UNIVERSIDADE DE SÃO PAULO–USP
ESCOLA DE ENGENHARIA DE SÃO CARLOS
DEPARTAMENTO DE ENGENHARIA ELÉTRICA E DE COMPUTAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA

Carolina Toledo Ferraz

**Novos descritores de textura para
localização e identificação de objetos em
imagens usando “Bag-of-Features”**

Carolina Toledo Ferraz

**Novos descritores de textura para
localização e identificação de objetos em
imagens usando “Bag-of-Features”**

Tese de doutorado apresentada ao Programa de Engenharia Elétrica da Escola de Engenharia de São Carlos como parte dos requisitos para a obtenção do título de Doutor em Ciências.

Área de concentração: Processamento de Sinais e Instrumentação

Orientador: Prof. Dr. Adilson Gonzaga

São Carlos

2016

Trata-se da versão corrigida da tese. A versão original se encontra disponível na EESC/USP que aloja o Programa de Pós-Graduação de Engenharia Elétrica.

AUTORIZO A REPRODUÇÃO TOTAL OU PARCIAL DESTE TRABALHO,
POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS
DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

F381n Ferraz, Carolina Toledo
Novos descritores de textura para localização e
identificação de objetos em imagens usando
"Bag-of-Features" / Carolina Toledo Ferraz; orientador
Adilson Gonzaga. São Carlos, 2016.

Tese (Doutorado) - Programa de Pós-Graduação em
Engenharia Elétrica e Área de Concentração em
Processamento de Sinais e Instrumentação -- Escola de
Engenharia de São Carlos da Universidade de São Paulo,
2016.

1. Descritores Locais. 2. Bag-of-Features. 3.
Support Vector Machine. I. Título.

FOLHA DE JULGAMENTO

Candidata: Licenciada **CAROLINA TOLEDO FERRAZ**.

Título da tese: "Novos descritores de textura para localização e identificação de objetos em imagens usando "Bag-of-Features"".

Data da defesa: 02/09/2016.

Comissão Julgadora:

Resultado:

Prof. Associado **Adilson Gonzaga**
(Orientador)
(Escola de Engenharia de São Carlos/EESC)

APROVADA

Prof. Dr. **Marcelo Andrade da Costa Vieira**
(Escola de Engenharia de São Carlos/EESC)

APROVADA

Prof. Dr. **Aparecido Nilceu Marana**
(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP – Bauru)

APROVADA

Profa. Dra. **Inês Aparecida Gasparotto Boaventura**
(Universidade Estadual Paulista "Júlio de Mesquita Filho"/UNESP – São José do Rio Preto)

APROVADA

Prof. Associado **Rudinei Goulart**
(Instituto de Ciências Matemáticas e de Computação/ICMC-USP)

APROVADA

Coordenador do Programa de Pós-Graduação em Engenharia Elétrica:
Prof. Associado **Luis Fernando Costa Alberto**

Presidente da Comissão de Pós-Graduação:
Prof. Associado **Luis Fernando Costa Alberto**

*Este trabalho é dedicado aos meu pais e ao meu marido que
sempre me incentivaram.*

Agradecimentos

Inicialmente, agradeço a Deus, por me proporcionar saúde e energia para alcançar o meu objetivo.

Agradeço ao meu orientador Prof. Dr. Adilson Gonzaga, pelos momentos de discussão, reflexão, aprendizado, incentivo, e também pela amizade construída ao longo desses anos.

Agradeço aos meus queridos pais, Mercedes e João Baptista, que me educaram e me incentivaram na minha caminhada até aqui. Em especial ao meu pai (em memória) que sempre se orgulhava do meu trabalho e da minha dedicação. Também agradeço ao restante da família por me incentivarem a alcançar meu sonho.

Agradeço ao meu amado marido Marcelo que me apoiou em todos os momentos da minha vida e me incentivou a continuar os estudos. Agradeço os momentos de alegria, descontração e companheirismo nesses anos de doutorado.

Agradeço aos meus amigos de laboratório que também contribuíram para a realização deste trabalho.

Agradeço aos meus amigos que me proporcionaram momentos de distração, alegria e lazer.

Agradeço aos funcionários do departamento de Engenharia Elétrica e de Computação da USP São Carlos e a todos os professores que me auxiliaram durante o doutorado.

Agradeço, por fim, ao CNPq e à Fapesp pelo apoio financeiro na realização deste trabalho.

*“Cada escolha, uma oportunidade.
Cada queda, um aprendizado.
Cada atitude, uma consequência.”
Autor: Desconhecido*

Resumo

Ferraz, Carolina Toledo **Novos descritores de textura para localização e identificação de objetos em imagens usando “Bag-of-Features”**. 151 p. Tese de doutorado – Escola de Engenharia de São Carlos, Universidade de São Paulo, 2016.

Descritores de características locais de imagens utilizados na representação de objetos têm se tornado muito populares nos últimos anos. Tais descritores têm a capacidade de caracterizar o conteúdo da imagem em dados compactos e discriminativos. As informações extraídas dos descritores são representadas por meio de vetores de características e são utilizados em várias aplicações, tais como reconhecimento de faces, cenas complexas e texturas. Neste trabalho foi explorada a análise e modelagem de descritores locais para caracterização de imagens invariantes a escala, rotação, iluminação e mudanças de ponto de vista. Esta tese apresenta três novos descritores locais que contribuem com o avanço das pesquisas atuais na área de visão computacional, desenvolvendo novos modelos para a caracterização de imagens e reconhecimento de imagens. A primeira contribuição desta tese é referente ao desenvolvimento de um descritor de imagens baseado no mapeamento das diferenças de nível de cinza, chamado *Center-Symmetric Local Mapped Pattern* (CS-LMP). O descritor proposto mostrou-se robusto a mudanças de escala, rotação, iluminação e mudanças parciais de ponto de vista, e foi comparado aos descritores *Center-Symmetric Local Binary Pattern* (CS-LBP) e *Scale-Invariant Feature Transform* (SIFT). A segunda contribuição é uma modificação do descritor CS-LMP, e foi denominada *Modified Center-Symmetric Local Mapped Pattern* (MCS-LMP). O descritor inclui o cálculo do pixel central na modelagem matemática, caracterizando melhor o conteúdo da mesma. O descritor proposto apresentou resultados superiores aos descritores CS-LMP, SIFT e LIOP na avaliação de reconhecimento de cenas complexas. A terceira contribuição é o desenvolvimento de um descritor de imagens chamado *Mean-Local Mapped Pattern* (M-LMP) que captura de modo mais fiel pequenas transições dos pixels na imagem, resultando em um número maior de “matches” corretos do que os descritores CS-LBP e SIFT.

Além disso, foram realizados experimentos para classificação de objetos usando as base de imagens Caltech e Pascal VOC2006, apresentando melhores resultados comparando aos outros descritores em questão. Tal descritor foi proposto com a observação de que o descritor LBP pode gerar ruídos utilizando apenas a comparação dos vizinhos com o pixel central. O descritor M-LMP insere em sua modelagem matemática o cálculo da média dos pixels da vizinhança, com o objetivo de evitar ruídos e deixar as características mais robustas. Os descritores foram desenvolvidos de tal forma que seja possível uma redução de dimensionalidade de maneira simples e sem a necessidade de aplicação de técnicas como o PCA. Os resultados desse trabalho mostraram que os descritores propostos foram robustos na descrição das imagens, quantificando a similaridade entre as imagens por meio da abordagem *Bag-of-Features* (BoF), e com isso, apresentando resultados computacionais relevantes para a área de pesquisa.

Palavras-chave: Descritores locais. *Bag-of-Features*. *Support Vector Machine*.

Abstract

Ferraz, Carolina Toledo **New texture descriptors for locating and identifying objects in images using “Bag-of-Features”**. 151 p. Ph.D. Thesis – São Carlos School of Engineering, University of São Paulo, 2016.

Local feature descriptors used in objects representation have become very popular in recent years. Such descriptors have the ability to characterize the image content in compact and discriminative data. The information extracted from descriptors is represented by feature vectors and is used in various applications such as face recognition, complex scenes and textures. In this work we explored the analysis and modeling of local descriptors to characterize invariant scale images, rotation, changes in illumination and viewpoint. This thesis presents three new local descriptors that contribute to the current research advancement in computer vision area, developing new models for the characterization of images and image recognition. The first contribution is the development of a descriptor based on the mapping of gray-level-differences, called Center-Symmetric Local Mapped Pattern (CS-LMP). The proposed descriptor showed to be invariant to scale change, rotation, illumination and partial changes of viewpoint and compared to the descriptors Center-Symmetric Local Binary Pattern (CS-LBP) and Scale-Invariant Feature Transform (SIFT). The second contribution is a modification of the CS-LMP descriptor, which we call Modified Center-Symmetric Local Mapped Pattern (MCS-LMP). The descriptor includes the central pixel in mathematical modeling to better characterize the image content. The proposed descriptor presented superior results to CS-LMP, SIFT and LIOP descriptors in evaluating recognition of complex scenes. The third proposal includes the development of an image descriptor called Mean-Local Mapped Pattern (M-LMP) capturing more accurately small transitions of pixels in the image, resulting in a greater number of “matches” correct than CS-LBP and SIFT descriptors. In addition, experiments for classifying objects have been achieved by using the images based Caltech and Pascal VOC2006, presenting better results compared to other descriptors in question.

This descriptor was proposed with the observation that the LBP descriptor can generate noise using only the comparison of the neighbors to the central pixel. The M-LMP descriptor inserts in their mathematical modeling the averaging of the pixels of the neighborhood, in order to avoid noise and leave the more robust features. The results of this thesis showed that the proposed descriptors were robust in the description of the images, quantifying the similarity between images using the Bag-of-Features approach (BoF), and thus, presenting relevant computational results for the research area.

Keywords: Local descriptors. *Bag-of-Features*. *Support Vector Machine*.

Lista de ilustrações

Figura 1	Representação do procedimento de obtenção das Diferenças de Gaussianas DoG para diversas oitavas de uma imagem (LOWE, 2004).	39
Figura 2	Detecção de extremos no espaço-escala (LOWE, 2004).	39
Figura 3	A imagem da esquerda são os gradientes da imagem. A imagem da direita o descritor de pontos-chave. (LOWE, 2004).	40
Figura 4	Características dos métodos LBP e CS-LBP para uma vizinhança de 8 <i>pixels</i> (Adaptado de (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009))	41
Figura 5	ILBP (MARCEL; RODRIGUES; HEUSCH, 2007)	42
Figura 6	M-LBP	43
Figura 7	Diferenças de níveis de cinza como um padrão local	44
Figura 8	Diferença do mapeamento dos níveis de cinza pela Equação 10. (a) <i>Local Mapped Pattern</i> . (b) <i>Local Binary Pattern</i>	45
Figura 9	O método LIOP (WANG; FAN; WU, 2011)	46
Figura 10	Estratégia de geração do dicionário no algoritmo BoF	49
Figura 11	Algoritmos de agrupamento (JAIN; MURTY; FLYNN, 1999)	50
Figura 12	Operação do algoritmo <i>k-means</i> (FONTANA; NALDI, 2009)	52
Figura 13	Exemplo da abordagem <i>hard assignment</i>	53
Figura 14	Exemplo da abordagem <i>soft assignment</i>	54
Figura 15	Exemplo da pirâmide espacial de Lazebnik et al. (2006)	54
Figura 16	Exemplo de geração de um código CS-LMP para uma vizinhança de 8 <i>pixels</i>	58
Figura 17	Construção do descritor CS-LMP	59
Figura 18	Características do MCS-LMP para uma vizinhança de 8 <i>pixels</i>	61
Figura 19	Descritor MCS-LMP	61
Figura 20	Exemplo de geração de código M-LMP para uma vizinhança de 8 <i>pixels</i>	63

Figura 21	Imagens utilizadas para o ajuste do parâmetro β : <i>graf</i> (mudança de ponto de vista), <i>bark</i> (escala e rotação), <i>trees</i> (imagem borrada), <i>leuven</i> (mudança de iluminação), <i>ubc</i> (compressão JPEG), <i>boat</i> (escala e rotação), <i>bike</i> (imagem borrada), <i>wall</i> (mudança de ponto de vista) . . .	66
Figura 22	Imagens do conjunto de teste: <i>resid</i> , <i>east park</i> , <i>east south</i> , <i>ensimag</i> , <i>laptop</i> e <i>Inria</i> (escala e rotação), <i>zmars</i> , <i>zmonet</i> , <i>new york</i> (rotação), <i>axterix</i> , <i>belledone</i> , <i>bip</i> , <i>crolle</i> , <i>lap</i> e <i>vangogh</i> (escala)	67
Figura 23	Exemplo de uma imagem de cada categoria da Base de Dados Caltech 2004	68
Figura 24	Exemplo de imagens de cada uma das seis classes da base de dados Caltech 2004, usadas no segundo experimento.	68
Figura 25	Exemplo de imagens de cada classe da base de dados Pascal VOC 2006.	68
Figura 26	Exemplo de uma imagem de cada categoria da Base de Dados Caltech 101	68
Figura 27	Exemplo de uma imagem de cada categoria da Base de Dados Caltech 101	69
Figura 28	Exemplo de uma imagem de cada categoria da Base de faces Feret . . .	69
Figura 29	Exemplo de 3 imagens de 6 categorias da Base de faces ORL	70
Figura 30	Exemplo de 3 imagens de 6 categorias da Base de faces <i>Yale Face</i> . . .	70
Figura 31	Exemplos de imagens de cada classe da Base de Imagens Caltech 101 .	71
Figura 32	Exemplos de imagens de cada classe da Base de Imagens ImageNet . .	72
Figura 33	Exemplos de imagens de cada classe da base de imagens <i>Indoor Scene Recognition dataset</i>	73
Figura 34	<i>Framework</i> do método <i>Bag of Features (BoF)</i>	75
Figura 35	<i>Framework</i> da metodologia proposta para o reconhecimento de faces humanas	76
Figura 36	Resultados das curvas <i>recall</i> $\times$ <i>precision</i> para os descritores CS-LMP, M-LMP, CS-LBP e SIFT usando o detector de regiões de interesse <i>Hessian-Affine</i>	81
Figura 37	Resultados das curvas <i>recall</i> $\times$ <i>precision</i> para os descritores M-LMP, CS-LBP e SIFT usando o detector de regiões de interesse <i>Harris-Affine</i>	82
Figura 38	Curvas <i>recall</i> $\times$ <i>precision</i> para os descritores CS-LMP, CS-LBP e SIFT para a imagem do conjunto de teste chamada <i>bark</i>	83
Figura 39	Curvas <i>recall</i> $\times$ <i>precision</i> para os descritores M-LMP, CS-LBP e SIFT para a imagem do conjunto de teste chamada <i>bark</i>	84
Figura 40	Tempo de processamento (segundos) versus número de <i>bins</i> (<i>b</i>) do descritor usando a distância Euclideana nos métodos CS-LMP e M-LMP na imagem do conjunto de treinamento chamada <i>leuven</i>	84

Figura 41	Curvas <i>recall</i> $\times$ <i>precision</i> para os descritores CS-LMP, M-LMP, CS-LBP e SIFT com a redução no número de <i>bins</i> (<i>b</i>) para <i>b</i> =8 (CS-LMP) e <i>b</i> =7 (M-LMP).	87
Figura 42	Correspondências corretas entre um par de imagens “ <i>east park</i> ”. Total de 361 correspondências retornadas.	88
Figura 43	Variação da sensibilidade com o tamanho de <i>cluster</i> (6 classes da base de imagens Caltech 2004)	90
Figura 44	<i>b</i> <i>versus</i> sensibilidade para quatro classes da base de imagens Caltech 101	94
Figura 45	<i>b</i> <i>versus</i> taxa de erro para quatro classes da base de imagens Caltech 101	94
Figura 46	Matriz de Confusão para o descritor M-LMP usando <i>dense sampling</i> (Caltech 101)	105
Figura 47	Matriz de confusão para o descritor SIFT usando <i>dense sampling</i> (Caltech 101)	106
Figura 48	Matriz de confusão para o descritor LIOP usando <i>dense sampling</i> (Caltech 101)	106
Figura 49	Matriz de confusão para o descritor CS-LMP usando <i>dense sampling</i> (Caltech 101)	107
Figura 50	Exemplos de imagens de 5 classes mostrando a melhor performance do descritor M-LMP na base de imagens Caltech 101 (30 acertos = 100% classificações corretas)	108
Figura 51	Exemplos de imagens de 5 classes mostrando a melhor performance do descritor LIOP na base de imagens Caltech 101 (30 acertos = 100% de classificações corretas)	108
Figura 52	Exemplos de imagens de 6 classes mostrando a melhor performance do descritor CS-LMP na base de imagens Caltech 101 (30 acertos = 100% classificações corretas)	108
Figura 53	Exemplo de imagens de 5 classes mostrando a melhor performance do descritor SIFT na base de imagens Caltech 101 (30 acertos = 100% classificações corretas)	109
Figura 54	A pior performance para todos os descritores: exemplos da classe <i>background_google</i> (Caltech 101)	109
Figura 55	Classes com um número grande de imagens classificadas incorretamente pelos descritores M-LMP e CS-LMP para a base de imagens Caltech 101 (confusões interclasses)	109
Figura 56	Classes com um grande número de imagens classificadas incorretamente pelos descritores LIOP e SIFT para a base de imagens Caltech 101 (confusões interclasses)	109

Figura 57	Matriz de confusão para o descritor M-LMP usando <i>Hessian Multiscale</i> como detector de regiões de interesse (Caltech 101)	112
Figura 58	Matriz de confusão para o descritor SIFT usando <i>Hessian Multiscale</i> como detector de regiões de interesse (Caltech 101)	113
Figura 59	Matriz de confusão para o descritor LIOP usando <i>Hessian Multiscale</i> como detector de regiões de interesse (Caltech 101)	114
Figura 60	Matriz de confusão para o descritor CS-LMP usando <i>Hessian Multiscale</i> como detector de regiões de interesse (Caltech 101)	115
Figura 61	Matriz de confusão para o descritor M-LMP usando <i>dense sampling</i> (ImageNet)	116
Figura 62	Matriz de confusão para o descritor SIFT usando <i>dense sampling</i> (ImageNet)	117
Figura 63	Matriz de confusão para o descritor LIOP usando <i>dense sampling</i> (ImageNet)	118
Figura 64	Matriz de confusão para o descritor CS-LMP usando <i>dense sampling</i> (ImageNet)	119
Figura 65	Matriz de confusão para o descritor M-LMP usando <i>Hessian Multiscale</i> (ImageNet)	120
Figura 66	Matriz de confusão para o descritor SIFT usando <i>Hessian Multiscale</i> (ImageNet)	121
Figura 67	Matriz de confusão para o descritor LIOP usando <i>Hessian Multiscale</i> (ImageNet)	121
Figura 68	Matriz de Confusão para o descritor CS-LMP usando o detector <i>Hessian Multiscale</i> (ImageNet)	122
Figura 69	Matriz de confusão para o descritor MCS-LMP usando <i>Intersection Kernel</i> como um núcleo para o SVM aplicado sobre a base <i>Indoor Scene Recognition</i>	123
Figura 70	Matriz de confusão para o descritor CS-LMP usando <i>Intersection Kernel</i> como um núcleo para o SVM aplicado sobre a base <i>Indoor Scene Recognition</i>	124
Figura 71	Matriz de confusão para o descritor SIFT usando <i>Intersection Kernel</i> como um núcleo para o SVM aplicado sobre a base <i>Indoor Scene Recognition</i>	124
Figura 72	Matriz de confusão para o descritor LIOP usando <i>Intersection Kernel</i> como um núcleo para o SVM aplicado sobre a base <i>Indoor Scene Recognition</i>	125
Figura 73	Matriz de confusão para o descritor MCS-LMP usando <i>Chi-Square kernel</i> para o classificador SVM aplicado sobre a base <i>Indoor Scene Recognition</i>	126

- Figura 74 Matriz de confusão para o descritor CS-LMP usando *Chi-Square kernel* para o classificador SVM aplicado sobre a base *Indoor Scene Recognition*.126
- Figura 75 Matriz de confusão para o descritor SIFT usando *Chi-Square kernel* para o classificador SVM aplicado sobre a base *Indoor Scene Recognition*.127
- Figura 76 Matriz de confusão para o descritor LIOP usando *Chi-Square kernel* para o classificador SVM aplicado sobre a base *Indoor Scene Recognition*.127

