Em *The IEEE International Conference on Computer Vision (ICCV)*.
- Shi, Z.; Hospedales, T. M. & Xiang, T. (2015). Transferring a semantic representation for person re-identification and search. Em *Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on*, pp. 4184--4193.
- Sivic, J. & Zisserman, A. (2003). Video google: a text retrieval approach to object matching in videos. Em *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, volume 2, pp. 1470--1477.
- Song, M.; Tao, D. & Maybank, S. J. (2013). Sparse camera network for visual surveillance – A comprehensive survey. *CoRR*, abs/1302.0446.
- Sugiyama, M. (2006). Local fisher discriminant analysis for supervised dimensionality reduction. Em *Proceedings of the 23rd International Conference on Machine Learning, ICML '06*, pp. 905--912, New York, NY, USA. ACM.
- Sun, C.; Arr, G.; Ramachandran, R. & Ritchie, S. (2004). Vehicle reidentification using multidetector fusion. *Intelligent Transportation Systems, IEEE Transactions on*, 5(3):155--164.
- Tu, P.; Doretto, G.; Krahnstoeber, N.; Perera, A. G. A.; Wheeler, F.; Liu, X.; Rittscher, J.; Sebastian, T.; Yu, T. & Harding, K. (2007). An intelligent video framework for

- homeland protection. Relatório técnico, GE Global Research, Niskayuna, NY, USA. Visualization and Computer Vision Laboratory.
- Vezzani, R.; Baltieri, D. & Cucchiara, R. (2013). People reidentification in surveillance and forensics: A survey. *ACM Comput. Surv.*, 46(2):29:1--29:37.
- Wold, H. (1985). Partial least squares. Em Kotz, S. & Johnson, N., editores, *Encyclopedia of Statistical Science*, pp. 581–591. New York: Wiley.
- Xiong, F.; Gou, M.; Camps, O. & Sznaiar, M. (2014). *Computer Vision – ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part VII*, capítulo Person Re-Identification Using Kernel-Based Metric Learning Methods, pp. 1–16. Springer International Publishing, Cham.
- Yang, Y.; Liao, S.; Lei, Z.; Yi, D. & Li, S. Z. (2014a). Color models and weighted covariance estimation for person re-identification. Em *22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24-28, 2014*, pp. 1874--1879.
- Yang, Y.; Yang, J.; Yan, J.; Liao, S.; Yi, D. & Li, S. (2014b). Salient color names for person re-identification. Em Fleet, D.; Pajdla, T.; Schiele, B. & Tuytelaars, T., editores, *Computer Vision ECCV 2014*, volume 8689 of *Lecture Notes in Computer Science*, pp. 536–551. Springer International Publishing.
- Ye, M.; Chen, J.; Leng, Q.; Liang, C.; Wang, Z. & Sun, K. (2015). *MultiMedia Modeling: 21st International Conference, MMM 2015, Sydney, NSW, Australia, January 5-7, 2015, Proceedings, Part I*, capítulo Coupled-View Based Ranking Optimization for Person Re-identification, pp. 105--117. Springer International Publishing, Cham.
- Yi, D.; Lei, Z. & Li, S. Z. (2014). Deep metric learning for practical person re-identification. *CoRR*, abs/1407.4979.
- Zeng, M.; Wu, Z.; Tian, C.; Zhang, L. & Hu, L. (2015). Efficient person re-identification by hybrid spatiogram and covariance descriptor. Em *Computer Vision and Pattern Recognition Workshops (CVPRW), 2015 IEEE Conference on*, pp. 48–56.
- Zhang, L.; Yang, M.; Feng, X.; Ma, Y. & Zhang, D. (2012). Collaborative representation based classification for face recognition. *CoRR*, abs/1204.2358.
- Zhang, Z.; Chen, Y. & Saligrama, V. (2014). A novel visual word co-occurrence model for person re-identification. *CoRR*, abs/1410.6532.