Lista de tabelas

Tabela 1	Avaliação do parâmetro β para a imagem exemplo <i>bark</i> usando <i>Hessian-Affine</i>	80
Tabela 2	Avaliação do parâmetro β para a imagem exemplo <i>trees</i> usando <i>Harris-Affine</i>	80
Tabela 3	<i>Area under curve</i> (AUC) para as imagens do conjunto de teste da Figura 22 usando o detector de regiões de interesse <i>Hessian-Affine</i> . . .	81
Tabela 4	<i>Area under curve</i> (AUC) para as imagens do conjunto de teste da Figura 22, usando o detector de regiões de interesse <i>Harris-Affine</i> . . .	83
Tabela 5	<i>Area under curve</i> para a imagem do conjunto de teste chamada <i>bark</i> , variando-se o número de <i>bins</i> (<i>b</i>).	84
Tabela 6	Tempo de processamento para comparação de dois vetores de um par de imagens com $b=8$ para o CS-LMP e $b=7$ para o M-LMP (milisegundos)	85
Tabela 7	<i>Area under curve</i> (AUC) para as imagens do banco de teste, calculadas com $b=8$ para o CS-LMP e $b=7$ para o M-LMP	86
Tabela 8	Razão de correspondências corretas obtidas para as imagens do conjunto de teste.	86
Tabela 9	O melhor tamanho de <i>cluster</i> (4 classes da base de dados Caltech 2004)	88
Tabela 10	Matriz de Confusão para localização de objetos em quatro classes de imagens da base Caltech 2004, utilizando o descritor M-LMP	89
Tabela 11	Matriz de Confusão para localização de objetos em quatro classes de imagens da base Caltech 2004, utilizando o descritor CS-LBP	89
Tabela 12	Matriz de Confusão para localização de objetos em quatro classes de imagens da base Caltech 2004, utilizando o descritor SIFT	89
Tabela 13	Taxas de erro para quatro classes de imagens da base Caltech 2004 . .	90
Tabela 14	Matriz de confusão para o descritor M-LMP na classificação de objetos (6 classes da base de imagens Caltech 2004)	91
Tabela 15	Matriz de confusão para o descritor CS-LBP na classificação de objetos (6 classes da base de imagens Caltech 2004)	91

Tabela 16	Matriz de confusão para o descritor SIFT na classificação de objetos (6 classes da base de imagens Caltech 2004)	91
Tabela 17	Taxas de erro (6 classes da base de imagens Caltech 2004)	92
Tabela 18	Matriz de Confusão para o descritor M-LMP na classificação de objetos (4 classes da base de imagens Pascal VOC 2006)	92
Tabela 19	Matriz de Confusão para o descritor CS-LBP na classificação de objetos (4 classes da base de imagens Pascal VOC 2006)	92
Tabela 20	Matriz de Confusão para o descritor SIFT na classificação de objetos (4 classes da base de imagens Pascal VOC 2006)	93
Tabela 21	Taxas de erro (4 classes da base de imagens Pascal VOC 2006)	93
Tabela 22	Comparação da sensibilidade dos três descritores	93
Tabela 23	Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor M-LMP com $b = 15$.	94
Tabela 24	Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor M-LMP com $b = 4$. .	95
Tabela 25	Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor CS-LBP	95
Tabela 26	Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor SIFT	95
Tabela 27	Taxas de erro (4 classes do banco de imagens Caltech 101)	95
Tabela 28	Tempo de processamento em segundos para a construção do vocabulário visual utilizando o descritor M-LMP variando o parâmetro b	96
Tabela 29	Tempo de processamento em segundos para a construção do vocabulário visual utilizando os descritores CS-LBP e SIFT	96
Tabela 30	Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 5 classes, utilizando o descritor M-LMP com $b = 4$. .	97
Tabela 31	Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 5 classes, utilizando o descritor CS-LBP	97
Tabela 32	Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 5 classes, utilizando o descritor SIFT	97
Tabela 33	Taxas de erro (5 classes do banco de imagens Caltech 101)	97
Tabela 34	Performance do descritor M-LMP durante a sintonização do parâmetro β para 40 imagens por classes (base de imagens Feret)	99
Tabela 35	Desempenho dos métodos comparados (sensibilidade) para 45 classes da base de imagens Feret variando-se o tamanho do <i>cluster</i>	100
Tabela 36	Comparação da metodologia proposta com a metodologia <i>Face Fuzzy</i> para a base de imagens Feret	100
Tabela 37	Comparação da metodologia proposta com outras metodologias usando a base de imagens ORL	101

Tabela 38	Comparação da metodologia proposta com outras metodologias usando a base de imagens <i>Yale Face</i>	102
Tabela 39	Desempenho do descritor M-LMP durante o ajuste do parâmetro β para diferentes tamanhos de descritores (Caltech 101)	103
Tabela 40	Desempenho do descritor M-LMP durante o ajuste do parâmetro β variando o número de imagens por classe (Caltech 101)	104
Tabela 41	Desempenho do descritor CS-LMP durante o ajuste do parâmetro β para 80 imagens por classe (Caltech 101)	104
Tabela 42	Sensibilidade (27 classes da base de imagens Caltech 101) usando <i>dense sampling</i>	105
Tabela 43	O pior desempenho (<i>hit-rate</i>) dos métodos comparados para a base de imagens Caltech 101 (<i>dense sampling</i>)	109
Tabela 44	Tempo de processamento (em segundos) para a construção do vocabulário visual para os descritores M-LMP, SIFT, LIOP e CS-LMP usando <i>dense sampling</i> como detector de regiões de interesse	110
Tabela 45	Tempo de processamento (em segundos) para a tarefa de classificação dos descritores M-LMP, SIFT, LIOP and CS-LMP usando <i>dense sampling</i> como detector de regiões de interesse	110
Tabela 46	Desempenho dos descritores M-LMP e CS-LMP durante o ajuste do parâmetro β usando <i>Hessian Multiscale</i> e um descritor de 80 elementos (Caltech 101)	110
Tabela 47	Desempenho do método (taxa de acerto) usando o detector <i>Hessian Multiscale</i> (para 27 classes de imagens da base de dados Caltech 101) .	111
Tabela 48	O pior desempenho (taxa de acerto) dos métodos comparados (<i>Hessian Multiscale</i>)	113
Tabela 49	Tempo de processamento (em segundos) para classificação de objetos da base Caltech 101 usando os descritores M-LMP, SIFT, LIOP e CS-LMP (<i>Hessian Multiscale</i>)	114
Tabela 50	Desempenho dos métodos	114
Tabela 51	Ajuste do parâmetro β para o descritor M-LMP usando 80 imagens por classe (ImageNet)	115
Tabela 52	Desempenho dos descritores comparados (sensibilidade) usando o detector <i>dense sampling</i> (para 27 classes da base de imagens ImageNet) .	116
Tabela 53	Sintonização do parâmetro β na base de imagens ImageNet para o descritor M-LMP	118
Tabela 54	Sintonização do parâmetro β da base de imagens ImageNet para o descritor CS-LMP	119

Tabela 55	Desempenho dos descritores comparados (Sensibilidade) usando o detector de regiões <i>Hessian Multiscale</i> (para 27 classes de imagens da base de imagens ImageNet)	120
Tabela 56	Ajuste do parâmetro β para o descritor MCS-LMP.	123
Tabela 57	Ajuste do parâmetro β para o descritor CS-LMP.	123
Tabela 58	Tempo de processamento em segundos para a construção do vocabulário visual usando os descritores MCS-LMP, CS-LMP, SIFT e LIOP. . .	125
Tabela 59	Imagens selecionadas para o <i>Codebook</i> (Pascal VOC 2006).	149
Tabela 60	Imagens selecionadas para o conjunto de treinamento (Pascal VOC 2006). .	150
Tabela 61	Imagens selecionadas para o conjunto de teste (Pascal VOC 2006).	150

Lista de siglas

AUC *Area Under Curve*

CS-LBP *Center-Symmetric Local Binary Patterns*

CS-LMP *Center-Symmetric Local Mapped Pattern*

CSIFT *Colored SIFT*

FUNED *Fuzzy Number Edge Detector*

GLOH *Gradient Location and Orientation Histogram*

ILBP *Improved Local Binary Pattern*

KPB-SIFT *Kernel Projection Based SIFT*

LBP *Local Binary Pattern*

LFP *Local Fuzzy Pattern*

LMP *Local Mapped Pattern*

M-LMP *Mean Local Mapped Pattern*

M-LBP *Modified Local Binary Pattern*

PCA *Principal Component Analysis*

PCA-SIFT *Principal Component Analysis SIFT*

SIFT *Scale-Invariant Feature Transform*

SURF *Speeded Up Robust Features*

SVM *Support Vector Machine*

MB-LBP *Multi-scale Block Local Binary Pattern*

LEDTD *Local Edge Direction and Texture Description*

LLBP *Local Line Binary Pattern*

LIOP *Local Intensity Order Pattern*

FLDA *Fisher Linear Discriminant Analysis*

LFDA *Local Fisher Discriminant Analysis*

ISOMAP *Isometric Mapping*

LLE *Locally Linear Embedding*

LPP *Locality Preserving Projections*

HEP *Histograms of Equivalent Patterns*

LIO *Local Intensity Order*

LGOP *Local Gradient Local Pattern*

Sumário

1	Introdução	29
1.1	Contextualização	29
1.2	Motivação	32
1.3	Definição do Problema	33
1.4	Objetivos e Contribuições	34
1.5	Organização da Tese	34
2	Fundamentação Teórica	37
2.1	Introdução	37
2.2	Extração de características	37
2.2.1	<i>Scale Invariant Feature Transform</i> (SIFT)	37
2.2.2	<i>Center Symmetric Local Binary Pattern</i> (CS-LBP)	40
2.2.3	Os métodos ILBP e M-LBP	42
2.2.4	<i>Local Mapped Pattern</i> (LMP)	43
2.2.5	<i>Local Intensity Order Pattern</i> (LIOP)	45
2.3	<i>Bag-of-Features</i> (BoF)	46
2.3.1	Deteção de regiões de interesse e descritores de imagens	48
2.3.2	Dicionário de palavras visuais	49
2.3.3	Histograma de palavras visuais	53
2.3.4	Classificadores	55
2.4	Considerações Finais	55
3	Descritores Propostos	57
3.1	Introdução	57
3.2	<i>Center-Symmetric Local Mapped Pattern</i> (CS-LMP)	57
3.3	<i>Modified Center-Symmetric Local Mapped Pattern</i> (MCS-LMP)	59
3.4	<i>Mean Local Mapped Pattern</i> (M-LMP)	62
3.5	Considerações Finais	62

4	Material e Metodologia	65
4.1	Banco de Imagens	65
4.1.1	Banco de imagens para Correspondência de Imagens	65
4.1.2	Banco de imagens utilizados na metodologia BoF	65
4.2	Metodologia	74
4.2.1	Metodologia utilizada para correspondência de imagens	74
4.2.2	Metodologia utilizada para classificação das imagens	74
4.3	Critério de Avaliação	77
4.3.1	Critério de avaliação para correspondência de imagens	77
4.3.2	Critério de avaliação para a classificação das imagens	78
4.4	Considerações Finais	78
5	Resultados	79
5.1	Introdução	79
5.2	Correspondência de imagens	79
5.2.1	Redução do tamanho dos descritores	82
5.3	Resultados obtidos na localização e identificação de objetos	87
5.3.1	Resultados obtidos com o descritor M-LMP	87
5.3.2	Resultados obtidos com os descritores M-LMP e CS-LMP	101
5.3.3	Resultados obtidos com o descritor MCS-LMP	120
5.4	Considerações Finais	128
6	Conclusão	129
6.1	Contribuições	134
6.2	Trabalhos Futuros	135
6.3	Artigos aceitos e submetidos	135
	Referências	137
	Apêndices	147
	APÊNDICE A Tabelas	149

Introdução

1.1 Contextualização

Os fundamentos da teoria da visão publicados por David Marr em 1982 (MARR, 1982) baseiam-se na necessidade de abordar a compreensão dos requisitos da descrição da cena. Esta concepção de visão prevê a análise das imagens em estágios iniciais de processamento do particular para o geral (processamento *bottom-up* ou guiado por dados sensoriais), havendo uma grande quantidade de processamento e numerosas representações simbólicas. No entanto, na fase final da análise é normalmente armazenado conhecimento sobre os dados informativos particulares (processamento *top-down* ou guiadas conceitualmente). A visão computacional clássica evoluiu a partir desses conceitos indicando fases para o reconhecimento de objetos em cenas dentro de um *framework* amplamente utilizado e que consiste do processamento inicial envolvendo filtragens e modificações nos níveis de intensidade dos pixels, segmentação de continuidades e descontinuidades dos níveis de cinza, seleção e extração de características, descrição destes segmentos e reconhecimento e interpretação da cena. Este modelo pode ser referenciado como “orientado a segmentação”, pois a etapa essencial e que guia todo o processo de visão reside exatamente em se definir quais os segmentos deverão ser extraídos para permitir o reconhecimento dos objetos na cena. Esta definição está intimamente relacionada com o conhecimento humano sobre o problema o que introduz uma interferência *top-down* já na fase de segmentação.

Outro modelo de visão computacional que tem tido grandes avanços nos últimos anos parte de uma análise global da imagem e procura detectar similaridades que sejam significativas para objetos semelhantes sem a necessidade de segmentação da cena. A partir da década de 80, os estudos relacionados à fisiologia da visão e da atenção visual intensificaram-se. Recentes progressos em Neurociência clarearam a ideia da estrutura e organização dos centros cerebrais que cooperam com as funções visuais. Especialização, modularidade e organização hierárquica parecem ser princípios importantes para entender o funcionamento desse sistema extremamente complexo (MILANESE, 1993). Nesse sentido diversas tendências se verificam, tais como, a atenção visual seletiva e os mapas

de conspicuidades ou mapas de saliências, a detecção de Regiões de Interesse (RoI) e de pontos-chave (*keypoints*). Os modelos de atenção visual seletiva inspirados biologicamente seguem a hipótese de Kock e Ullman (1987) (KOCH; ULLMAN, 1987), no qual mapas de características alimentam um mapa de saliências. Este mapa representa topograficamente a conspicuidade visual (ITTI; KOCH, 2001), ou seja, características que por algum motivo foram localizadas na imagem como salientes ou importantes para a classificação de objetos presentes na cena.

A partir do trabalho de Lowe (1999) (LOWE, 1999) com a definição de *keypoints* por meio da abordagem da Transformada de Características Invariantes à Escala (SIFT) (LOWE, 2004) outros modelos foram propostos permitindo a análise de uma imagem baseada em regiões similares e de objetos presentes em cenas mesmo que sofram transformações afins, de escala, rotação posição e oclusões parciais. Dentre estes modelos destacam-se o SURF (BAY; TUYTELAARS; GOOL, 2006), o CS-LBP (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2006; HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009) e os descritores propostos nesta tese.

Os pontos-chave de uma imagem, utilizados para reconhecimento de objetos em uma cena, tem paralelo imediato com a maneira como seres humanos realizam a busca por informação visual utilizando os movimentos sacádicos do olho (RAYNER, 2009). No entanto, variações nas imagens interferem drasticamente nas tarefas de reconhecimento e detecções de objetos, cenas e faces. Tais variações incluem oclusões, iluminação, escala e distorções geométricas. Além disso, há o caso de imagens ruidosas, com rotações e deformações fotométricas. Com isso, classificar uma imagem, ou melhor, uma classe de imagens cujo objeto principal da cena é o fator determinante, torna-se uma tarefa desafiadora. Outro fator que interfere nesta classificação é o fundo sobre o qual o objeto principal está “assentado”. A diversidade dos elementos do fundo tem muitas vezes papel preponderante na definição dos pontos chaves e consequentemente na atribuição da classe correta à qual pertence a imagem. Para interpretar seu conteúdo é necessário construir um descritor capaz de extrair uma quantidade significativa de características que a discrimine de maneira robusta, ou seja, mesmo com as possíveis alterações em escala, rotação, iluminação e outras afins, consiga reconhecer eficientemente cada imagem avaliada.

Um dos descritores mais conhecidos e consolidados na literatura é o *Scale-Invariant Feature Transform* (SIFT). Ele foi introduzido por Lowe (1999) e é caracterizado por um histograma 3D de gradientes orientados para descrever as vizinhanças locais de pontos de interesse. Ele armazena os *bins* em um vetor de 128 posições. Algumas variações do SIFT foram propostas, tais como *Principal Component Analysis SIFT* (PCA-SIFT) (KE; SUKTHANKAR, 2004), *Gradient Location and Orientation Histogram* (GLOH) (K.MIKOLAJCZYK; SCHMID, 2005), *Speeded Up Robust Features* (SURF) (BAY; TUYTELAARS; GOOL, 2006), *Colored SIFT* (CSIFT) (ABDEL-HAKIM; FARAG, 2006), *Center-Symmetric Local Binary Patterns* (CS-LBP) (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009) e *Kernel Projection Based SIFT* (KPB-SIFT) (ZHAO et al., 2010). A técnica PCA-SIFT

combina a metodologia SIFT com o redutor de dimensionalidade *Principal Component Analysis* (PCA). O método PCA é capaz de projetar linearmente descritores de altas dimensões em um espaço de baixa dimensão e reduzir ruídos de alta frequência. Contudo, os descritores baseados em PCA precisam de um estágio *offline* para treinar e estimar a matriz de covariância utilizada para a projeção. Sendo assim, a aplicação baseada no método PCA é limitada. O descritor GLOH também usa PCA para reduzir o tamanho do descritor. Ele é muito similar ao SIFT, contudo, utiliza-se um *grid* polar ao invés do *grid* cartesiano. O método SURF se aproxima do SIFT utilizando imagem integral para calcular respostas à *wavelet* de Haar em sentidos horizontal e vertical. O tamanho do descritor é de 64 posições, o que permite um menor tempo de processamento. Como o método SIFT trata apenas imagens em níveis de cinza, o método CSIFT surgiu com o intuito de abranger descritores de imagens coloridas. Ele usa um modelo de invariância de cor para construir o descritor. O método KBP-SIFT reduz a dimensão do descritor de 128 para somente 36 *bins*. A abordagem é tolerante a distorções geométricas e não requer um estágio de treinamento; contudo, a avaliação de desempenho não é superior em todos os casos com relação ao SIFT.

Outro descritor local amplamente utilizado é o *Local Binary Pattern* (LBP) (OJALA; M.PIETIKÄINEN; D.HARWOOD, 1996). Uma variação do tradicional LBP é o *Modified Local Binary Pattern* (M-LBP) (BAI; ZHU; DING, 2008) que não considera o *pixel* central como é feito no método original. Jin et al. (2004) propuseram o *Improved Local Binary Pattern* (ILBP) que compara todos os *pixels* (incluindo o *pixel* central) com a intensidade média. O descritor chamado de CS-LBP (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009) combina os pontos fortes do SIFT e o operador de textura LBP. A construção do CS-LBP é mais simples do que o SIFT, mas gera um vetor de características de 256 posições que é duas vezes o tamanho do vetor do SIFT. *Local Line Binary Pattern* (LLBP) (PETPON; SRISU, 2009) foi introduzido por Petpon e Srisuk (2009) para reconhecimento de expressões faciais e faces com variações de iluminação. *Multi-scale Block Local Binary Pattern* (MB-LBP) (LIAO et al., 2007) também é utilizada para reconhecimento facial e codifica as macroestruturas e microestruturas dos padrões na imagem fornecendo uma representação mais robusta do que o LBP. *Local Edge Direction and Texture Description* (LEDTD) (LI; SANG; GAO, 2016) explora a direção da aresta e a informação de textura disponível localmente.

Em um trabalho recente, Vieira et al. (2012) apresentaram um novo descritor baseado em números *fuzzy* chamado *Local Fuzzy Pattern* (LFP). Este método interpreta os valores de nível de cinza de uma vizinhança da imagem como um conjunto *fuzzy* e cada nível como um número *fuzzy*. Uma função de pertinência é usada para descrever a pertinência do *pixel* central para uma vizinhança. Além disso, a metodologia LFP é uma generalização das técnicas já publicadas anteriormente tais como o LBP, o *Texture Unit* (HE; WANG, 1990), o *Fuzzy Number Edge Detector* (FUNED) (BOAVENTURA; GONZAGA,

2009) e a Transformada Census (ZABIH; WOODFILL, 1994). Uma evolução do método LFP foi o *Local Mapped Pattern* (LMP) (FERRAZ; JUNIOR; GONZAGA, 2014), onde os autores consideram a soma das diferenças de cada nível de cinza de uma vizinhança para o *pixel* central como um padrão local que pode ser mapeado para um *bin* do histograma usando uma função de mapeamento.

1.2 Motivação

Descritores de características que representam uma imagem são úteis para diversas aplicações na área de visão computacional, tais como recuperação e classificação de imagens. Muitas metodologias foram desenvolvidas com representações globais e locais. As metodologias consideradas globais (N.; NAIKA; DAS, 2014; TALEB; OUIS; MAMMAR, 2014; MOON et al., 2015) ignoram as características locais das imagens, o que pode-se tornar um problema quando há muitas variações na imagem, como em face humana por exemplo (ZHANG; YAN; LADES, 1997). Já as metodologias locais, que é o caso dos descritores já citados, extraem características mais discriminantes e robustas (FAUDZI; YAHYA, 2014; ZHANG et al., 2015). Sendo assim, há um grande interesse por parte dos pesquisadores na investigação de descritores locais para as tarefas de reconhecimento e detecção de objetos.

Há um enorme desafio na construção de um descritor devido a fatores que influenciam na descrição de uma imagem, tais como, iluminação, escala, oclusões de objetos, pose, rotações, etc. Além disso, os descritores devem ter tamanhos reduzidos para não impactar no tempo de processamento do sistema e armazenamento em disco.