- Zhao, R.; Ouyang, W. & Wang, X. (2013a). Person re-identification by salience matching. Em *IEEE International Conference on Computer Vision (ICCV)*, Sydney, Australia.
- Zhao, R.; Ouyang, W. & Wang, X. (2013b). Unsupervised salience learning for person re-identification. Em *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Portland, USA.
- Zhao, R.; Ouyang, W. & Wang, X. (2014a). Learning mid-level filters for person re-identification. Em *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, USA.
- Zhao, R.; Ouyang, W. & Wang, X. (2014b). Learning mid-level filters for person re-identification. Em *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, pp. 144–151.
- Zhao, R.; Ouyang, W. & Wang, X. (2014c). Person re-identification by saliency learning. *CoRR*, abs/1412.1908.
- Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Bu, J. & Tian, Q. (2015). Person re-identification meets image search. *CoRR*, abs/1502.02171.
- Zheng, W.-S.; Gong, S. & Xiang, T. (2011). Person re-identification by probabilistic relative distance comparison. Em *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pp. 649–656.
- Zheng, W.-S.; Gong, S. & Xiang, T. (2013). Reidentification by relative distance comparison. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(3):653–668.
- Zheng, W.-S.; Gong, S. & Xiang, T. (2014). Group association: Assisting re-identification by visual context. Em Gong, S.; Cristani, M.; Yan, S. & Loy, C. C., editores, *Person Re-Identification*, Advances in Computer Vision and Pattern Recognition, pp. 183–201. Springer London.
- Zobel, J. & Moffat, A. (2006). Inverted files for text search engines. *ACM Comput. Surv.*, 38(2). ISSN 0360-0300.

# Apêndice A

## Multi-Task Cross-View Collaborative Representation based Classification

O algoritmo Cross-view Collaborative Representation based Classification (X-CRC) criado por de Carvalho Prates & Schwartz [2015] ainda não foi publicado e se encontra em processo de revisão. Uma breve descrição do algoritmo é apresentado a seguir.

Considerando que diferentes campos de visão de um mesmo indivíduo são tarefas relacionadas, o X-CRC estima simultaneamente os coeficientes  $\alpha_1$  e  $\alpha_2$  que descrevem os indivíduos das amostras de teste e galeria, respectivamente. Os autores introduziram um termo de similaridade  $\|\alpha_1 - \alpha_2\|_2^2$  na formulação multi-tarefa, resultando no seguinte problema de otimização

$$\min_{\alpha_1, \alpha_2} \|y - D_1\alpha_1\|_2^2 + \|x - D_2\alpha_2\|_2^2 + \lambda \|\alpha_1\|_2^2 + \lambda \|\alpha_2\|_2^2 + \gamma \|\alpha_1 - \alpha_2\|_2^2 \quad (\text{A.1})$$

que analiticamente foi derivado em relação à  $\alpha_1$  e  $\alpha_2$  obtendo

$$\alpha_1 = \gamma P^{-1}\alpha_2 + P^{-1}D_1^T y \quad (\text{A.2})$$

$$\alpha_2 = \gamma Q^{-1}\alpha_1 + Q^{-1}D_2^T x \quad (\text{A.3})$$

as projeções das matrizes  $P$  e  $Q$  são dadas por

$$P = D_1^T D_1 + (\lambda + \gamma) I \text{ e } Q = D_2^T D_2 + (\lambda + \gamma) I, \quad (\text{A.4})$$

onde as equações A.2 e A.3 são interdependentes, e  $I$  é a matriz identidade. Logo, é

possível substituir  $\alpha_2$  pela equação A.3 e isolar  $\alpha_1$  obtendo

$$\alpha_1 = \gamma W^{-1} P^{-1} Q^{-1} D_2^T x + W^{-1} P^{-1} D_1^T y \quad (\text{A.5})$$

com a matriz de projeção  $W$  correspondendo a

$$W = I - \gamma^2 P^{-1} Q^{-1}. \quad (\text{A.6})$$

Então, a similaridade entre a imagem de teste  $y$  e a galeria  $x$  é calculada pelo erro de reconstrução  $\|y - D_1 \alpha_1\|_2^2$  como descrito no Algoritmo 1.

---

**Algorithm 1:** Multi-Task Cross-View Collaborative Representation based Classification (X-CRC)

---

**Entrada:** Dicionários ( $D_1$  e  $D_2$ ), galeria ( $\mathbf{X}$ ) e amostra de teste ( $\mathbf{Y}$ )

**Saída :** Lista de candidatos das imagens  $\mathbf{R}$  da galeria

Calcule as matrizes  $\mathbf{Q}$  e  $\mathbf{P}$  de acordo com a equação A.4

Calcule  $\mathbf{W}$  usando a equação A.6

Calcule  $Z_x$  e  $Z_y$  de acordo com:

$$Z_x \leftarrow W^{-1} P^{-1} Q^{-1} D_2^T$$

$$Z_y \leftarrow W^{-1} P^{-1} D_1^T$$

**for**  $y_j \in Y$  **do**

**for**  $x_i \in X$  **do**

$\alpha_1 \leftarrow Z_x X - i + Z_y y_j$

$sim(i) \leftarrow \|y_j - D_1 \alpha_1\|_2^2$

$R_j \leftarrow ordena(sim, ascend)$

**return**  $R$

---