Em muitos casos são utilizadas técnicas como o PCA para reduzir a dimensionalidade dos descritores (K.MIKOLAJCZYK; SCHMID, 2005; KE; SUKTHANKAR, 2004; HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009). O trabalho de Valenzuela e Pedrini (2013) apresenta a comparação de diferentes redutores de dimensionalidade para os descritores SIFT e SURF, englobando a abordagem *Bag-of-Features* (BoF). A utilização de metodologias que não utilizem método *offline* para a redução dos descritores aumentaria a eficácia da utilização de tais descritores em sistemas em tempo real.

Os descritores são englobados dentro de metodologias para a execução de tarefas (YANG; KURITA, 2013; ALWANI; CHAHIR, 2014; JANSEN; SHANTIA; WIERING, 2015). A metodologia BoF pode ser usada como uma abordagem global para análise de descritores. Nos últimos anos, o método BoF tem demonstrado ser um método promissor e altamente adaptável a vários tipos de problemas, alcançando boa performance (JURIE; TRIGGS, 2005; LAZEBNIK; SCHMID; PONCE, 2006; ZHANG et al., 2009; SENNA; DRUMMOND; BASTOS, 2015; CUI et al., 2015; LI; WANG; ZHANG, 2016). Esta técnica tem como objetivo representar uma imagem por vetores de características, e utilizando técnicas de agrupamento de dados, construir um vocabulário visual que será aplicado na análise, recuperação e classificação de imagens.

1.3 Definição do Problema

O principal problema investigado nesta tese de doutorado é o desenvolvimento de descritores de regiões de interesse que sejam robustos, de modo a extrair as características locais de imagens e, utilizar a abordagem BoF para localização de objetos semelhantes em imagens com variações de escala, iluminação e rotação. Além disso, os descritores foram desenvolvidos de tal forma que seja possível uma redução de dimensionalidade de maneira simples e sem a necessidade de aplicação de técnicas como PCA.

A metodologia para a resolução do problema a ser investigado consiste em vários passos que serão descritos a seguir, utilizando a abordagem BoF.

- ❑ Encontrar pontos-chave nas imagens, utilizando detectores tais como *dense sampling*, *Hessian-Affine*, *Hessian-Laplace*, *Harris-Affine* e *Hessian-Multiscale*. A utilização de detectores em imagens é um passo primordial do sistema, pois dependendo da quantidade de pontos-chave detectados, uma quantidade enorme de dados podem ser gerados, prejudicando o desempenho do sistema. Por outro lado, uma quantidade maior de pontos-chave poderá aumentar a sensibilidade de detecção e reconhecimento das imagens.
- ❑ Considerar a região ao redor de cada ponto-chave e aplicar um bom descritor local com características importantes, tais como invariância a iluminação, rotação e escala. Além disso, é necessário que o descritor seja de tamanho reduzido. O desenvolvimento de descritores locais é a motivação principal deste trabalho.
- ❑ Construir um vocabulário visual que englobe tais descritores. A construção de dicionários envolve a utilização de algoritmos para agrupamento, sendo que a análise desses algoritmos para que o tempo de busca seja reduzido é de extrema importância.
- ❑ Representar as imagens com histogramas de palavras visuais, indicando com que frequência cada palavra visual aparece na imagem. A representação de histogramas como um método para descrever imagens pode ser contruído por uma pirâmide espacial, onde a imagem é particionada em sub-regiões menores e para cada sub-região, um histograma é calculado. Tal representação tem se mostrado uma alternativa com relação a representação usual, com um ganho na acurácia em aplicações para classificação de objetos.
- ❑ Utilizar metodologias para classificar e detectar objetos em imagens. *Support Vector Machine* (SVM) é uma técnica de aprendizado que vem apresentando resultados superiores a outros algoritmos de classificação como Redes Neurais Artificiais.

1.4 Objetivos e Contribuições

O principal objetivo da tese de doutorado é o desenvolvimento de descritores de regiões de interesse que possuam características importantes, tais como, invariância a escala, rotação, iluminação e mudanças de ponto de vista, e utilizar a abordagem BoF para localização de objetos semelhantes em imagens. Além disso, a redução de dimensionalidade dos descritores deve ocorrer de uma maneira simples e sem a necessidade de aplicação de técnicas como PCA.

A primeira contribuição foi o desenvolvimento do descritor *Center-Symmetric Local Mapped Pattern* (CS-LMP). O descritor proposto combina as boas propriedades dos descritores CS-LBP e SIFT e utiliza uma função de mapeamento sigmóide na modelagem matemática. Tal descritor encontrou um número maior de *matches* comparado ao descritor CS-LBP e mostrou-se robusto com relação a variação de escala, rotação, iluminação e mudanças de pontos de vista.

A segunda contribuição foi uma extensão do descritor CS-LMP, que foi chamado de MCS-LMP, na qual inclui o *pixel* central nos cálculos para melhor caracterizar o conteúdo das imagens. O descritor proposto foi comparado com outros três descritores: CS-LMP, SIFT e *Local Intensity Order Pattern* (LIOP) na tarefa de reconhecimento de cenas complexas. O descritor proposto alcançou uma melhor sensibilidade comparado aos outros descritores estudados. Além disso, seu tamanho foi reduzido pela quantização de *bins* do histograma.

A terceira e última contribuição foi o desenvolvimento do descritor *Mean Local Mapped Pattern* (M-LMP) que compara os *pixels* da vizinhança com a média de todos os *pixels* da mesma vizinhança. Tal descritor foi desenvolvido com o objetivo de eliminar possíveis ruídos causados pela comparação dos *pixels* da vizinhança apenas com o pixel central (como é feito no descritor LBP). Este descritor apresentou resultados superiores ao CS-LBP e SIFT, em termos de *matches* corretos e também no reconhecimento de objetos, faces e cenas complexas.

1.5 Organização da Tese

Os demais capítulos desta tese estão estruturados como segue.

Capítulo 2: apresenta a fundamentação teórica do trabalho. O objetivo é discutir o estado da arte e apresentar as limitações encontradas nas técnicas atuais, focalizando nos principais descritores já existentes na literatura e nas etapas para a descrição de uma imagem utilizando a metodologia BoF.

Capítulo 3: discute os três descritores propostos neste trabalho de doutorado, que permitem enriquecer a representação semântica de uma imagem.

Capítulo 4: apresenta o material e a metodologia utilizada para correspondência de imagens e classificação e reconhecimento de objetos.

Capítulo 5: sumariza os resultados e contribuições alcançadas nesta tese.

Capítulo 6: recapitula o trabalho apresentado nesta tese com as conclusões e apresenta os artigos publicados e submetidos. Além disso, discute-se as limitações da abordagem proposta e apresenta-se as sugestões de trabalhos futuros visando a continuidade deste trabalho.

Fundamentação Teórica

2.1 Introdução

Neste capítulo são abordados alguns descritores de imagens amplamente estudados na literatura, que são considerados base para este trabalho (Seção 2.2). O descritor é responsável por extrair propriedades relevantes das imagens e armazená-las em vetores de características. Além disso, ele deve ser também robusto com relação a variações de escala, rotação e oclusões. A metodologia *Bag-of-Features* (BoF) também é apresentada em detalhes. A idéia de tal abordagem é quantizar os descritores locais em um conjunto de palavras visuais, em que um vetor de frequência da ocorrência das palavras visuais (que representa a imagem) é utilizado para detecção e classificação. Os detalhes da metodologia estão na Seção 2.3.

2.2 Extração de características

Na busca pela análise das características de uma imagem com o objetivo de descrever as características de um objeto estão os descritores de imagens. Nesta Seção são discutidos os seguintes descritores: SIFT (Sub-seção 2.2.1), CS-LBP (Sub-seção 2.2.2), os descritores ILBP e M-LBP (Sub-seção 2.2.3), LMP (Sub-seção 2.2.4) e o descritor LIOP (Sub-seção 2.2.5).

2.2.1 *Scale Invariant Feature Transform* (SIFT)

SIFT (*Scale Invariant Feature Transform*) é uma técnica que permite a detecção de pontos-chave (*keypoints*) e a extração de características locais em imagens.

A obtenção dos descritores SIFT pode ser dividida nas seguintes etapas:

1. Detecção de extremos.
2. Localização de pontos-chave.

3. Definição de orientação.
4. Descritor dos pontos-chave.

A primeira etapa consiste em buscar pontos que sejam invariantes a mudanças de escala da imagem. Para este objetivo, o método utiliza a função Gaussiana, que também é chamada de espaço de escala.

É realizada a convolução da imagem I com a função Gaussiana G , com desvio padrão σ conforme mostram as equações 1 e 2.

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (1)$$

sendo que

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad (2)$$

O método também amplia a busca por pontos-chave utilizando a função DoG (*Difference of Gaussians*). A função DoG é dada pela diferença de imagens filtradas pela Gaussiana em escalas próximas, separadas por uma constante k . Tal função é definida de acordo com a Equação 3.

$$DoG = G(x, y, k\sigma) - G(x, y, \sigma) \quad (3)$$

Realiza-se, então, a convolução da função DoG com a imagem I conforme mostra a Equação 4.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (4)$$

A criação da Diferença das Gaussianas é esquematizada na Figura 1 de acordo com os seguintes passos:

1. Um fator de escala k no espaço de escala é utilizado para produzir na imagem inicial convoluções incrementais com filtros gaussianos. A verificação deste fato são representados na coluna da esquerda.
2. É necessário fazer a convolução da imagem até 2σ para que os descritores sejam invariantes a escala. Portanto, para se obter em s intervalos, o fator de escala k é definido por $k = 2^{1/s}$, produzindo assim $s + 3$ imagens na oitava.
3. As imagens produzidas na coluna da esquerda são subtraídas para produzir as imagens da Diferença das Gaussianas mostradas à direita.
4. Logo após o processamento da oitava, a resolução da imagem é diminuída, considerando-se cada segundo *pixel* da imagem no centro da oitava, gerando uma nova oitava. Volta-se ao passo 1.

Figura 1: Representação do procedimento de obtenção das Diferenças de Gaussianas DoG para diversas oitavas de uma imagem (LOWE, 2004).

Na segunda etapa, realiza-se a detecção dos extremos em cada intervalo da oitava. Um extremo é definido por valores de máximo ou mínimo locais para cada $D(x, y, \sigma)$. Compara-se a intensidade de cada ponto com as intensidades de seus oito vizinhos na escala atual, e com os nove vizinhos na escala superior e inferior, representados na Figura 2, onde o ponto “X” é comparado com seus vizinhos marcados como “bolinhas verdes”.

Figura 2: Detecção de extremos no espaço-escala (LOWE, 2004).

Na terceira etapa, localiza-se então, os pontos-chave e excluem-se os pontos instáveis. Para calcular a localização exata dos pontos-chave é preciso ajustar uma função quadrática 3D do ponto de amostragem local de modo a determinar uma localização interpolada do máximo.

Após a determinação dos pontos-chave são utilizadas as informações de magnitude e orientação do gradiente da região ao redor de cada ponto. As magnitudes $m(x, y)$ do gradiente e orientações $\theta(x, y)$ são calculadas ao redor dos pontos-chave $L(x, y)$ pelas equações 5 e 6 respectivamente.

$$m(x, y) = \sqrt{(L(x+1, y) - L(x-1, y))^2 + (L(x, y+1) - L(x, y-1))^2} \quad (5)$$

$$\theta(x, y) = \tan^{-1}((L(x, y+1) - L(x, y-1)) / (L(x+1, y) - L(x-1, y))) \quad (6)$$

Gera-se um histograma das orientações para cada região ao redor do ponto-chave. Cada ponto da vizinhança do ponto-chave é adicionado ao histograma com um valor de peso já estipulado. O descritor é formado por um vetor contendo as magnitudes de todas as orientações dos histogramas, como apresentado na Figura 3. As matrizes utilizadas são de tamanho 4×4 , resultando em um vetor de 128 posições.

Figura 3: A imagem da esquerda são os gradientes da imagem. A imagem da direita o descritor de pontos-chave. (LOWE, 2004).

Apesar do descritor SIFT ser um dos descritores mais aplicados na literatura, ele é muito sensível a mudanças geométricas e de iluminação. Alguns trabalhos tentam suprir essa deficiência encontrando mais pontos na imagem com diferentes mudanças de iluminação (ALITAPPEH; SARAVI; MAHMOUDI, 2012; ALITAPPEH; MAHMOUDI, 2013). A redução de dimensionalidade do descritor SIFT é bastante discutida na literatura e envolve técnicas como PCA, *Fisher Linear Discriminant Analysis* (FLDA), *Local Fisher Discriminant Analysis* (LFDA), *Isometric Mapping* (ISOMAP), *Locally Linear Embedding* (LLE) e *Locality Preserving Projections* (LPP) (VALENZUELA; SCHWARTZ; PEDRINI, 2012; ZHUO; CHENG; ZHANG, 2014) e a escolha de um *kernel* apropriado para projeção (ZHAO et al., 2010).

2.2.2 Center Symmetric Local Binary Pattern (CS-LBP)

O método CS-LBP é um descritor similar ao SIFT, com características individuais e diferentes (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2006). As características são extraídas

por um operador similar ao *Local Binary Pattern* (LBP) e possuem propriedades desejáveis, tais como tolerância à mudança de iluminação, robustez em áreas finas das imagens e simplicidade computacional (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009). Afirmar-se em (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009) que o LBP produz histogramas muito grandes e, portanto, difíceis de serem usados no contexto de descritor de imagens. Para resolver este problema, ao invés de comparar ao *pixel* central toda uma vizinhança, compara-se apenas os pares de *pixels* simétricos em relação ao centro, conforme é ilustrado na Figura 4. Nota-se que para 8 vizinhos, o LBP produz no máximo 256 padrões binários diferentes ($padrao_{LBP} \in [0, 255]$), enquanto que para o CS-LBP, este número se reduz a 16 ($padrao_{CS-LBP} \in [0, 15]$).

Figura 4: Características dos métodos LBP e CS-LBP para uma vizinhança de 8 *pixels* (Adaptado de (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009))

Os detectores de regiões de interesse, tais como *Hessian-Affine*, *Harris-Affine*, *Hessian-Laplace* e *Harris-Laplace* (MIKOLAJCZYK et al., 2005; MIKOLAJCZYK; SCHMID, 2004), fornecem as regiões que são usadas para calcular os descritores. Uma vez detectadas, as regiões são mapeadas para uma região circular de raio constante para se obter invariância à escala e transformação afim. Estas regiões são do tamanho 41×41 *pixels* e os valores das intensidades dos *pixels* são normalizados entre o intervalo $[0, 1]$. Para se obter invariância à rotação, cada região é rotacionada em relação ao gradiente dominante.

Para a construção do descritor, as regiões da imagem são divididas em células e determina-se um histograma de características para cada célula. Para a determinação deste histograma, extraem-se as características de cada *pixel* da região utilizando o CS-LBP por meio da Equação 7:

$$CS - LBP_{R,N,T}(x, y) = \sum_{i=0}^{(N/2)-1} s(n_i - n_{i+(N/2)})2^i, \quad (7)$$

sendo que n_i e $n_{i+(N/2)}$ correspondem aos valores de níveis de cinza dos pares simétricos em relação ao centro, com N *pixels* igualmente espaçados em um círculo de raio R e

$$s(x) = \begin{cases} 1, & \text{se } x > T, \\ 0, & \text{caso contrário.} \end{cases}$$

Uma interpolação bilinear é usada para distribuir o peso de cada característica em *bins* adjacentes do histograma, evitando, assim, os efeitos de contorno. O resultado do descritor é um histograma 3D de localizações e características. O histograma da região de interesse é gerado por meio da concatenação dos histogramas obtidos para cada célula e posterior normalização para obter nível máximo unitário para cada *bin*. Com o intuito de minimizar a influência dos elementos de alta repetibilidade (os quais nem sempre possuem maior informação sobre a região), limita-se o nível de cada *bin* em 0,2 e normaliza-se novamente o histograma tal como antes. Para uma malha cartesiana 4×4 , o vetor de características final conterà 256 posições ($4 \times 4 \times 16$) (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009).

2.2.3 Os métodos ILBP e M-LBP

O método ILBP é uma versão modificada do LBP e sua aplicação está em detecção de faces (JIN et al., 2004). Como o operador LBP compara apenas os *pixels* da vizinhança com o *pixel* central, o método ILBP propõe comparar todos os *pixels* (incluindo o *pixel* central) com a média de intensidade dos *pixels* da região.

A formulação do ILBP é apresentada na equação 8.

$$ILBP_{N,R}(x, y) = \sum_{p=0}^{N-1} s(g_p - g_{mean})2^P + s(g_c - g_{mean})2^N, \quad (8)$$

na qual

$$g_{mean} = \frac{1}{N+1} \left(g_c + \sum_{p=0}^{N-1} g_p \right) \quad (9)$$

e

$$s(x) = \begin{cases} 1, & \text{se } x > T, \\ 0, & \text{caso contrário.} \end{cases}$$

onde g_p são os pixels da vizinhança, g_c é o pixel central e N a quantidade de vizinhos.

O método ILBP considera N amostras de pontos uniformemente distribuídas em um círculo com raio R do *pixel* central, conforme mostra a Figura 5.

83	75	126	$\longrightarrow$ Comparação com a média (100.1)	0	0	1
99	95	141		0	0	1
91	91	100		0	0	0

Figura 5: ILBP (MARCEL; RODRIGUES; HEUSCH, 2007)

O método *Mean Local Binary Pattern*(M-LBP) é similar ao ILBP, mas não considera o *pixel* central. Neste método, a média da região é calculada e usada como um limiar,

ao invés de usar o nível de cinza do *pixel* central (BAI; ZHU; DING, 2008). Um exemplo é apresentado na Figura 6.

Figura 6: M-LBP

2.2.4 Local Mapped Pattern (LMP)

Baseado em estudos anteriores sobre a metodologia LFP, recentemente foi proposta a abordagem *Local Mapped Pattern* (CHIERICI et al., 2013) que descreve uma imagem utilizando uma função de mapeamento específica.

Considerando o número de níveis de cinza de uma imagem digitalizada igual a 256, há 256^9 padrões possíveis para uma vizinhança 3×3 . Histogramas de altas dimensões fornecem uma estimativa incerta da distribuição de características e não tem nenhum poder discriminante na descrição da imagem (LATEGAHN et al., 2010). A maneira mais simples de reduzir a dimensionalidade do histograma seria diminuir o número de níveis de cinza. Contudo, o número de *bins* requerido aumenta significativamente mesmo com poucos níveis de cinza. Por exemplo, usando quatro níveis de cinza em uma vizinhança 3×3 , o resultado seria $4^9 = 262144$ *bins*. Todos os métodos baseados em diferenças de níveis de cinza tem o problema de definir uma partição do espaço de padrões em padrões discriminantes, juntando os *bins* do histograma. O número de classes no qual o espaço de padrão é subdividido representa a dimensionalidade de cada método.

O problema central é definir um critério desejável para particionar o espaço de distribuição de características. Muitas soluções já foram propostas para gerar descritores de textura que compartilham o mesmo princípio. Estes métodos são conceitualmente mais simples, fáceis de implementar e razoavelmente rápidos. Alguns deles são muito populares na comunidade de visão computacional, tais como o LBP e suas variações.

Para a abordagem *Histograms of Equivalent Patterns* (HEP) (FERNANDEZ; ALVAREZ; BIANCONI, 2013), a geração de um descritor de textura é uma questão de determinar uma função f que defina uma partição desejável do espaço de distribuição de características. Contudo, a discretização multi-nível proposta é limitada e o método ignora as pequenas diferenças dentro de uma vizinhança.

A metodologia LMP assume que cada distribuição de nível de cinza dentro de uma vizinhança da imagem é um padrão local. Este padrão pode ser representado pela diferença de níveis de cinza ao redor do *pixel* central. A Figura 7 mostra um exemplo de uma vizinhança 3×3 e um gráfico 3D do respectivo padrão local.

Figura 7: Diferenças de níveis de cinza como um padrão local

Cada padrão definido por uma vizinhança $W \times W$ será mapeado para um *bin* do histograma h_b usando a Equação 10, sendo que $f_{g(i,j)}$ é uma função de mapeamento, $P(k,l)$ é uma matriz de pesos de valores pré-definidos para cada posição do *pixel* dentro de uma vizinhança, e B é o número de *bins* do histograma. Esta equação representa a soma do peso de cada diferença do nível de cinza entre os *pixels* da vizinhança e o *pixel* central, mapeado no intervalo $[0, 1]$ por uma função de mapeamento e arredondado para B *bins* possíveis.

$$h_b = \text{round} \left(\frac{\sum_{k=1}^W \sum_{l=1}^W (f_{g(i,j)} P(k,l))}{\sum_{k=1}^W \sum_{l=1}^W P(k,l)} (B - 1) \right), \quad (10)$$

Pela Equação 10 é possível derivar algumas abordagens já publicadas anteriormente para análise de padrões locais. O método *Local Binary Pattern* (LBP) pode ser derivado da Equação 10 usando a função Heaviside apresentada na Equação 11:

$$f_{g(i,j)} = H[A(k,l) - g(i,j)] \quad (11)$$

sendo que

$$H[A(k,l) - g(i,j)] = \begin{cases} 0 & \text{se } [A(k,l) - g(i,j)] < 0; \\ 1 & \text{se } [A(k,l) - g(i,j)] \geq 0. \end{cases}$$

Levando em consideração o LBP básico com uma vizinhança de 3×3 *pixels*, a matriz de pesos será:

$$P(k,l) = \begin{bmatrix} 1 & 2 & 4 \\ 128 & 0 & 8 \\ 64 & 32 & 16 \end{bmatrix}$$

O LBP é assim um caso particular da abordagem LMP.

Usando diferentes funções de mapeamento, a metodologia LMP pode extrair características específicas da imagem usando padrões locais de processamento.

Para análise de textura, a abordagem LMP usa uma aproximação suave para a função Heaviside por uma função logística, uma curva logística, ou uma curva sigmóide, como apresentado na Equação 12.

$$f_{g(i,j)} = \frac{1}{1 + e^{\frac{-[A(k,l) - g(i,j)]}{\beta}}}, \quad (12)$$

sendo que β é a inclinação da curva e $[A(k,l) - g(i,j)]$ são as diferenças de níveis de cinza dentro de uma vizinhança centrada em $g(i,j)$. A matriz de pesos proposta é:

$$P(k,l) = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Usando a curva sigmóide, o método é chamado LMP-s.

Funções sigmóides, nas quais os gráficos são curvas em forma de S, aparecem em uma grande variedade de contextos, tais como funções de transferências de muitas redes neurais. Sua ubiquidade não é acidental, estas curvas estão entre as curvas não lineares mais simples, encontrando um equilíbrio entre o comportamento linear e não linear. A Figura 8 mostra o mapeamento $f_{g(i,j)}$ usado na Equação 10 para as abordagens LMP-s (sigmóide) e LBP (função Heaviside).

(a) Função de mapeamento para o método LMP-s

(b) Função de mapeamento para o método LBP

Figura 8: Diferença do mapeamento dos níveis de cinza pela Equação 10. (a) *Local Mapped Pattern*. (b) *Local Binary Pattern*

2.2.5 Local Intensity Order Pattern (LIOP)

O método LIOP tem como objetivo codificar a informação de ordem local de cada *pixel* e essa informação é utilizada para dividir a região em sub-regiões que são usadas para acumular os LIOPs. Como a divisão das regiões e o cálculo do LIOP são baseados em relações relativas de intensidade, o descritor é invariante a mudanças de rotação e intensidade monotônicas. Porém, o tamanho do seu descritor é maior do que o SIFT, gerando um tempo maior de processamento das imagens (HUANG et al., 2013).

A Figura 9 ilustra o funcionamento do algoritmo. Primeiramente, a imagem é suavizada por um filtro Gaussiano, e um detector de regiões de interesse é utilizado (tal como *Harris-Affine* ou *Hessian-Affine*). Como as regiões de interesse detectadas têm variações de tamanho e formas, elas são normalizadas para regiões circulares de tamanho fixo. Para a divisão das regiões, o método baseia-se na ordem de intensidade. Especificamente, todos os *pixels* na região são ordenados por sua intensidade em ordem decrescente. Então, a região local é igualmente quantizada em B *bins* ordinários de acordo com sua ordem. A Figura 9c) ilustra a divisão da região baseada na ordem de intensidades, onde cada *bin* é marcado com uma cor diferente.

Figura 9: O método LIOP (WANG; FAN; WU, 2011)

Diversos estudos já foram feitos com o intuito de investigar a proposta dos autores com aplicações em imagens com mudanças de intensidade (QI et al., 2013; LEI; YI; LI, 2014; SALEEM; BAIS; SABLATNIG, 2014). Os resultados mostraram que o descritor *Local Intensity Order* (LIO) (QI et al., 2013) é robusto a variação de iluminação, porém foi testado apenas em imagens de faces. Já o descritor *Local Gradient Local Pattern* (LGOP) (LEI; YI; LI, 2014) considera a ordem das respostas do gradiente na região local para obter uma representação robusta da face. E por fim, vários descritores foram testados com relação a correspondência de imagens e comparados ao descritor LIOP (SALEEM; BAIS; SABLATNIG, 2014) em imagens de cenas, apresentando um desempenho inferior na avaliação das curvas *recall* $\times$ *precision*.

2.3 *Bag-of-Features* (BoF)

O modelo BoF tem mostrado um desempenho excepcional em uma ampla variedade de tarefas, tais como reconhecimento de ações (ULLAH; PARIZI; LAPTEV, 2010), textura (ZHANG et al., 2007), gestos (DARDAS; CHEN; GEORGANAS, 2010), classificação de imagens (NOWAK; JURIE; TRIGGS, 2006a), etc. Sua maior vantagem é devido à sua simplicidade e

facilidade de implementação, onde imagens podem ser representadas por palavras visuais, agrupadas e representadas por um histograma com as respectivas ocorrências.

Esta técnica foi introduzida na área de recuperação de imagens por Sivic e Zisserman (2003). Os autores propuseram esta abordagem para localizar todas as ocorrências de um objeto nas cenas de um vídeo. O objeto é representado por um conjunto de descritores de regiões invariantes ao ponto de vista e eles utilizam o método de agrupamento *k-means* para a construção do vocabulário visual. A partir daí novos trabalhos foram propostos para as mais diversas aplicações. No trabalho de Batista et al. (2009) foi utilizada a metodologia BoF para identificação automática de imagens que contêm fachadas e edifícios no acervo digitalizado do Arquivo Público Mineiro. Já no trabalho de Lopes et al. (2009a) os autores utilizaram o método para identificar nudez em vídeos; Li, Imai e Kaneko (2010) utilizaram BoF baseado em blocos para reconhecimento de faces; Lopes et al. (2009b) compararam três descritores utilizando a abordagem BoF para reconhecimento de ações humanas. Wang et al. (2011c) aplicaram o BoF pesando as palavras visuais na recuperação de imagens médicas. Já em Wang et al. (2011b) os autores desenvolveram um novo algoritmo para o BoF chamado *Joint Vivo*, e classificam imagens de densidade do tecido mamário em mamografias. Ainda na área médica, Pedrosa et al. (2012) reduziram o tamanho do descritor SIFT utilizando a entropia de Shannon, e utilizaram o método BoF para classificar imagens médicas de raio-x e ressonância magnética. Outros trabalhos, como o de Sun et al. (2012) utilizaram o método para recuperação de fragmentos similares baseado em palavras visuais extraídas de um conjunto limitado de dados, para proteção de direitos autorais de animações. Zhang e Sawchuk (2011) apresentaram um *framework* baseado no método BoF para a tarefa de reconhecimento de atividade humana. Mais recentemente, na área de multimídia, Abid, Melo e Petriu (2013) trabalharam com linguagem de sinais e reconhecimento de voz para aplicações em *smart home*. Os autores utilizaram o descritor HOG3D para extração de características, o agrupamento *k-means++* e o classificador SVM para classificar gestos da mão. Wu et al. (2012) apresentaram um novo modelo que enriquece o conceito de reconhecimento de objetos. Os autores utilizaram o modelo BoF para o reconhecimento de placas de carros, extraíndo as características pelo descritor SIFT. Barata, Marques e Mendonça (2013) propuseram dois sistemas para a detecção de melanomas em imagens dermatópicas, onde o primeiro sistema usou métodos globais para classificar lesões de pele, e o segundo utilizou características locais e o método BoF para classificação das imagens. Li, Wang and Zhang (2016) utilizaram o descritor SURF e a pirâmide espacial numa metodologia Bag-of-Features para melhorar o reconhecimento e classificação de imagens.

A abordagem BoF é uma técnica que modela as características locais de uma imagem. Cria-se um dicionário visual, onde as características locais extraídas das imagens são tratadas como palavras visuais deste dicionário. As características locais são mapeadas à sua respectiva palavra visual no dicionário, podendo-se descrever uma imagem como

um vetor-de-características de acordo com uma contagem da frequência com que cada palavra visual aparece na imagem. Sendo assim, o processo necessário para representar uma imagem pelo método BoF pode ser definido pelas seguintes fases:

- ❑ Detecção de regiões de interesse de uma imagem (*Hessian-Affine*, *Harris-Affine*, etc).
- ❑ Representação de cada região de interesse, como um vetor de características, por meio de um descritor de características (SIFT, SURF, LBP, etc).
- ❑ Agrupamento dos vetores de características por um método de agrupamento, como por exemplo o *k-means*, gerando o dicionário de palavras visuais.
- ❑ Atribuição das características a uma palavra visual do dicionário.
- ❑ Contagem das ocorrências de cada palavra visual contida na imagem, e geração do histograma.

As próximas seções descrevem com mais detalhes cada passo do processo de implementação do método BoF.

2.3.1 Detecção de regiões de interesse e descritores de imagens

A primeira fase da abordagem BoF é a detecção de regiões de interesse de uma imagem. Tais regiões são padrões locais representativos, e a partir deles são criados os vetores que representam diversas partes da imagem.

Existem diversas técnicas na literatura para a detecção de tais regiões de interesse. Uma delas é a divisão por quadrantes (MARÉ et al., 2005), chamada de grade regular. Outro método seria os pontos de interesse ou *keypoints*, que são caracterizados por terem alguma semântica envolvida.

As pesquisas mostram que o maior desafio é encontrar pontos invariantes a rotação, escala, iluminação, possibilitando a localização dos mesmos pontos em imagens distintas.

Os detectores de pontos de interesse mais comuns na literatura são: Diferença de Gaussianas (DOG) (LOWE, 1999), Laplaciano da Função Gaussiana (LOG) (LINDEBERG, 1998), *Harris-Laplace* e *Affine* (MIKOLAJCZYK; SCHMID, 2004) e *Hessian-Laplace* e *Affine* (MIKOLAJCZYK et al., 2005).

Os descritores *Harris-affine* e *Hessian-Affine* são algoritmos quase idênticos, ambos derivados dos trabalhos de Krystian Mikolajczyk e Cordelia Schmid em 2002 (MIKOLAJCZYK; SCHMID, 2002). O detector *Harris-affine* utiliza múltiplas escalas usando a medida de *Harris corner* na matriz do segundo momento relacionada a autocorrelação. O detector *Hessian-Affine* também utiliza um algoritmo em múltiplas escalas, contudo em cada escala o detector escolhe pontos de interesse baseado na matriz Hessiana.

Após a detecção de regiões de interesse de uma imagem, um vetor de características deverá ser gerado, representando numericamente cada região de interesse. Esse vetor é chamado de descritor de imagem. Neste sentido, cada região de interesse é representada por um ponto no espaço $\mathbb{R}^n$, para n características. É desejável que um descritor seja invariante a transformações afins, para que haja uma melhor correspondência entre imagens semelhantes. Tais descritores já foram abordados na Seção 2.2.

2.3.2 Dicionário de palavras visuais

A determinação do dicionário de palavras visuais é de extrema importância no método BoF, pois ela será responsável por determinar quais características representam a estrutura de uma imagem.

A estratégia de geração do dicionário de palavras visuais está esquematizada na Figura 10 e é definida pelos seguintes passos: primeiramente um conjunto de imagens é escolhido da base de dados; para cada imagem é utilizado um detector de regiões de interesse (*Hessian-Affine*, *Harris-Affine*) e as regiões são descritas com algum descritor (SIFT, CS-LBP, etc), tal como discutido na Seção anterior; por fim, é realizado um agrupamento de dados desse espaço de características utilizando algum algoritmo de agrupamento. Cada agrupamento terá um centróide, o qual será considerado uma palavra visual do dicionário. Os algoritmos de agrupamento podem ser divididos em três categorias: hierárquicos, por particionamento e incrementais. Tais algoritmos serão descritos na Sub-seção 2.3.2.1.

(a) Detector *Hessian-Affine*

(b) Construção do descritor e quantização pelo *k-means*

Figura 10: Estratégia de geração do dicionário no algoritmo BoF

Um fator importante que afeta o desempenho do método BoF é o tamanho do dicionário de palavras visuais. Um vocabulário pequeno tem pouco poder discriminativo, pois dois agrupamentos podem ser atribuídos a uma mesma palavra visual. Por outro lado, um vocabulário grande é pouco generalista e afeta o desempenho da abordagem (SIVIC; ZISSERMAN, 2003; ZHANG et al., 2007).

2.3.2.1 Algoritmos de agrupamento

Algoritmos de agrupamento são técnicas de classificação não supervisionada de dados em grupos. Tal agrupamento é baseado na similaridade entre os dados, sendo que dados de um mesmo grupo são mais similares entre si do que dados pertencendo a outros grupos. Eles são divididos em duas categorias: hierárquicos e por particionamento (JAIN; MURTY; FLYNN, 1999).

Nos algoritmos hierárquicos, os grupos vão sendo formados por aglomerações ou divisões de elementos, normalmente sendo representados por uma árvore. Uma vantagem destes algoritmos seria a facilidade em lidar com qualquer medida de similaridade utilizada e sua consequente aplicabilidade a qualquer atributo. As desvantagens seriam em relação ao critério de parada impreciso e ao não refinamento das soluções obtidas durante a execução do algoritmo (BERKHIN, 2002).

Os algoritmos mais comuns são o *single-link* e o *complete-link* (JAIN; MURTY; FLYNN, 1999) que são diferentes na maneira de como calculam a similaridade entre um par de dados. Os formatos dos grupos também serão diferentes. Pelo *complete-link* os grupos são mais compactos e pelo *single-link* são mais alongados (Figura 11).

(a) Algoritmo *single-link* para 2 classes

(b) Algoritmo *complete-link* para 2 classes

Figura 11: Algoritmos de agrupamento (JAIN; MURTY; FLYNN, 1999)

O algoritmo *single-link* é um algoritmo do tipo aglomerativo, ou seja, todos os dados são grupos individuais e, recursivamente os dados vão se agrupando.

O algoritmo *single-link* (LACHI; ROCHA, 2005) pode ser descrito da seguinte maneira:

1. Defina cada padrão como sendo um grupo;
2. Construa uma lista das distâncias entre padrões, para todos os pares de padrões;
3. Ordene esta lista em ordem crescente;

4. Percorra essa lista de distâncias do seguinte modo: para cada valor d_k da lista de distâncias, construa um grafo onde, os pares de padrões cujos valores são mais próximos de d_k são conectados por uma aresta;
5. Se todos os padrões são membros de um grafo conexo, então, pare. Caso contrário repita todos esses passos.

Veja que no *single-link* a similaridade é definida como máxima entre os membros dos grupos. No algoritmo *complete-link*, é utilizado o mesmo esquema do algoritmo *single-link*, porém a similaridade é definida como mínima entre os membros dos grupos.

Nos algoritmos por particionamento o conjunto de elementos é dividido em k subconjuntos, alocando inicialmente os elementos aos grupos. Tais algoritmos são vantajosos em aplicações que envolvem um grande número de conjuntos, devido a capacidade computacional de se criar a hierarquia (LACHI; ROCHA, 2005; PIRES, 2008).

Um dos grandes problemas desses algoritmos é a escolha do número ideal de grupos. Para minimizar tais problemas, são utilizados índices de validação, ou seja, os algoritmos são executados repetidas vezes, utilizando diferentes valores para k , e o melhor resultado obtido com o índice é utilizado como resultado do agrupamento. Alguns índices utilizados são: *PBM-index* (PAKHIRAA; BANDYOPADHYAY; MAULIK, 2004), Xie-Beni (XIE; BENI, 1991) e Calinski-Harabasz (CALINSKI; HARABASZ., 1974).

Um dos métodos bem conhecidos é o *k-means* (MACQUEEN, 1967) que utiliza os centróides para representar os grupos, o qual é calculado pela média de todos os objetos do grupo.

O algoritmo *k-means* pode ser descrito pelos seguintes passos (Figura 12) (FONTANA; NALDI, 2009):

- ❑ Atribuem-se valores iniciais para os protótipos, como por exemplo, sorteio aleatório desses valores dentro dos limites de domínio de cada atributo;
- ❑ Atribui-se cada objeto ao grupo cujo protótipo possua maior similaridade com o objeto;
- ❑ Recalcula-se o valor do centróide de cada grupo, determinando a média dos objetos atuais do grupo;
- ❑ Repetem-se os passos 2 e 3 até a estabilização dos grupos;

Um dos maiores problemas do *k-means* é que ele é sensível a seleção da partição inicial e pode convergir para um mínimo local caso a partição inicial não seja bem escolhida (JAIN; MURTY; FLYNN, 1999). Outra desvantagem é a sua sensibilidade em relação a erros ou elementos muito diferentes dos demais. Caso haja um dado errôneo, ele pode influenciar drasticamente o centro do seu grupo, interferindo na análise do resultado.

(a) Sorteio dos valores iniciais dos centróides

(b) Atribuição dos objetos aos grupos cujo centróides possuem maior similaridade

(c) Recalcula os valores dos protótipos como sendo a média dos objetos atuais pertencentes ao grupo

(d) Preparação para repetição dos passos 2 e 3. Disvincula-se os objetos dos grupos.

(e) Repetição do passo 2 - Atribuição dos objetos aos grupos cujo protótipo possua maior similaridade

(f) Repetição do passo 3 - Recalcula os valores dos protótipos como sendo a média dos objetos atuais pertencentes ao grupo

Figura 12: Operação do algoritmo *k-means* (FONTANA; NALDI, 2009)

Elkan *k-means* (ELKAN, 2003) tem como objetivo acelerar o método *k-means* aplicando uma desigualdade triangular de duas maneiras diferentes para evitar desnecessários cálculos de distâncias. Lloyd *k-means* (LLOYD, 1982) usa um diagrama de Voronoi ao invés de determinar o centro mais próximo do conjunto de pontos.

2.3.3 Histograma de palavras visuais

Definido o dicionário de palavras visuais, o próximo passo é definir o histograma das palavras visuais. Para cada imagem, é definido um histograma, que indica com que frequência as palavras visuais (caracterizadas pelos centróides) deste dicionário ocorrem na mesma.

Esta fase pode ser realizada por duas abordagens diferentes: *hard-assignment* (SIVIC; ZISSERMAN, 2003) ou *soft assignment* (JIANG; NGO; YANG, 2007; PHILBIN et al., 2008). Na primeira abordagem, também chamada de *nearest-neighbor*, cada palavra visual (representada por seu descritor) é associada ao grupo mais próximo. A Figura 13 ilustra bolas azuis representando os centros dos grupos, e a bola vermelha seria a palavra visual que deverá ser associada a algum dos grupos (no caso, o grupo número 2).

Figura 13: Exemplo da abordagem *hard assignment*

Na abordagem *soft assignment*, um descritor pode estar próximo de mais de um centro do grupo. Neste caso atribuem-se os k centros mais próximos a ele. Esta abordagem também é chamada de *k-nearest neighbor*. A Figura 14 apresenta um exemplo do método. Supondo $k = 4$, o histograma mostra que os 4 centros contribuem para a construção do histograma. Neste caso a abordagem atribui um peso de tal forma que um centro do grupo mais próximo tem maior relevância do que um mais distante.

O trabalho de Lazebnik et al. (2006) realiza uma extensão da abordagem BoF e constrói histogramas baseado no modelo “pirâmide espacial”. O processo subdivide repetidamente uma imagem e calcula histogramas das características das imagens em cada sub-região. A Figura 15 mostra o exemplo da construção da pirâmide em três níveis.

Figura 14: Exemplo da abordagem *soft assignment*

A imagem possui três tipos de características que foram subdivididas em três níveis de resolução. Para cada nível de resolução, e para cada bloco conta-se as características que se enquadram para cada *bin* espacial com o objetivo de contruir o histograma para cada bloco.

Figura 15: Exemplo da pirâmide espacial de Lazebnik et al. (2006)

Os resultados desta metodologia mostram uma melhor performance em comparação ao BoF original e evidencia a importância de se trabalhar com o contexto de *patches* locais para categorização visual.

Muitos pesquisadores estudaram esta abordagem para diversas aplicações envolvendo *tracking* (WANG et al., 2011a), reconhecimento de objetos (WANG et al., 2014) e detecção de objetos (LIU; LI, 2015).

2.3.4 Classificadores

As técnicas de classificação de imagens consistem em separá-las colocando-as em grupos pré-definidos. Os critérios de classificação baseiam-se nos atributos das imagens. O tipo de classificação pode ser dividido em duas categorias: supervisionada e não-supervisionada (DUDA; HART; STORK, 2000). A técnica não-supervisionada funciona sem a etapa de treinamento, utilizando critérios de semelhança de objetos, como por exemplo, a métrica de distância. Já na técnica supervisionada, as classes são previamente definidas pelo usuário, englobando os dados de treinamento no classificador. Estes classificadores se dividem em paramétricos e não-paramétricos (LANDGREBE, 2003).

Nos métodos paramétricos uma função densidade de probabilidade se supõe conhecida descrevendo o comportamento dos dados, como por exemplo o método Máxima Verossimilhança (MAXVER). Já nos métodos não-paramétricos a função densidade de probabilidade não é conhecida, como por exemplo nos Sistemas Fuzzy, SVM e k-NN.

No método Máxima Verossimilhança, utiliza-se de estatísticas de treinamento para calcular a probabilidade de um *pixel* pertencer a uma determinada classe. Examina-se o *pixel* para cada classe e atribui-se o *pixel* à classe com maior probabilidade. Porém, para que a classificação seja precisa o suficiente é necessário um número elevado de *pixels* para o conjunto de treinamento. O método deverá ser aplicado quando o analista conhecer bem as imagens a serem classificadas, para que se possa determinar classes que sejam representativas (CRÓSTA, 1993). Além disso, a classificação de dados em alta dimensionalidade pode se tornar um problema para métodos paramétricos, por causa do grande número de parâmetros a serem estimados. Sendo assim, desperta-se o interesse para métodos não-paramétricos como o SVM (HUANG; DAVIS; TOWNSHEND, 2002). Esta técnica busca minimizar o erro com relação ao conjunto de treinamento e de teste. O objetivo é obter um equilíbrio entre esses erros, minimizando o excesso de ajustes com relação às amostras de treinamento e aumentando a capacidade de generalização do classificador (VAPNIK, 1999). Alguns estudos mostram que SVM binários, aplicados em problemas multi-classe, apresentam uma acurácia melhor que outras técnicas de reconhecimento de padrões utilizadas na literatura (HUANG; DAVIS; TOWNSHEND, 2002; MELGANI; BRUZZONE, 2004; FOODY; MATHUR, 2008). Outras propostas, utilizando algoritmos genéticos para estimação de parâmetros, são apresentadas para aumentar a acurácia do classificador SVM (GUO et al., 2008; BAZI; MELGANI, 2006).

2.4 Considerações Finais

Neste capítulo foi apresentada uma visão geral de alguns descritores existentes na literatura, e como eles são utilizados numa metodologia BoF. A metodologia BoF apresenta muitos parâmetros que afetam diretamente seu desempenho. Sendo assim, a análise das técnicas envolvidas é de extrema importância para o bom funcionamento do sistema.

Na metodologia, primeiramente detectam-se as regiões de interesse de uma imagem. Dependendo do tipo de aplicação, há uma necessidade de se averiguar a melhor técnica para este propósito, pois, dependendo da técnica escolhida, há um número muito grande de pontos-chave gerados, aumentando o tempo computacional. Logo após, um descritor deverá ser escolhido e gerado para cada ponto-chave encontrado no primeiro passo. O descritor deverá ter um tamanho reduzido para um melhor processamento, e satisfazer algumas características importantes, como invariância a iluminação, escala e rotação. Tais descritores serão agrupados para a geração do dicionário visual. Quanto menor o dicionário, mais rápido o processamento. Porém, o dicionário tem que caracterizar da melhor maneira possível as imagens em análise. Normalmente a técnica utilizada para este propósito é o *k-means*. O histograma de palavras visuais será gerado de acordo com a abordagem escolhida, com evidências que a contrução espacial apresenta melhores resultados do que a abordagem original do BoF. Foram apresentadas duas abordagens que são amplamente utilizadas na metodologia. E, por fim, um classificador deverá ser escolhido para classificar de uma maneira rápida e precisa as imagens em múltiplas classes.

Descritores Propostos

3.1 Introdução

Para se alcançar os objetivos deste trabalho, foram desenvolvidos três descritores de imagens, *Center-Symmetric Local Mapped Pattern* (CS-LMP), *Modified Center-Symmetric Local Mapped Pattern* (MCS-LMP) e *Mean Local Mapped Pattern* (M-LMP). Os descritores propostos passaram pelo processo de redução de dimensionalidade, e analisou-se sua eficiência por meio de curvas $recall \times precision$ e pela metodologia BoF para a classificação e reconhecimento de objetos em imagens. Tais descritores foram comparados aos métodos CS-LBP, SIFT e LIOP, mostrando serem robustos nas bases de dados escolhidas. Para o agrupamento foram aplicados os métodos *k-means*, *Elkan k-means* e *Lloyd k-means*, sendo que o parâmetro k foi escolhido empiricamente por sucessivas execuções do algoritmo à procura da melhor sensibilidade. Para a classificação das imagens foram utilizados os métodos *k-nearest neighbor* e SVM. Os descritores propostos foram testados para o reconhecimento de objetos, cenas e faces.

3.2 *Center-Symmetric Local Mapped Pattern* (CS-LMP)

O descritor CS-LBP apresentado na Seção 2.2.2 apresenta a seguinte desvantagem: a diferença entre os pares simétricos em relação ao centro são limiarizadas para 0 ou 1, não capturando nuances que podem ocorrer na imagem.

Analisando a vantagem de flexibilidade do método LMP (Seção 2.2.4), pela sua maleabilidade na mudança da função de mapeamento, no número de *bins* e na matriz de pesos da Equação 10, e a desvantagem do descritor CS-LBP (rigidez na utilização da função degrau), propõe-se um novo descritor, CS-LMP, utilizando a metodologia do CS-LBP associada a uma função de mapeamento apropriada proposta pelo descritor LMP.

A Equação 13 define a extração de características de cada *pixel* da região de interesse, definida em uma vizinhança $W \times W$,

$$CS - LMP_{R,N,f}(x, y) = \text{round} \left(\frac{\sum_{i=0}^{N/2-1} f(i)P(i)}{\sum_{i=0}^{N/2-1} P(i)} b \right) \quad (13)$$

na qual b representa o número de *bins* do histograma menos um, $P(i)$ é o i -ésimo elemento da matriz de pesos P definida na Equação 14, e $f(i)$ é o mapeamento definido pela função sigmóide da Equação 15.

$$\mathbf{P} = [p_0 p_1 p_2 \dots p_{N/2-1}]^t \quad (14)$$

sendo que

$$p_i = 2^i$$

$$f(i) = \frac{1}{1 + e^{\frac{-[n_i - n_{i+(N/2)}]}{\beta}}} \quad (15)$$

na qual n_i e $n_{i+(N/2)}$ são as intensidades de nível de cinza dos pares simétricos ao centro, com N *pixels* igualmente espaçados em um círculo de raio R e β define a inclinação da curva sigmóide. Se as coordenadas de n_i do círculo não correspondem às coordenadas da imagem, o nível de cinza deste ponto é interpolado.

A Figura 16 ilustra o funcionamento do método CS-LMP fixando-se o valor de b em 15 (16 *bins* menos um).

Figura 16: Exemplo de geração de um código CS-LMP para uma vizinhança de 8 *pixels*

Neste trabalho, as regiões de interesse da imagem são detectadas utilizando-se os detectores *Hessian-Affine*, *Harris-Affine* e *Hessian-Laplace*. Estas regiões são então mapeadas para uma região circular de raio constante e rotacionadas na direção do gradiente dominante (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009; LOWE, 2004). Divide-se cada região

em células com uma malha de tamanho 4×4 . Extraem-se as características para cada *pixel* da região de interesse utilizando-se o descritor CS-LMP e constrói-se um histograma de características para cada célula.

Os parâmetros R , N e β foram estabelecidos empiricamente, e os melhores resultados obtidos serão discutidos no Capítulo 5.

Realiza-se, ainda, uma interpolação bilinear (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2006), obtendo como resultado um histograma tridimensional de localizações x e y e de características do tipo CS-LMP em z , conforme ilustrado na Figura 17. O histograma final (descritivo da região de interesse) é formado pela concatenação dos histogramas determinados para cada célula, obtendo-se um vetor de 256 posições ($4 \times 4 \times 16$). Normaliza-se o histograma final tal como realizado para o CS-LBP, apresentado na Seção 2.2.2.

Figura 17: Construção do descritor CS-LMP

3.3 Modified Center-Symmetric Local Mapped Pattern (MCS-LMP)

Baseado no descritor CS-LMP, o descritor MCS-LMP tem como objetivo utilizar a informação do *pixel* central nos cálculos matemáticos para melhor caracterizar o conteúdo

da imagem. O novo descritor é definido na Equação 16 e tem como objetivo a extração de características para cada *pixel* da região de interesse definida na vizinhança: $W \times W$:

$$\text{MCS-LMP}_{R,N,f}(x, y) = \text{round} \left(\left(\frac{\sum_{i=0}^{N/2-1} f(i)P(i) + g(n_c)P(N/2)}{\sum_{i=0}^{N/2-1} P(i) + P(N/2)} \right) b \right) \quad (16)$$

onde b é o número de *bins* do histograma menos um, $P(i)$ é o i th elemento da matriz de peso P definida na Equação 17, $f(i)$ é o mapeamento da função sigmóide definida na Equação 18 e $g(n_c)$ é o mapeamento da função sigmóide definida na Equação 19.

$$\mathbf{P} = [p_0 p_1 p_2 \dots p_{N/2-1} p_{N/2}]^t \quad (17)$$

onde

$$p_i = 2^i$$

$$f(i) = \frac{1}{1 + e^{\frac{-[n_i - n_{i+(N/2)}]}{\beta}}} \quad (18)$$

onde n_i e $n_{i+(N/2)}$ correspondem aos valores do nível de cinza dos pares simétricos ao centro de N *pixels* igualmente espaçados em um círculo de raio R e β é a inclinação da curva sigmóide. Se as coordenadas de n_i no círculo não corresponderem as coordenadas das imagens, os valores dos níveis de cinza neste ponto é interpolado.

$$g(n_c) = \frac{1}{1 + e^{\frac{-[n_c - n_{mean}]}{\beta}}} \quad (19)$$

onde n_{mean} é definido pela Equação 20 e n_c é o *pixel* central.

$$n_{mean} = \frac{1}{N+1} \sum_{i=0}^{N-1} n_i + n_c \quad (20)$$

A Figura 18 ilustra a operação do MCS-LMP quando o parâmetro b foi ajustado para o valor 7.

Adotou-se o detector *Hessian-Laplace* proposto por Mikolajczyk e Schmid (2004), que usa a matriz Hessiana para localizar os pontos no espaço e a função Laplaciana para calcular sua escala. As regiões são mapeadas em regiões circulares de raio constante e rotacionadas na direção do gradiente dominante (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009) e (LOWE, 2004). Cada região é dividida em células com o tamanho da malha de 4×4 . As características são extraídas para cada *pixel* da região de interesse pelo descritor MCS-LMP e um histograma é construído para cada célula.

Uma interpolação bilinear é feita sobre as dimensões x e y para balancear o peso das características entre os quatro vizinhos mais próximos (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2006). Portanto, um histograma 3D das localizações das características e valores do

Figura 18: Características do MCS-LMP para uma vizinhança de 8 *pixels*.

MCS-LMP é construído, conforme mostra a Figura 19. O descritor é construído pela concatenação dos histogramas calculados para as células e um vetor de 112 posições ($4 \times 4 \times 7$) é obtido. O descritor é então normalizado conforme o descritor CS-LBP (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2006).

Figura 19: Descritor MCS-LMP

3.4 Mean Local Mapped Pattern (M-LMP)

No LBP original, as características são obtidas apenas comparando-se os níveis de cinza dos *pixels* de uma vizinhança com o nível de cinza do *pixel* central. Isto pode gerar ruídos no resultado final. Para evitar este problema e tornar as características mais robustas, é proposto neste trabalho calcular a média dos *pixels* da região ao invés de utilizar apenas o nível de cinza de um único *pixel* como valor de comparação (BAI; ZHU; DING, 2008). Assim, o método M-LMP é determinado pela comparação dos *pixels* de uma vizinhança circular com a média dos *pixels* desta região, conforme é mostrado na Equação 21,

$$M - LMP_{R,N,f}(x, y) = \text{round} \left(\frac{\sum_{i=0}^{N-1} f(i)P(i)}{\sum_{i=0}^{N-1} P(i)} b \right) \quad (21)$$

na qual b representa o número de *bins* do histograma menos um, $P(i)$ é o i -ésimo elemento da matriz de pesos P definida na Equação 22, e $f(i)$ é o mapeamento definido pela função sigmóide na Equação 23,

$$\mathbf{P} = [p_0 p_1 p_2 \dots p_{N-1}]^t \quad (22)$$

sendo

$$p_i = 2^i$$

$$f(i) = \frac{1}{1 + e^{\frac{-(n_i - n_{mean})}{\beta}}} \quad (23)$$

na qual n_i corresponde ao valores de níveis de cinza de uma vizinhança de N *pixels* igualmente espaçados em um círculo de raio R , β é a inclinação da função sigmóide e n_{mean} é definida pela Equação 24.

$$n_{mean} = \frac{1}{N} \sum_{i=0}^{N-1} n_i \quad (24)$$

A Figura 20 mostra um exemplo da geração das características M-LMP para uma região circular de raio $R = 2$ e $N = 8$ *pixels*. Neste exemplo, o número de *bins* é 16, e assim $b = 15$ na Equação 21, gerando códigos no intervalo $[0, 15]$.

A construção do descritor é realizada da mesma maneira que o descritor CS-LMP apresentado na Seção 3.2. Os parâmetros R , N e β foram empiricamente estabelecidos e serão detalhados no Capítulo 5.

3.5 Considerações Finais

Neste capítulo foram apresentados três novos descritores de imagens baseado na análise de padrões locais. O primeiro denominado CS-LMP foi inspirado nos descritores LMP e

Figura 20: Exemplo de geração de código M-LMP para uma vizinhança de 8 *pixels*

CS-LBP. Neste caso, a generalização que o operador LMP proporciona, amplia o leque de possibilidades na escolha da função utilizada para a descrição das imagens. Enquanto o CS-LBP trabalha com a função degrau, foi escolhido para o CS-LMP a função sigmóide, que é uma suavização da função degrau. Sendo assim, para regiões não homogêneas, consegue-se minimizar os picos altos, e consequentemente ruídos, para a construção dos histogramas.

O segundo descritor, chamado MCS-LMP foi uma extensão do descritor CS-LMP, incluindo a informação do *pixel* central para melhor caracterizar o conteúdo das imagens. O terceiro descritor, chamado M-LMP, também segue a linha do descritor CS-LMP, mas ao invés de comparar os pares de *pixels* simétricos ao centro, ele compara os *pixels* com a média da vizinhança. O cálculo da média apesar de ser mais custoso tem a vantagem de suavizar a região, e consequentemente eliminar ruídos que possam interferir no processo de classificação.

Os descritores propostos nesta tese tem muitas características que os tornam adequados para a tarefa de detecção de objetos: um histograma relativamente pequeno, baixo custo computacional e habilidade para mudar o tamanho do descritor.

Material e Metodologia

4.1 Banco de Imagens

4.1.1 Banco de imagens para Correspondência de Imagens

Nesta Sub-seção é apresentado o banco de imagens utilizado para o critério de avaliação dos descritores CS-LMP e M-LMP baseado nos experimentos utilizando correspondência de imagens.

Para a avaliação das correspondências de imagens, foram retiradas imagens de dois bancos de imagens disponíveis na Internet ^{1 2}. As imagens utilizadas para o ajuste do parâmetro β são mostradas na Figura 21, e possuem diferentes tipos de transformações: mudança de ponto de vista, mudança de escala, rotação, mudança de iluminação, compressão JPEG e efeito de imagem borrada (totalizando 16 imagens). A Figura 22 apresenta as imagens de teste utilizadas neste experimento, com transformações de escala e rotação (totalizando 30 imagens).

4.1.2 Banco de imagens utilizados na metodologia BoF

O banco de imagens Caltech 2004 ³ foi utilizado para dois experimentos com o descritor M-LMP. Primeiramente foram utilizadas quatro classes de imagens da base Caltech 2004 que incluem 1074 aviões, 1155 carros, 450 faces e 826 motos. A Figura 23 mostra um exemplo de uma imagem de cada categoria.

Para o segundo experimento com a base Caltech 2004 o número de classes foi aumentado para 6 incluindo-se as classes “guitarra” e “casa”. A Figura 24 mostra um exemplo de cada classe.

O banco de imagens Pascal VOC 2006 ⁴ foi utilizado para a realização de um expe-

¹<http://www.robots.ox.ac.uk/vgg/research/affine/>

²<http://lear.inrialpes.fr/people/mikolajczyk/>

³<http://www.robots.ox.ac.uk/vgg/data3.html>

⁴<http://www.pascal-network.org/challenges/VOC/voc2006>

Figura 21: Imagens utilizadas para o ajuste do parâmetro β : *graf* (mudança de ponto de vista), *bark* (escala e rotação), *trees* (imagem borrada), *leuven* (mudança de iluminação), *ubc* (compressão JPEG), *boat* (escala e rotação), *bike* (imagem borrada), *wall* (mudança de ponto de vista)

rimento com o descritor M-LMP utilizando a metodologia BoF. Foi utilizada a base de dados Pascal VOC 2006 com as seguintes classes: ônibus, vaca, cavalo e ovelha. A Figura 25 mostra um exemplo de cada classe.

Foram realizados dois experimentos com as imagens da base Caltech 101 (FEI; FERGUS; PERONA, 2004) utilizando o descritor M-LMP. Primeiramente foram escolhidas 4 classes da base Caltech 101: *bonsai*, *faces_easy*, *leopards* e *watch*, com imagens com variação de rotação, iluminação, escala, compressão e *aliasing*. Exemplos das imagens de cada classe são mostradas na Figura 26.

Foi também analisado o desempenho do descritor M-LMP utilizando a metodologia proposta nesta tese para a classificação de objetos de 5 classes da base Caltech 101: *bonsai*, *faces_easy*, *leopards*, *watch* e *airplanes*. Algumas das imagens são mostradas na Figura 27.

Três banco de imagens de faces foram utilizados para experimentos com a metodologia proposta utilizando o descritor M-LMP. A base de faces Feret (PHILLIPS; RAUSS, 1997) com 45 classes e 40 imagens por classe, onde as imagens escolhidas para cada classe incluem: expressão facial regular (fa), expressão facial alternativa (fb), posição esquerda

(a) *resid1* (b) *resid4* (c) *east park1* (d) *east park4* (e) *east south1* (f) *east south4* (g) *ensimag1* (h) *ensimag4* (i) *laptop1* (j) *laptop4* (k) *zmars1* (l) *zmars4* (m) *zmonet1* (n) *zmonet4* (o) *new york1* (p) *new york4* (q) *axterix1* (r) *axterix4* (s) *belledone1* (t) *belledone4* (u) *bip1* (v) *bip4* (w) *crolle1* (x) *crolle4* (y) *lap1* (z) *lap4* (aa) *vangogh1* (ab) *vangogh4* (ac) *inria1* (ad) *inria10*

Figura 22: Imagens do conjunto de teste: *resid*, *east park*, *east south*, *ensimag*, *laptop* e *Inria* (escala e rotação), *zmars*, *zmonet*, *new york* (rotação), *axterix*, *belledone*, *bip*, *crolle*, *lap* e *vangogh* (escala)

Figura 23: Exemplo de uma imagem de cada categoria da Base de Dados Caltech 2004

Figura 24: Exemplo de imagens de cada uma das seis classes da base de dados Caltech 2004, usadas no segundo experimento.

Figura 25: Exemplo de imagens de cada classe da base de dados Pascal VOC 2006.

Figura 26: Exemplo de uma imagem de cada categoria da Base de Dados Caltech 101

Figura 27: Exemplo de uma imagem de cada categoria da Base de Dados Caltech 101

e direita (hl e hr), perfil esquerdo e direito (ql e qr), imagens randômicas (ra, rb, rc, rd e re) e algumas de faces com mudanças de escala. A Figura 28 mostra alguns exemplos por classe.

Figura 28: Exemplo de uma imagem de cada categoria da Base de faces Feret

Para a base de faces ORL ⁵ foram utilizadas 40 classes com 10 imagens por classe. As imagens possuem diversas características tais como mudança de iluminação, expressão facial e detalhes faciais (com óculos e sem óculos), e tiradas com um fundo escuro, com tolerância para algumas mudanças de pose. A Figura 29 apresenta três exemplos de 6 classes da base de faces ORL.

E por fim, a base de faces *Yale Face* ⁶ foi utilizada nos experimentos e contém 15 classes com 11 imagens por classe. As configurações das faces em cada classe são: com

⁵<http://www.cl.cam.ac.uk/>

⁶<http://vision.ucsd.edu/content/yale-face-database>

Figura 29: Exemplo de 3 imagens de 6 categorias da Base de faces ORL

iluminação centralizada, com óculos, sem óculos, feliz, com iluminação a esquerda, normal, com iluminação a direita, triste, sonolento, surpreso e piscando os olhos. A Figura 30 mostra três exemplos de 6 classes da base de imagens *Yale Face*.

Figura 30: Exemplo de 3 imagens de 6 categorias da Base de faces *Yale Face*

Para a avaliação do desempenho dos descritores M-LMP e CS-LMP foram utilizadas duas bases de imagens: Caltech 101 ⁷ e ImageNet ⁸, onde foram escolhidas aleatoriamente 27 classes contendo 80 imagens por classe. A Figura 31 mostra 3 exemplos de cada classe para a base de imagens Caltech 101. As classes utilizadas foram: airplane, background_google, bonsai, butterfly, brain, buddha, car_side, chandelier, ewer, faces, face_easy, grand_piano, hawksbill, helicopter, ibis, kangaroo, ketch, laptop, leopards,

⁷[http://www.vision.caltech.edu/Image Datasets/Caltech101/](http://www.vision.caltech.edu/Image%20Datasets/Caltech101/)

⁸<http://www.image-net.org/>

menorah, motorbikes, revolver, scorpion, starfish, sunflower, trilobite e watch. Já a Figura 32 mostra 3 exemplos de cada classe para a base de imagens ImageNet. As classes utilizadas foram: *airplane*, *blender*, *bonsai*, *butterfly*, *car_elevator_car*, *car_wheel*, *train*, *chandelier*, *coffee*, *ewer*, *face*, *fish*, *grand_piano*, *hawksbill*, *helicopter*, *horse*, *ketch*, *laptop*, *leopard*, *menorah*, *pen*, *revolver*, *school_bus*, *scorpion*, *sunflower*, *watch* e *wine*.

Figura 31: Exemplos de imagens de cada classe da Base de Imagens Caltech 101

Figura 32: Exemplos de imagens de cada classe da Base de Imagens ImageNet

O último banco de imagens utilizado nos experimentos foi o banco *Indoor Scene Recognition dataset*⁹. Foram utilizadas 8 classes, com 200 imagens por classes escolhidas aleatoriamente. O descritor MCS-LMP foi avaliado aplicando a metodologia proposta para classificação de imagens. A Figura 33 mostra alguns exemplos das imagens de cada classe onde imagens intra-classes são altamente diferentes umas das outras.

⁹<http://web.mit.edu/torralba/www/indoor.html>

(a) *Bakery* (b) *Bedroom* (c) *Casino* (d) *Corridor* (e) *Grocery Store* (f) *Kitchen* (g) *Restaurant* (h) *Toystore*

Figura 33: Exemplos de imagens de cada classe da base de imagens *Indoor Scene Recognition dataset*.

4.2 Metodologia

4.2.1 Metodologia utilizada para correspondência de imagens

Dois dos descritores de imagens propostos neste trabalho, CS-LMP e M-LMP, foram comparados com os descritores da literatura CS-LBP e SIFT aplicados a correspondência de pares de imagens, sendo utilizada a estratégia *1-nearest-neighbor* (1-NN), em que, para um determinado par de imagens, cada região na imagem 1 tem apenas um correspondente na imagem 2 e vice-versa.

O banco de imagens utilizado foi apresentado na Seção 4.1.1.

Para os testes, foram utilizados dois detectores de regiões de interesse: *Hessian-Affine* e *Harris-Affine*.

4.2.2 Metodologia utilizada para classificação das imagens

Foram desenvolvidas duas metodologias para os experimentos com o descritores propostos.

Metodologia-1: A Figura 34 mostra o *framework* utilizado para os experimentos com os descritores propostos nesta tese. As imagens foram divididas entre *codebook*, conjunto de treinamento e conjunto de testes. Foi aplicado o detector de regiões de interesse *Hessian-Affine* em cada uma das imagens. Para cada região foram calculados os descritores M-LMP, CS-LBP e SIFT. Os descritores foram agrupados usando o método *k-means*. Tal método tem sido amplamente usado para agrupar descritores em modelos BoF (STREICHER; BURKHARDT; FEHR, 2009; YU et al., 2013; ELSHOUBAGY; HEMAYED; FAYEK, 2016).

Durante o treinamento foram utilizadas as imagens do conjunto de treinamento e aplicado o mesmo procedimento utilizado na geração do *codebook*. A frequência de ocorrência de cada palavra no *codebook* é armazenada em um histograma, descrevendo um objeto. Em outras palavras, cada imagem (ou objeto) é representado por um histograma de frequência de “palavras visuais”.

Na fase de teste foram utilizadas imagens do conjunto de testes e aplicados os mesmos procedimentos para a geração do *codebook*. Cada imagem foi representada por um histograma de frequências da ocorrência das “palavras visuais” armazenadas no *codebook*. Usando a busca pelo vizinho mais próximo (*k-nearest neighbor*) encontra-se o objeto mais próximo a ele.

Metodologia-2: A metodologia anteriormente utilizada para reconhecimento de objetos foi a metodologia BoF com o classificador K-NN, porém tal classificador é muito sensível a *outliers*. Como a metodologia utilizada neste trabalho extrai “características de textura globais”, a seleção de características é uma tarefa difícil por causa dos *outliers*. O classificador SVM têm a tendência de trabalhar melhor do que o K-NN pois ele usa

Figura 34: *Framework* do método *Bag of Features (BoF)*

pontos relevantes para encontrar a separação das classes. Sendo assim, propõe-se uma variante da metodologia BoF baseado em histograma espacial e SVM.

A Figura 35 mostra os passos utilizados para a metodologia proposta nesta tese.

Primeiramente foram extraídas regiões de interesse nas imagens utilizando detectores de regiões de interesse tais como: *Hessian-Affine*, *Hessian Multiscale* (MIKOLAJCZYK; SCHMID, 2002), *dense sampling* (ISCEN et al., 2015) ou *Hessian-Laplace*. *Hessian Multiscale* usa o determinante da matriz Hessiana em multi-escala para detectar características em múltiplas escalas. No detector *dense sampling*, o centro da região de interesse é estabelecido fazendo um deslocamento a cada 2 *pixels* e assim é construído uma malha regular de tamanho e forma fixa. Cada região de interesse de tamanho 40×40 *pixels* é dividida em uma malha de tamanho 4×4 células, e as características são extraídas para cada região de interesse utilizando o descritor M-LMP. Alguns descritores usam pontos-chave multi-escala para detectar as regiões de interesse (LOWE, 2004), mas alguns trabalhos mostram que *dense sampling* ou *random sampling* superam os métodos de pontos-chave (ISCEN et al., 2015; NOWAK; JURIE; TRIGGS, 2006b).

Foram selecionadas aleatoriamente 30 imagens e diferentes pontos-chave foram localizados para cada imagem, gerando diferentes descritores ao redor deles. Tais descritores foram clusterizados utilizando as metodologias *Elkan k-means* (versão otimizada do método *k-means* (MACQUEEN, 1967)) ou *Lloyd k-means* (LLOYD, 1982) que utiliza o diagrama

(a) Detector de regiões por *Hessian-Affine*

(b) Construção do descritor e quantização do vetor por *Elkan k-means*

(c) Representação do histograma de um objeto (palavras visuais)

Figura 35: *Framework* da metodologia proposta para o reconhecimento de faces humanas

de Voronoi ao invés de determinar o centro mais próximo dos conjuntos de pontos. O número de *clusters* variam entre 200 e 600 e será discutido no Capítulo 5.

Depois da geração do vocabulário foi contruído o histograma de palavras visuais usando a representação de pirâmide espacial descrita no trabalho de (LAZEBNIK; SCHMID; PONCE, 2006). Tal representação divide a imagem em sucessivos blocos em cada nível e calcula os descritores para cada bloco. Os histogramas espaciais são usados como entrada para o classificador SVM (VEDALDI; FULKERSON, 2010). Foi calculado um mapeamento de característica explícito que aproxima um *kernel* não linear χ^2 como um kernel linear (MIKOLAJCZYK; SCHMID, 2004) usando transformação de dados. A regra um-para-todos foi utilizada, ou seja, um classificador é treinado para separar cada classe do restante e a imagem teste que tiver a mais alta resposta é associada ao *label* do classificador.

4.3 Critério de Avaliação

Nesta Seção é apresentado o critério de avaliação para os descritores propostos baseado nos experimentos utilizando correspondência de imagens e a metodologia BoF. Na Sub-seção 4.3.1 é apresentado o critério de avaliação utilizando os descritores CS-LMP e M-LMP para correspondência de imagens, e na Sub-seção 4.3.2 o critério de avaliação utilizado para a classificação das imagens.

4.3.1 Critério de avaliação para correspondência de imagens

O desempenho dos descritores CS-LMP e M-LMP são avaliados e comparados aos métodos CS-LBP e SIFT, utilizando-se um critério de avaliação baseado nos trabalhos apresentados por Heikkilä, Pietikäinen e Schmid (2006) e Heikkilä, Pietikäinen e Schmid (2009). Tal critério considera o número de correspondências corretas e falsas para um par de imagens, sendo conhecido *a priori* o conjunto de verdadeiras correspondências (*ground-truth*). Heikkilä Pietikäinen e Schmid (2009) estabelecem o *ground-truth* por meio do erro de sobreposição de duas regiões, que determina o quanto uma região A da imagem 1 se sobrepõe a uma região B da imagem 2, sendo a imagem 2 projetada sobre a imagem 1 por uma matriz de homografia H. O erro de sobreposição é definido pelo raio de intersecção e união das duas regiões (HEIKKILÄ; PIETIKÄINEN; SCHMID, 2009), conforme descrito na Equação 25.

$$\epsilon_S = 1 - (A \cap H^T B H) / (A \cup H^T B H) \quad (25)$$

O conjunto de verdadeiras correspondências é formado pelas correspondências cujo erro de sobreposição é menor que 50%, ou seja, $\epsilon_S < 0.5$.

Para avaliar o desempenho dos descritores propostos, os resultados são apresentados no formato de curvas *recall* $\times$ *precision* (BAEZA-YATES; RIBEIRO-NETO, 1999) e suas respectivas *area under curve* (AUC).

Precision, definido pela Equação 26, representa o número de corretas correspondências n_c em relação ao número total de possíveis correspondências retornadas n (corretas + falsas).

$$precision = \frac{n_c}{n} \quad (26)$$

Nesta equação n representa o número total de possíveis correspondências para um dado limiar de distância Euclideana entre os vetores de características do par de regiões das imagens, ou seja, é o número de pares de regiões cuja distância Euclideana entre os vetores de características é menor do que um limiar t_d ; e n_c é o número de pares correspondentes cuja distância Euclideana entre os vetores de características é menor do que t_d e ainda cujo erro de sobreposição ϵ_S é menor que 0.5.

O valor *recall*, definido pela Equação 27, representa a razão entre o número de correspondências corretas n_c e o número total de verdadeiras correspondências M (*ground-truth*).

$$recall = \frac{n_c}{M}, \quad (27)$$

Para a construção da curva *recall* $\times$ *precision*, cada ponto é definido escolhendo-se $n \in [1, N]$, sendo que N representa o número de possíveis correspondências.

Na análise da curva *recall* $\times$ *precision*, quanto mais próxima a curva estiver do canto superior direito do gráfico, melhor é o descritor, sendo desejável *recall* próximo de 1 com boa precisão.

4.3.2 Critério de avaliação para a classificação das imagens

Para todos os experimentos realizados foram geradas as matrizes de confusão com os verdadeiros positivos (VP), ou o número de imagens corretamente classificadas, e falsos negativos (FN), ou as imagens classificadas erroneamente, para todo o conjunto de imagens. Para os resultados quantitativos, foi utilizada a sensibilidade (Equação 28), e a proporção de imagens classificadas erroneamente conforme a Equação 29.

$$S = \frac{VP}{VP + FN} \quad (28)$$

$$Err = \frac{FP + FN}{VP + FN + FP + VN} \quad (29)$$

sendo que FP são os falsos positivos e VN são os verdadeiros negativos.

4.4 Considerações Finais

Os métodos envolvidos para a análise dos descritores propostos envolve critérios importantes para a correspondência de imagens, juntamente com algumas métricas já conhecidas na literatura. Além disso, duas metodologias baseadas na abordagem BoF foram utilizadas para o reconhecimento de objetos, cenas e faces. No próximo Capítulo serão apresentados os resultados obtidos nesta tese.

Resultados

5.1 Introdução

Os descritores propostos foram comparados com os descritores CS-LBP, SIFT e LIOP, e os resultados são apresentados neste Capítulo.

As imagens das bases testadas foram utilizadas na sua totalidade, ou seja, não houve nenhum pré-processamento ou recorte das imagens, como é feito por vários trabalhos da literatura (KRIZHEVSKY; SUTSKEVER; HINTON, 2012; SIMONYAN; VEDALDI; ZISSERMAN, 2013; ZEILER; FERGUS, 2014; SERMANET et al., 2014; SZEGEDY et al., 2015; SIMONYAN; ZISSERMAN, 2015). Sendo assim, há um grande desafio em classificar as imagens em uma base de imagens quando o “objeto principal” que está presente em algum lugar da imagem define a classe da imagem. O amplo espectro das categorias dos objetos motiva a necessidade do desenvolvimento de algoritmos capazes de distinguir classes que são visualmente similares. Neste trabalho foram utilizadas todas as imagens em suas resoluções originais. O objetivo foi localizar o “objeto principal” presente na imagem toda, mesmo que o tamanho do objeto seja pequeno, e classificar cada imagem em sua devida classe.

5.2 Correspondência de imagens

Os descritores CS-LMP e M-LMP foram comparados com dois descritores amplamente conhecidos e divulgados na literatura (SIFT e CS-LBP) para geração dos resultados.

A Tabela 1 apresenta a avaliação do parâmetro β nos métodos CS-LMP e M-LMP para a imagem *bark*. O parâmetro β foi testado em intervalos de valores pré-definidos gerando-se as curvas *recall* $\times$ *precision* e medindo suas respectivas *area under curve* (AUC). Para a maioria das imagens da Figura 21, o melhor valor de β foi próximo de 0.6 para o CS-LMP e 1.2 para o M-LMP.

O detector de regiões de interesse usado foi o *Hessian-Affine*. Os parâmetros estabelecidos para o CS-LMP foram $\beta = 0,6$, $b = 15$, $R = 2$ e $N = 8$. Os parâmetros para

Tabela 1: Avaliação do parâmetro β para a imagem exemplo *bark* usando *Hessian-Affine*

CS-LMP								
β	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8
AUC	0,8993	0,9073	0,9100	0,9173	0,9173	0,9196	0,9163	0,8960
M-LMP								
β	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7
AUC	0,8862	0,8862	0,9032	0,8862	0,8791	0,8960	0,8980	0,8862

o M-LMP foram $\beta = 1, 2$, $b = 15$, $R = 2$ and $N = 8$. Para o CS-LBP, usaram-se os parâmetros $R = 2$, $N = 8$ and $T = 0,01$ conforme definidos no trabalho de Heikkilä, Pietikäinen e Schmid (2006).

A Tabela 2 apresenta a avaliação do parâmetro β para o descritor M-LMP usando o detector de regiões de interesse *Harris-Affine* para a imagem *trees*. Usando-se o conjunto de treinamento variou-se o parâmetro β da mesma maneira que no caso anterior, gerando-se as curvas *recall* $\times$ *precision*, e medindo suas respectivas *area under curve* (AUC). Para a maioria das imagens do conjunto de treinamento, o melhor valor de β foi próximo 1,0 para o M-LMP.

Tabela 2: Avaliação do parâmetro β para a imagem exemplo *trees* usando *Harris-Affine*

M-LMP								
β	1,0	1,1	1,2	1,3	1,4	1,5	1,6	1,7
AUC	0,3439	0,3422	0,3422	0,3414	0,336	0,3322	0,3332	0,3291

Os parâmetros estabelecidos para o M-LMP foram: $\beta = 1, 0$, $b = 15$, $R = 2$ e $N = 8$.

A Figura 36 mostra as curvas *recall* $\times$ *precision* resultantes da aplicação dos descritores CS-LMP, M-LMP, SIFT e CS-LBP sobre quatro pares de imagens do conjunto de testes da Figura 22 utilizando o detector *Hessian-Affine*. Percebe-se que todos os descritores garantem boa precisão para baixo *recall*, ou seja, os primeiros pares de regiões retornadas são correspondências corretas; no entanto, à medida que se aumenta o número de pares de regiões retornadas, diminui-se a precisão, sendo mais acentuada a queda para o CS-LBP e SIFT, ou seja, para uma maior quantidade de correspondências retornadas (n), o CS-LMP e o M-LMP apresentam um maior número de correspondências corretas (n_c) que o CS-LBP e SIFT.

A *Area Under Curve* (AUC) são apresentadas na Tabela 3 para os descritores CS-LMP, M-LMP, CS-LBP e SIFT. Nota-se que para todas as imagens, os descritores propostos CS-LMP e M-LMP tem um valor AUC maior do que os métodos SIFT e CS-LBP. O CS-LMP retorna os melhores resultados para 9 imagens testes, enquanto que o M-LMP é melhor para as outras 6 imagens.

Utilizando o detector de regiões de interesse *Harris-Affine* nas imagens do conjunto de teste da Figura 22, foram comparados os métodos M-LMP, CS-LBP e SIFT. A Figura

Figura 36: Resultados das curvas $recall \times precision$ para os descritores CS-LMP, M-LMP, CS-LBP e SIFT usando o detector de regiões de interesse *Hessian-Affine*

Tabela 3: *Area under curve* (AUC) para as imagens do conjunto de teste da Figura 22 usando o detector de regiões de interesse *Hessian-Affine*

Imagem	CS-LMP	M-LMP	CS-LBP	SIFT
<i>resid</i>	0,9043	0,9090	0,8332	0,6846
<i>east park</i>	0,8758	0,8801	0,7842	0,7021
<i>east south</i>	0,9027	0,9021	0,7816	0,7762
<i>ensimag</i>	0,7459	0,7260	0,5994	0,5533
<i>laptop</i>	0,9219	0,9374	0,8891	0,7972
<i>zmars</i>	0,9659	0,9702	0,9499	0,8684
<i>zmonet</i>	0,9659	0,9669	0,9364	0,8323
<i>new york</i>	0,9731	0,9667	0,9584	0,8536
<i>axterix</i>	0,9577	0,9322	0,9237	0,8164
<i>belledonne</i>	0,7878	0,7561	0,6704	0,3891
<i>bip</i>	0,8889	0,8876	0,8495	0,7069
<i>crolle</i>	0,9161	0,9085	0,8724	0,6281
<i>lap</i>	0,9702	0,9668	0,9483	0,8695
<i>vangogh</i>	0,9362	0,9342	0,9246	0,7955
<i>inria</i>	0,8748	0,8963	0,8110	0,6012

37 mostra o resultado da aplicação dos descritores M-LMP, SIFT e CS-LBP sobre quatro imagens do banco de imagens de teste. Nas quatro imagens o desempenho do descritor proposto M-LMP ultrapassou o desempenho dos outros descritores. Isto significa que o M-LMP gera um número maior de correspondências corretas comparado aos outros métodos.

Figura 37: Resultados das curvas $recall \times precision$ para os descritores M-LMP, CS-LBP e SIFT usando o detector de regiões de interesse *Harris-Affine*

A AUC são apresentadas na Tabela 4 para os descritores M-LMP, CS-LBP e SIFT. Note para que todas as imagens, M-LMP tem um valor AUC maior do que os métodos SIFT e CS-LBP.

5.2.1 Redução do tamanho dos descritores

As Figuras 38 e 39 mostram as curvas $recall \times precision$ para a imagem *bark*, variando-se os valores de b para os descritores CS-LMP e M-LMP respectivamente. O parâmetro b (número de *bins* menos um) das Equações 16 e 21 foi variado no intervalo $[7,15]$ para o CS-LMP e $[6,15]$ para o M-LMP. Na Tabela 5 são apresentados os valores da *area under curve* (AUC) para os métodos CS-LMP e M-LMP para a imagem do conjunto de teste chamada *bark* quando se varia o número de *bins* (b). A AUC para o método CS-LBP é 0,8491 e para o SIFT é 0,7925. Foi então escolhido o menor valor de b na qual a AUC é maior do que o CS-LBP e SIFT. Foi escolhido $b = 8$ para o CS-LMP e $b = 7$ para o

Tabela 4: Area under curve (AUC) para as imagens do conjunto de teste da Figura 22, usando o detector de regiões de interesse *Harris-Affine*

Imagem	M-LMP	CS-LBP	SIFT
<i>resid</i>	0,8431	0,7713	0,6531
<i>east park</i>	0,8172	0,7359	0,6375
<i>east south</i>	0,7779	0,6084	0,6442
<i>ensimgag</i>	0,5866	0,3478	0,4036
<i>laptop</i>	0,9130	0,8221	0,7641
<i>zmars</i>	0,9151	0,8836	0,7957
<i>zmonet</i>	0,9204	0,9112	0,7727
<i>new york</i>	0,9372	0,9053	0,7896
<i>axterix</i>	0,8439	0,8386	0,7469
<i>belledonne</i>	0,5509	0,5113	0,2635
<i>bip</i>	0,8616	0,8172	0,6926
<i>crolle</i>	0,9504	0,8973	0,7351
<i>lap</i>	0,9512	0,9252	0,8471
<i>vangogh</i>	0,8715	0,8377	0,6389
<i>inria</i>	0,7405	0,6474	0,4213

M-LMP. Os valores marcados em negrito referem-se aos menores valores de b que geram melhores valores de AUC que os dois métodos comparados (CS-LBP e SIFT).

(a) CS-LMP

(b) (Figura 38a ampliada) CS-LMP

Figura 38: Curvas $recall \times precision$ para os descritores CS-LMP, CS-LBP e SIFT para a imagem do conjunto de teste chamada *bark*.

A Figura 40 mostra o tempo de processamento para se comparar dois vetores de um par de imagens usando os descritores CS-LMP e M-LMP. Estas curvas foram obtidas usando-se a distância Euclideana na imagem do conjunto de treinamento chamada “*leuven*”. Conforme b aumenta, o tamanho do descritor aumenta e o tempo de comparação também cresce. Assim, pode-se notar que quanto menor o tamanho do descritor, mais rápido será o seu processamento.

A Tabela 6 mostra o tempo de processamento em milissegundos da distância Euclideana entre dois vetores de um par de imagens para as imagens do banco de teste *inria*, *new york* e *resid* para todos os descritores comparados. O valor usado para o parâmetro b foi

(a) M-LMP

(b) (Figura 39a ampliada) M-LMP

Figura 39: Curvas $recall \times precision$ para os descritores M-LMP, CS-LBP e SIFT para a imagem do conjunto de teste chamada *bark*.

Tabela 5: *Area under curve* para a imagem do conjunto de teste chamada *bark*, variando-se o número de *bins* (b).

<i>b</i>	CS-LMP	M-LMP
6	-	0,8353
7	0,8328	0,8561
8	0,8808	0,8760
9	0,9189	0,8468
10	0,8720	0,8516
11	0,9196	0,8570
12	0,8880	0,8687
13	0,9165	0,8888
14	0,8791	0,8960
15	0,9196	0,9032

(a) CS-LMP

(b) M-LMP

Figura 40: Tempo de processamento (segundos) versus número de *bins* (*b*) do descritor usando a distância Euclidiana nos métodos CS-LMP e M-LMP na imagem do conjunto de treinamento chamada *leuven*

8 para o CS-LMP e 7 para o M-LMP. O tempo de comparação para os métodos M-LMP e SIFT são muito próximos pois seus vetores tem o mesmo tamanho (128 posições). O CS-LMP tem o segundo menor tempo de processamento (seu vetor tem 144 posições), enquanto que o CS-LBP tem o pior tempo de processamento (256 posições no vetor de características).

Tabela 6: Tempo de processamento para comparação de dois vetores de um par de imagens com $b=8$ para o CS-LMP e $b=7$ para o M-LMP (milisegundos)

Imagem	CS-LMP	M-LMP	CS-LBP	SIFT
<i>inria</i>	0,014794	0,014677	0,017666	0,014717
<i>new york</i>	0,015166	0,014903	0,017987	0,014905
<i>resid</i>	0,017544	0,017307	0,020852	0,017558

Os parâmetros utilizados para o descritor CS-LBP foram $R = 2$, $N = 8$ e $T = 0,01$. Conforme mostrado na Tabela 1, utilizou-se $\beta = 0,6$ para o CS-LMP e $\beta = 1,2$ para o M-LMP.

Foram geradas as curvas $recall \times precision$ para todas as imagens do banco de imagens de teste e calculou-se a *area under curve* para cada uma delas. Na Figura 41 são apresentadas as curvas para quatro dessas imagens. Nota-se que há uma boa precisão para um baixo *recall* para os quatro descritores, isto é, os primeiros pares de regiões retornadas são correspondências corretas. Contudo, conforme o número de pares de regiões retornadas aumenta, a precisão diminui. Pode-se notar que a redução na precisão é mais acentuada para o SIFT e CS-LBP. Conforme o número de correspondências retornadas aumenta (n), os descritores CS-LMP e M-LMP alcançam um número maior de correspondências corretas (n_c) quando comparados ao SIFT e CS-LBP.

A AUC foi calculada pelas curvas $recall \times precision$ para todas as imagens do conjunto de teste da Figura 22. Estes resultados são mostrados na Tabela 7. Note que para todas as imagens, os descritores CS-LMP e M-LMP tem uma AUC maior do que o SIFT e CS-LBP, e ambos retornam os melhores resultados para 8 imagens do conjunto de imagens de teste.

A Tabela 8 apresenta as n melhores correspondências retornadas pelos quatro descritores. Por exemplo, para a imagem *resid*, considerando as primeiras 1127 correspondências retornadas, tem-se 573 corretas para o CS-LMP, 547 para o CS-LBP e 490 para o descritor SIFT. A imagem “*zmars*” apresenta a diferença de 169 correspondências corretas favoráveis ao CS-LMP e para a imagem *laptop*, o descritor M-LMP se destaca por 50 correspondências corretas sobre o CS-LBP.

A Figura 42 apresenta as correspondências corretas entre um par de imagens “*east park*”.

Tabela 7: *Area under curve* (AUC) para as imagens do banco de teste, calculadas com $b=8$ para o CS-LMP e $b=7$ para o M-LMP

Imagem	CS-LMP	M-LMP	CS-LBP	SIFT
<i>resid</i>	0,8909	0,8808	0,8332	0,6846
<i>east park</i>	0,8717	0,8670	0,7842	0,7021
<i>east south</i>	0,8689	0,8779	0,7816	0,7762
<i>ensimag</i>	0,7182	0,7404	0,5994	0,5533
<i>laptop</i>	0,9164	0,9214	0,8891	0,7972
<i>zmars</i>	0,9625	0,9680	0,9499	0,8684
<i>zmonet</i>	0,9704	0,9500	0,9364	0,8323
<i>new york</i>	0,9694	0,9679	0,9584	0,8536
<i>axterix</i>	0,9497	0,9398	0,9237	0,8164
<i>belledonne</i>	0,7561	0,7561	0,6704	0,3891
<i>bip</i>	0,8839	0,8581	0,8495	0,7069
<i>crolle</i>	0,9161	0,8828	0,8724	0,6281
<i>lap</i>	0,9497	0,9616	0,9483	0,8695
<i>vangogh</i>	0,9287	0,9362	0,9246	0,7955
<i>inria</i>	0,8444	0,8785	0,8110	0,6012

Tabela 8: Razão de correspondências corretas obtidas para as imagens do conjunto de teste.

Imagem	CS-LMP	M-LMP	CS-LBP	SIFT
<i>resid</i>	573/1127	569/1127	547/1127	490/1127
<i>east park</i>	638/1307	638/1307	589/1307	550/1307
<i>east south</i>	717/1305	726/1305	671/1305	644/1305
<i>ensimag</i>	378/1070	382/1070	357/1070	321/1070
<i>laptop</i>	299/523	342/523	292/523	263/523
<i>zmars</i>	1677/22001	1685/22001	1660/22001	1516/22001
<i>zmonet</i>	410/620	401/620	395/620	356/620
<i>new york</i>	555/674	556/674	549/674	490/674
<i>axterix</i>	275/385	271/385	266/385	236/385
<i>belledonne</i>	14/116	14/116	13/116	10/116
<i>bip</i>	476/828	465/828	460/828	402/828
<i>crolle</i>	56/160	55/160	54/160	41/160
<i>lap</i>	455/570	454/570	448/570	414/570
<i>vangogh</i>	271/389	274/389	270/389	236/389
<i>inria</i>	83/292	86/292	81/292	67/292

Figura 41: Curvas $recall \times precision$ para os descritores CS-LMP, M-LMP, CS-LBP e SIFT com a redução no número de *bins* (*b*) para $b=8$ (CS-LMP) e $b=7$ (M-LMP).

5.3 Resultados obtidos na localização e identificação de objetos

5.3.1 Resultados obtidos com o descritor M-LMP

Os resultados apresentados nesta Seção referem-se aos experimentos realizados com o descritor M-LMP com as seguintes base de imagens: Caltech 2004, Pascal VOC 2006, Caltech 101, Feret, *Yale Face* e ORL (Seção 4.1).

Foram realizados sete experimentos: dois experimentos com imagens da base Caltech 2004 ¹, um experimento com a base Pascal VOC 2006 ², dois experimentos com as imagens da base Caltech 101 (FEI; FERGUS; PERONA, 2004) e mais dois experimentos com as bases de imagens Feret (PHILLIPS; RAUSS, 1997), *Yale Face* ³ e ORL ⁴.

¹<http://www.robots.ox.ac.uk/vgg/data3.html>

²<http://www.pascal-network.org/challenges/VOC/voc2006>

³<http://vision.ucsd.edu/content/yale-face-database>

⁴<http://www.cl.cam.ac.uk/>

(a) M-LMP - 355 correspondências corretas

(b) CS-LMP - 355 correspondências corretas

(c) CS-LBP - 349 correspondências corretas

(d) SIFT - 330 correspondências corretas

Figura 42: Correspondências corretas entre um par de imagens “east park”. Total de 361 correspondências retornadas.

Experimento-1: foram utilizadas quatro classes de imagens da base Caltech 2004 que incluem 1074 aviões, 1155 carros, 450 faces e 826 motos. A Figura 23 mostra um exemplo de uma imagem de cada categoria. Foi utilizada a Metodologia-1 para o desenvolvimento desse experimento, onde utilizou-se 200 imagens de cada classe para a construção do *codebook* (800 imagens), 200 imagens de cada classe para o conjunto de treinamento (800 imagens) e 50 imagens de cada classe para o conjunto de testes (200 imagens). Foi calculada a variação de sensibilidade para vários valores de k (Tabela 9). Com este resultado, foi definido o melhor valor de $k = 250$.

Tabela 9: O melhor tamanho de *cluster* (4 classes da base de dados Caltech 2004)

tamanho do <i>cluster</i>	sensibilidade
50	74,5%
100	76,5%
150	78,5%
200	82,5%
250	85,0%
300	80,5%
350	78,5%
400	74,0%
450	74,0%
500	75,5%

As matrizes de confusão apresentadas nas Tabelas 10, 11 and 12 mostram o número de classificações corretas de cada objeto para os três descritores comparados (M-LMP,

CS-LBP e SIFT). Os valores em negrito são as sensibilidades alcançadas pelos descritores e os valores da diagonal representam o número de Verdadeiros Positivos ou número de classificações corretas. Os outros valores nas colunas correspondentes são as classificações incorretas ou o número de Falsos Negativos em cada classe. Os erros de classificação para cada descritor testado, calculado pela Equação 29, são mostrados na Tabela 13.

Tabela 10: Matriz de Confusão para localização de objetos em quatro classes de imagens da base Caltech 2004, utilizando o descritor M-LMP

		Entradas			
		<i>motos</i>	<i>faces</i>	<i>carros</i>	<i>aviões</i>
Saídas	<i>motos</i>	49	10	8	4
	<i>faces</i>	1	39	2	0
	<i>carros</i>	0	0	38	2
	<i>aviões</i>	0	1	2	44
Sensibilidade	85,0%				

Tabela 11: Matriz de Confusão para localização de objetos em quatro classes de imagens da base Caltech 2004, utilizando o descritor CS-LBP

		Entradas			
		<i>motos</i>	<i>faces</i>	<i>carros</i>	<i>aviões</i>
Saídas	<i>motos</i>	50	10	12	10
	<i>faces</i>	0	39	4	8
	<i>carros</i>	0	1	34	3
	<i>aviões</i>	0	0	0	29
Sensibilidade	76,0%				

Tabela 12: Matriz de Confusão para localização de objetos em quatro classes de imagens da base Caltech 2004, utilizando o descritor SIFT

		Entradas			
		<i>motos</i>	<i>faces</i>	<i>carros</i>	<i>aviões</i>
Saídas	<i>motos</i>	48	12	8	11
	<i>faces</i>	1	36	1	4
	<i>carros</i>	1	0	39	4
	<i>aviões</i>	0	2	2	31
Sensibilidade	77,0%				

Os resultados mostram que o descritor proposto M-LMP é melhor do que o CS-LBP e o SIFT para localização (ou reconhecimento) de objetos em um modelo BoF. O descritor proposto supera o SIFT em 8,0% e 9,0% de sensibilidade (ou classificações corretas) sobre o CS-LBP.

Tabela 13: Taxas de erro para quatro classes de imagens da base Caltech 2004

M-LMP	CS-LBP	SIFT
15,0%	24,0%	23,0%

Uma observação interessante é que o CS-LBP tem excelente desempenho no reconhecimento de motos com nenhum Falso Positivo e o pior desempenho entre os três métodos para reconhecimento de aviões (29 de 50). O SIFT foi melhor para reconhecimento de carros (39 de 50). O descritor proposto é melhor no reconhecimento de aviões (44 de 50) e no reconhecimento de faces (39 de 50), ficando em segundo lugar em relação a outras duas classes (motos e carros). A taxa de erro foi somente de 15,0% na aplicação do descritor M-LMP para reconhecimento de objetos.

Experimento-2: o número de classes da mesma base Caltech 2004 foi aumentado para 6 incluindo-se as classes “guitarra” e “casa”. Foram selecionadas aleatoriamente 200 imagens de cada classe para a geração do *codebook* (1200 imagens), 200 imagens de cada classe para o conjunto de treinamento (1200 imagens) e 50 imagens de cada classe para o conjunto de teste (300 imagens).

O desenvolvimento do experimento foi o mesmo descrito anteriormente, ou seja, usando a Metodologia-1 descrita no Capítulo 4. Os valores usados para os parâmetros R , N , T , b e β foram também os mesmos.

Foi calculada a variação da sensibilidade com vários valores de *clusters* apresentada na Figura 43. Foi escolhido o tamanho do *cluster* como $k = 250$ como o melhor valor.

Figura 43: Variação da sensibilidade com o tamanho de *cluster* (6 classes da base de imagens Caltech 2004)

As matrizes de confusão são apresentadas nas Tabelas 14, 15 e 16. Os valores em negrito são as sensibilidades alcançadas pelos descritores e os valores da diagonal representam o número de Verdadeiros Positivos ou número de classificações corretas. Os outros valores nas colunas correspondentes são as classificações incorretas ou o número de Falsos Negativos em cada classe. Os erros de classificação para cada descritor testado, calculado

pela Equação 29, são mostrados na Tabela 17.

Tabela 14: Matriz de confusão para o descritor M-LMP na classificação de objetos (6 classes da base de imagens Caltech 2004)

		Entradas					
		motos	faces	carros	aviões	guitarra	casas
Saídas	motos	50	2	2	2	7	8
	faces	0	44	3	5	4	4
	carros	0	1	36	1	3	5
	aviões	0	0	3	36	0	6
	guitarra	0	1	3	3	35	9
	casas	0	2	3	3	1	18
Sensibilidade		73,0%					

Tabela 15: Matriz de confusão para o descritor CS-LBP na classificação de objetos (6 classes da base de imagens Caltech 2004)

		Entradas					
		motos	faces	carros	aviões	guitarra	casas
Saídas	motos	49	2	8	10	10	10
	faces	0	36	1	5	5	6
	carros	0	0	34	3	1	1
	aviões	0	0	0	29	2	4
	guitarra	0	11	5	2	32	10
	casas	1	1	2	1	0	19
Sensibilidade		66,0%					

Tabela 16: Matriz de confusão para o descritor SIFT na classificação de objetos (6 classes da base de imagens Caltech 2004)

		Entradas					
		motos	faces	carros	aviões	guitarra	casas
Saídas	motos	47	6	7	5	11	9
	faces	2	39	2	2	3	5
	carros	0	0	35	6	0	4
	aviões	0	0	3	29	2	4
	guitarra	1	3	0	1	33	7
	casas	0	2	3	7	1	21
Sensibilidade		68,0%					

Experimento-3: foi utilizada a base de dados Pascal VOC 2006 com as seguintes classes: ônibus, vaca, cavalo e ovelha. Foram selecionadas aleatoriamente 80 imagens de cada classe para o *codebook* (320 imagens), e para o conjunto de treinamento foram

Tabela 17: Taxas de erro (6 classes da base de imagens Caltech 2004)

M-LMP	CS-LBP	SIFT
27,0%	33,7%	32,0%

selecionados 558 imagens: 94 ônibus, 126 vacas, 167 cavalos, e 171 ovelhas. Para o conjunto de teste foram selecionadas 869 imagens: 180 imagens de ônibus, 197 imagens de vacas, 247 imagens de cavalos e 251 de ovelhas. A metodologia deste experimento foi a Metodologia-1 (Sub-Seção 4.2.2), usando $k = 250$ no método *k-means*. O apêndice A apresenta três tabelas com as informações de quais imagens foram utilizadas para este experimento.

O número de classificações corretas (Verdadeiros Positivos) e classificações incorretas (Falsos Negativos) são apresentados nas matrizes de confusão das Tabelas 18, 19, e 20, e os valores em negrito são as sensibilidades de cada descritor. Os erros de classificação para cada método testado, calculados pela Equação 29, são mostrados na Tabela 21.

Tabela 18: Matriz de Confusão para o descritor M-LMP na classificação de objetos (4 classes da base de imagens Pascal VOC 2006)

		Entradas			
		ônibus	vaca	cavalo	ovelha
Saídas	ônibus	140	15	71	13
	vaca	2	108	57	61
	cavalo	36	24	98	40
	ovelha	2	50	28	124
Sensibilidade	55,08%				

Tabela 19: Matriz de Confusão para o descritor CS-LBP na classificação de objetos (4 classes da base de imagens Pascal VOC 2006)

		Entradas			
		ônibus	vaca	cavalo	ovelha
Saídas	ônibus	145	9	44	12
	vaca	4	93	59	84
	cavalo	25	40	114	48
	ovelha	6	55	37	94
Sensibilidade	51,32%				

Nota-se que o desempenho dos três descritores decaem quando aumenta-se o número de classes de quatro para seis, levando em conta a base de imagens Caltech 2004. Mudando a base de imagens para a Pascal VOC 2006 os três métodos reduzem drasticamente seus desempenhos, como pode ser visto na Tabela 21. Contudo é importante notar que

Tabela 20: Matriz de Confusão para o descritor SIFT na classificação de objetos (4 classes da base de imagens Pascal VOC 2006)

		Entradas			
		ônibus	vaca	cavalo	ovelha
Saídas	ônibus	126	3	24	6
	vaca	9	108	65	77
	cavalo	43	43	132	43
	ovelha	2	43	33	112
Sensibilidade	55,00%				

Tabela 21: Taxas de erro (4 classes da base de imagens Pascal VOC 2006)

M-LMP	CS-LBP	SIFT
45,00%	49,00%	45,00%

os parâmetros da nossa abordagem foram otimizados utilizando apenas as imagens da Figura 21, ou seja, foram utilizadas imagens que não pertencem ao conjunto de imagens testadas nestes experimentos. Isto é, os métodos CS-LBP e SIFT são metodologias não paramétricas e seus desempenhos não podem ser otimizados pela “sintonização” de parâmetros. Por outro lado, o M-LMP pode ter sua performance melhorada usando um conjunto de treinamento da mesma base de imagens.

A Tabela 22 mostra a comparação da sensibilidade dos três descritores para as diferentes base de imagens utilizada.

Tabela 22: Comparação da sensibilidade dos três descritores

Experimento	Número de classes - base de imagens	CS-LBP	SIFT	M-LMP
1	4 - Caltech 2004	76,00%	77,00%	85,00%
2	6 - Caltech 2004	66,00%	68,00%	73,00%
3	4 - Pascal VOC 2006	51,32%	55,00%	55,08%

Experimento-4: foram escolhidas 4 classes da base Caltech 101: *bonsai*, *faces easy*, *leopards* e *watch*, com imagens com variação de rotação, iluminação, escala, compressão e *aliasing*. O objetivo do Experimento-4 foi a redução do tamanho do descritor variando o parâmetro b da Equação 21 do descritor M-LMP no intervalo $[1, 15]$. Os outros parâmetros utilizados foram: $\beta = 1, 2$, $R = 2$ e $N = 8$ para o M-LMP.

Os parâmetros utilizados para o método CS-LBP foram: $R = 2$, $N = 8$ e $T = 0,01$.

A Figura 44 mostra a variação da sensibilidade para valores de b no intervalo $[1, 15]$, e a Figura 45 mostra a taxa de erro variando os valores de b no mesmo intervalo. Tem-se uma maior sensibilidade, juntamente com uma menor taxa de erro para $b = 4$ (5 bins), resultando em um descritor de tamanho 80 posições ($4 \times 4 \times (4 + 1)$).

Figura 44: b versus sensibilidade para quatro classes da base de imagens Caltech 101

Figura 45: b versus taxa de erro para quatro classes da base de imagens Caltech 101

As Tabelas 23, 24 mostram as matrizes de confusão para o M-LMP com $b = 15$ e $b = 4$. Para os métodos CS-LBP e SIFT os resultados estão nas Tabelas 25 e 26. Os valores em negrito são as sensibilidades alcançadas pelos descritores. Os erros de classificação para cada descritor são mostrados na Tabela 27.

Tabela 23: Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor M-LMP com $b = 15$

		Entradas			
		<i>leopards</i>	<i>faces</i>	<i>watch</i>	<i>bonsai</i>
Saídas	<i>leopards</i>	22	0	1	0
	<i>faces</i>	0	30	2	0
	<i>watch</i>	0	0	21	0
	<i>bonsai</i>	8	0	6	30
Sensibilidade		85,84%			

Os resultados mostram que mesmo com a redução do tamanho do descritor, o método

Tabela 24: Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor M-LMP com $b = 4$

		Entradas			
		<i>leopards</i>	<i>faces</i>	<i>watch</i>	<i>bonsai</i>
Saídas	<i>leopards</i>	24	0	0	0
	<i>faces</i>	0	30	2	0
	<i>watch</i>	0	0	22	1
	<i>bonsai</i>	6	0	6	29
Sensibilidade		87,5%			

Tabela 25: Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor CS-LBP

		Entradas			
		<i>leopards</i>	<i>faces</i>	<i>watch</i>	<i>bonsai</i>
Saídas	<i>leopards</i>	14	0	0	0
	<i>faces</i>	1	30	3	0
	<i>watch</i>	1	0	21	1
	<i>bonsai</i>	14	0	6	29
Sensibilidade		78,34%			

Tabela 26: Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 4 classes, utilizando o descritor SIFT

		Entradas			
		<i>leopards</i>	<i>faces</i>	<i>watch</i>	<i>bonsai</i>
Saídas	<i>leopards</i>	17	0	0	1
	<i>faces</i>	0	30	5	0
	<i>watch</i>	1	0	21	0
	<i>bonsai</i>	12	0	4	29
Sensibilidade		80,84%			

Tabela 27: Taxas de erro (4 classes do banco de imagens Caltech 101)

M-LMP $b = 15$	M-LMP $b = 4$	CS-LBP	SIFT
14,16%	12,5%	21,67%	19,17%

M-LMP apresenta bom desempenho na classificação de imagens, com uma sensibilidade maior para $b = 4$ (tamanho do descritor 80 posições) do que para $b = 15$ (tamanho do descritor 256 posições). Comparando o descritor M-LMP com $b = 4$ com o descritor M-LMP com $b = 15$ nota-se que o descritor M-LMP com $b = 4$ possui um número maior de acertos para *leopards* e *watch*, empata na classe *faces* e perde para a classe *bonsai* por apenas uma classificação incorreta. Os resultados do método M-LMP para $b = 4$ superam

os métodos SIFT e CS-LBP, apresentando uma sensibilidade maior e uma taxa de erro de apenas 12,5%.

As Tabelas 28 e 29 mostram o tempo de processamento para a construção do dicionário de palavras visuais. Para o método M-LMP verificou-se o tempo de processamento para cada valor de b . Nota-se que conforme o parâmetro b aumenta, o tempo para a construção do vocabulário visual também aumenta. Fixando $b = 4$ no descritor M-LMP tem-se um tempo de processamento menor comparado aos métodos CS-LBP e SIFT.

Tabela 28: Tempo de processamento em segundos para a construção do vocabulário visual utilizando o descritor M-LMP variando o parâmetro b

b	M-LMP
1	45,64s
2	56,11s
3	67,79s
4	78,01s
5	91,61s
6	101,47s
7	113,88s
8	127,53s
9	139,05s
10	150,69s
11	161,71s
12	173,98s
13	185,81s
14	204,20s
15	212,62s

Tabela 29: Tempo de processamento em segundos para a construção do vocabulário visual utilizando os descritores CS-LBP e SIFT

CS-LBP	SIFT
212,94s	117,67s

Experimento-5: foi analisado o desempenho do descritor M-LMP utilizando a Metodologia-1 para a classificação de objetos de 5 classes da base Caltech 101: *bonsai*, *faces_easy*, *leopards*, *watch* e *airplanes*. O descritor M-LMP é utilizado com o parâmetro $b = 4$ (escolhido pelo Experimento-4), e comparado com os descritores CS-LBP e SIFT. Os outros parâmetros utilizados foram idênticos ao do Experimento-4.

Foram utilizadas 98 imagens de cada classe para o agrupamento e treinamento e 30 imagens de cada classe para o conjunto de teste. Na fase de treinamento as regiões de interesse para cada imagem são obtidas utilizando o detector de regiões de interesse *Hessian-Affine* (MIKOLAJCZYK; SCHMID, 2004). Para cada região detectada são construídos os descritores M-LMP, CS-LBP e SIFT. Os descritores foram agrupados usando o

método *k-means* (MACQUEEN, 1967) com $k = 250$. Assim, são gerados 250 protótipos, que são os centro dos *clusters*.

Os resultados estão na matrizes de confusão apresentadas nas Tabelas 30, 31 e 32 para os três descritores comparados M-LMP, CS-LBP e SIFT respectivamente.

Tabela 30: Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 5 classes, utilizando o descritor M-LMP com $b = 4$

		Entradas				
		<i>leopards</i>	<i>faces_easy</i>	<i>watch</i>	<i>bonsai</i>	<i>airplanes</i>
Saídas	<i>leopards</i>	21	0	0	0	0
	<i>faces_easy</i>	0	30	1	1	6
	<i>watch</i>	1	0	20	0	0
	<i>bonsai</i>	8	0	7	29	7
	<i>airplanes</i>	0	0	2	0	17
Sensibilidade		78,00%				

Tabela 31: Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 5 classes, utilizando o descritor CS-LBP

		Entradas				
		<i>leopards</i>	<i>faces_easy</i>	<i>watch</i>	<i>bonsai</i>	<i>airplanes</i>
Saídas	<i>leopards</i>	14	0	0	0	0
	<i>faces_easy</i>	0	25	6	1	2
	<i>watch</i>	0	3	10	0	1
	<i>bonsai</i>	16	2	11	29	4
	<i>airplanes</i>	0	0	3	0	23
Sensibilidade		67,34%				

Tabela 32: Matriz de Confusão para classificação de objetos em imagens da base Caltech 101 para 5 classes, utilizando o descritor SIFT

		Entradas				
		<i>leopards</i>	<i>faces_easy</i>	<i>watch</i>	<i>bonsai</i>	<i>airplanes</i>
Saídas	<i>leopards</i>	13	0	0	0	0
	<i>faces_easy</i>	0	29	7	0	4
	<i>watch</i>	0	1	20	0	5
	<i>bonsai</i>	17	0	1	30	2
	<i>airplanes</i>	0	0	2	0	19
Sensibilidade		74,00%				

Os erros de classificação para cada descritor testado são mostrados na Tabela 33.

Tabela 33: Taxas de erro (5 classes do banco de imagens Caltech 101)

M-LMP	CS-LBP	SIFT
22,00%	32,67%	27,00%

Nota-se que o desempenho dos três descritores decaem quando aumenta-se o número de classes de quatro para cinco. Uma observação interessante é que o SIFT teve melhor

desempenho no reconhecimento de *bonsai* com nenhum falso negativo e o pior desempenho entre os três métodos no reconhecimento de *leopards* (13 de 30). O CS-LBP foi melhor no reconhecimento de *airplanes* (23 de 30). O descritor M-LMP é melhor no reconhecimento de *leopards* (21 de 30) e no reconhecimento de faces (30 de 30), empatando com o SIFT no reconhecimento de *watch*. Considerando-se as cinco classes, a sensibilidade global é maior do que a dos dois outros descritores.

Para os próximos experimentos foi utilizada a Metodologia-2 (Sub-Seção 4.2.2).

Experimento-6: foi avaliada a metodologia no reconhecimento de faces humanas, comparando o descritor M-LMP com os descritores SIFT e LIOP. Foi utilizada a base de faces Feret, onde foram escolhidas 45 classes contendo 40 imagens por classe. Subdividiu-se randomicamente as imagens da base de imagens Feret em 25 imagens para treinamento e 15 imagens para teste.

Primeiramente foram extraídas regiões de interesse nas imagens pelo detector de regiões *Hessian-Affine*. Cada região de interesse de tamanho 40×40 pixels é dividida em uma malha de tamanho 4×4 células, e as características são extraídas para cada região de interesse utilizando o descritor M-LMP. Tais descritores foram clusterizados utilizando a metodologia *Elkan k-means* para a clusterização. Os parâmetros utilizados para o descritor M-LMP foram: $R = 2$, $N = 8$ e $b = 4$. Para $b = 4$ há uma redução do tamanho do descritor para 80 posições.

O objetivo do Experimento-6 foi comparar o descritor M-LMP com outros três descritores já existentes na literatura e, além disso, comparar a metodologia proposta com outras metodologias na tarefa de reconhecimento de faces.

Neste experimento fixou-se o tamanho do *cluster* em $k = 450$, e ajustou-se o parâmetro β para o descritor M-LMP de acordo com sua sensibilidade. Os outros parâmetros foram configurados experimentalmente como $b = 4$, $R = 2$ e $N = 8$. A Tabela 34 mostra a variação do parâmetro β com sua respectiva sensibilidade.

O valor em negrito da Tabela 34 refere-se ao valor que será utilizado nos experimentos utilizando a base de imagens Feret. Com esse valor comparou-se o descritor M-LMP utilizando a metodologia BoF com outros dois descritores: SIFT e LIOP, variando o tamanho do *cluster* k (Tabela 35).

Observando a Tabela 35, a melhor sensibilidade ($S = 95,41\%$) foi alcançada para o descritor M-LMP quando $k = 450$, enquanto para o SIFT um número maior de *clusters* são necessários para se obter uma melhor sensibilidade ($S = 95,56\%$ para $k = 500$). Para o descritor LIOP a melhor sensibilidade também foi alcançada para $k = 550$ ($S = 92,00\%$).

Experimento-7: foi comparada a metodologia proposta nesta tese para o reconhecimento de faces humanas com outras quinze metodologias publicadas na literatura usando as seguintes base de imagens: Feret, ORL e *Yale Face*.

Inicialmente, foi utilizada a base de imagens Feret para comparar nossa metodologia com a metodologia *Face Fuzzy* (MENG; ZHAO; YUE, 2008), usando o teste de validação

Tabela 34: Performance do descritor M-LMP durante a sintonização do parâmetro β para 40 imagens por classes (base de imagens Feret)

β	Sensibilidade
1×10^{-10}	92,15%
1×10^{-6}	92,44%
1×10^{-5}	92,30%
1×10^{-4}	93,33%
1×10^{-3}	93,33%
1×10^{-2}	93,19%
2×10^{-4}	93,04%
3×10^{-4}	93,19%
4×10^{-4}	91,70%
5×10^{-4}	92,30%
6×10^{-4}	92,15%
7×10^{-4}	92,59%
8×10^{-4}	92,74%
9×10^{-12}	92,15%
9×10^{-4}	92,74%
21×10^{-5}	93,63%
22×10^{-5}	92,44%
23×10^{-5}	92,00%
34×10^{-3}	91,11%
71×10^{-5}	93,78%
72×10^{-5}	92,59%
73×10^{-5}	93,33%
89×10^{-5}	93,33%
91×10^{-5}	93,19%
92×10^{-5}	93,04%
93×10^{-5}	93,19%
94×10^{-5}	92,30%
95×10^{-5}	92,89%
96×10^{-5}	94,07%
111×10^{-6}	93,33%
795×10^{-5}	92,59%
899×10^{-6}	93,78%
911×10^{-6}	92,00%
951×10^{-6}	93,63%
952×10^{-6}	93,48%
953×10^{-6}	92,74%
954×10^{-6}	92,89%
955×10^{-6}	94,22%
956×10^{-6}	92,89%
989×10^{-8}	92,15%
998×10^{-6}	95,41%
999×10^{-6}	93,78%
9531×10^{-7}	92,15%
9532×10^{-7}	94,07%
9533×10^{-7}	92,74%
9981×10^{-7}	93,63%
9982×10^{-7}	93,48%
9999×10^{-7}	92,74%
99999×10^{-8}	92,44%

Tabela 35: Desempenho dos métodos comparados (sensibilidade) para 45 classes da base de imagens Feret variando-se o tamanho do *cluster*

Tamanho do <i>cluster</i>	M-LMP	LIOP	SIFT
k=250	92,15%	91,26%	92,89%
k=250	92,89%	90,96%	93,63%
k=300	92,74%	91,11%	94,07%
k=350	92,30%	90,81%	95,11%
k=400	93,93%	90,81%	94,81%
k=450	95,41%	90,52%	94,96%
k=500	93,78%	91,85%	95,56%
k=550	93,19%	92,00%	95,11%
k=600	93,19%	91,85%	94,37%

cruzada para 2, 5 e 10-*fold* dividido em 4×4 sub-imagens. Para este experimento também foram utilizadas todas as 40 imagens das 45 classes e subdivididas em 25 imagens para treinamento e 15 imagens para teste. Os resultados são apresentados na Tabela 36.

Tabela 36: Comparação da metodologia proposta com a metodologia *Face Fuzzy* para a base de imagens Feret

BoF (M-LMP) $\beta = 0.000998$	Face Fuzzy	
95,41%	validação cruzada 2-fold	63,33%
	validação cruzada 5-fold	70,55%
	validação cruzada 10-fold	71,39%

Observando a Tabela 36, os resultados mostram que a metodologia proposta ultrapassa a metodologia *Face Fuzzy* em 24,02% de sensibilidade.

Utilizando as bases de imagens ORL e *Yale Face* foi comparada a metodologia proposta nesta tese com outras catorze metodologias: OLPP2+SVM2, OLPP2+SVM1, OLPP2+NN, OLPP1+SVM2, OLPP2+SVM1, OLPP1+NN (CAI et al., 2006; SOLDERA; BEHAINE; SCHARCANSKI, 2015), *Enhanced ASM* (BEHAINE; SCHARCANSKI, 2012), SVM2 (CHANG; LIN, 2011; FAN; CHEN; LIN, 2005), SVM1 (CORTES; VAPNIK, 1995), SRKDA (SCHÖLKOPF; SMOLA, 2001), MvDA (KAN et al., 2015), *Eigenfaces* (TURK; PENTLAND, 1991), *Fisherfaces* (BELHUMEUR; HESPANHA; KRIEGMAN, 1997) e *Laplacianfaces* (HE et al., 2005). Foram utilizadas as 10 imagens das 40 classes e subdivididas em 9 imagens para treinamento e 1 imagem para teste.

Para a base de faces ORL comparou-se a metodologia proposta com mais catorze metodologias descritas na Tabela 37. A sensibilidade dos catorze métodos comparados com a metodologia proposta foram retirados do trabalho de Soldera et al. (2015). Neste experimento, o valor de β e k não foram previamente sintonizados, mas empiricamente escolhidos baseado nos outros experimentos.

Pode-se observar na Tabela 37 que a metodologia proposta usando BoF e o descritor M-LMP superou todas as outras metodologias, mesmo não ocorrendo a sintonização do

Tabela 37: Comparação da metodologia proposta com outras metodologias usando a base de imagens ORL

Métodos	Sensibilidade
BoF (M-LMP) $\beta = 0,0000001$ $k = 350$	97,50%
OLPP2+SVM2	94,00%
OLPP2+SVM1	91,50%
OLPP2+NN	93,50%
OLPP1+SVM2	93,70%
OLPP1+SVM1	84,60%
OLPP1+NN	93,00%
Enhanced ASM	10,50%
SVM2	93,50%
SVM1	86,30%
SRKDA	89,00%
MvDA	78,10%
Eigenfaces	93,00%
Fisherfaces	91,70%
Laplacianfaces	92,70%

parâmetro β .

A base de imagens *Yale Face* contém 15 classes com 11 imagens por classe. As configurações das faces em cada classe são: com iluminação centralizada, com óculos, sem óculos, feliz, com iluminação a esquerda, normal, com iluminação a direita, triste, sonolento, surpreso e piscando os olhos.

Utilizando a base de imagens *Yale Face*, comparou-se a metodologia proposta com as mesmas catorze metodologias com os resultados apresentados na Tabela 38. A sensibilidade dos catorze métodos comparados com a metodologia proposta foram retirados do trabalho de Soldera et al. (2015)(SOLDERA; BEHAINE; SCHARCANSKI, 2015). Também neste experimento, o valor de β e k foram empiricamente definidos baseado em experimentos anteriores.

Pode-se observar na Tabela 38 que a metodologia BoF usando o descritor M-LMP obteve uma sensibilidade de 100% e superou todas as outras metodologias, mesmo não ocorrendo a sintonia do parâmetro β . Observou-se também que não foi necessário a construção do vocabulário para um valor de k grande para a obtenção da sensibilidade máxima ocorrida.

5.3.2 Resultados obtidos com os descritores M-LMP e CS-LMP

Esta Sub-seção relata os resultados obtidos dos experimentos realizados com duas bases de imagens: Caltech 101 e ImageNet. O objetivo dos experimentos foi a redução do tamanho do descritor, onde os resultados são obtidos por uma variante da metodologia originalmente proposta baseada em BoF, mas utilizando histograma espacial e SVM. Os

Tabela 38: Comparação da metodologia proposta com outras metodologias usando a base de imagens *Yale Face*

Métodos	Sensibilidade
BoF (M-LMP) $\beta = 0.00000001$ $k = 200$	100%
OLPP2+SVM2	98,8%
OLPP2+SVM1	95,2%
OLPP2+NN	98,2%
OLPP1+SVM2	87,2%
OLPP1+SVM1	78,2%
OLPP1+NN	78,2%
Enhanced ASM	7,2%
SVM2	89,1%
SVM1	81,8%
SRKDA	98,2%
MvDA	71,7%
Eigenfaces	77,6%
Fisherfaces	98,1%
Laplacianfaces	77,7%

resultados são apresentados como uma avaliação comparativa do desempenho dos descritores M-LMP, CS-LMP, SIFT e LIOP na tarefa de classificação de objetos. Diferentemente da maioria dos trabalhos publicados nesta área, principalmente aqueles submetidos ao *ILSVRC-ImageNet Large Scale Visual Recognition Challenge* (RUSSAKOVSKY et al., 2015) as imagens utilizadas não são recortadas e nenhum pré-processamento é realizado. Além disso, para o desafio ILSVRC, eles calculam o erro de classificação utilizando a busca dos cinco elementos mais semelhantes (*Top-5*). Nos experimentos realizados neste trabalho foi utilizada a busca em *Top-1* que procura pela imagem mais similar na base de consulta, que deixa o tempo de execução do algoritmo mais rápido.

Para os experimentos realizados nesta Sub-seção, foram extraídas regiões de interesse utilizando duas abordagens: *Hessian Multiscale* (MIKOLAJCZYK; SCHMID, 2002) e *dense sampling* (ISCEN et al., 2015). *Hessian Multiscale* usa o determinante da matriz Hessiana em multi-escala para detectar características em múltiplas escalas. No detector *dense sampling*, o centro da região de interesse é estabelecido fazendo um deslocamento a cada 2 *pixels* e assim é construído uma malha regular de tamanho e forma fixa. O descritor é construído para cada região de interesse de tamanho 40×40 *pixels*. Alguns descritores usam pontos-chave multi-escala para detectar as regiões de interesse (LOWE, 2004), mas alguns trabalhos mostram que *dense sampling* ou *random sampling* superam os métodos de pontos-chave (ISCEN et al., 2015; NOWAK; JURIE; TRIGGS, 2006b).

O tamanho do descritor original é de 256 posições ($b=15$ na Equação 21). Para reduzir o tamanho do descritor, foram testados outros três tamanhos de descritores: 64 elementos ($b=3$), 80 elementos ($b=4$) e 96 elementos ($b=5$) e os outros parâmetros estabelecidos

foram $R = 2$ e $N = 8$. O parâmetro β foi empiricamente estabelecido.

Os descritores foram agrupados usando *Elkan k-means* (ELKAN, 2003). Foram selecionadas aleatoriamente 30 imagens para a geração do dicionário. O número de *clusters* foi variado entre 200 e 500.

Experimento-8: foram utilizadas 80 imagens de cada uma das 27 classes subdivididas aleatoriamente em 50 imagens de treinamento e 30 imagens de teste. O objetivo deste experimento foi escolher o tamanho do descritor ajustando o valor de β de tal forma que a maior sensibilidade seja encontrada.

Foi fixado o tamanho do cluster para $k = 400$ e os tamanhos dos descritores para 64, 80 e 96 elementos. A Tabela 39 apresenta os resultados da sensibilidade *versus* β para cada tamanho de descritor. O parâmetro b (número de *bins* do histograma) foi variado na Equação 21 para 64 elementos ($b = 3$), 80 elementos ($b = 4$) e 96 elementos ($b = 5$).

Tabela 39: Desempenho do descritor M-LMP durante o ajuste do parâmetro β para diferentes tamanhos de descritores (Caltech 101)

	64 elementos	80 elementos	96 elementos
β	Sensibilidade (S)	Sensibilidade (S)	Sensibilidade (S)
0,0007	77,28%	76,79%	75,56%
0,0008	77,16%	78,15%	75,68%
0,0009	77,04%	76,42%	75,93%
0,0010	75,93%	77,90%	75,68%
0,0011	76,05%	76,30%	76,91%
0,0012	76,91%	77,53%	77,41%
0,0013	77,04%	78,02%	76,91%

Como pode-se observar na Tabela 39, a melhor performance foi alcançada em $\beta = 0,0008$ com uma sensibilidade de $S = 78,15\%$. Sendo assim, foi escolhido o tamanho do descritor de 80 elementos como o mais discriminante para os outros experimentos.

Experimento-9: usou o tamanho do descritor obtido no Experimento-8 e diminuindo o número de imagens por classe, foi recalculado o valor do parâmetro β usando 30 imagens por classe (20 imagens para treinamento e 10 imagens para teste) e 40 imagens por classe (30 imagens de treinamento e 10 imagens de teste). O objetivo deste experimento foi verificar se todas as imagens devem ser usadas para o ajuste do parâmetro β .

A Tabela 40 mostra o parâmetro β versus sensibilidade usando 30 imagens por classe e 40 imagens por classe da base de imagens Caltech 101.

Analizando a Tabela 40, conforme o número de imagens por classe diminui, a sensibilidade também diminui. A sensibilidade máxima usando 30 imagens por classe foi 70,74% para $\beta = 0,0012$, e para 40 imagens por classe a sensibilidade máxima foi $S = 77,04\%$ com $\beta = 0,0007$. Portanto, para ajustar devidamente o parâmetro β para a mais alta sensibilidade, é necessário usar todas as imagens para cada classe.

Tabela 40: Desempenho do descritor M-LMP durante o ajuste do parâmetro β variando o número de imagens por classe (Caltech 101)

	30 imagens	40 imagens
β	Sensibilidade (S)	Sensibilidade (S)
0,0007	68,52%	77,04%
0,0008	69,26%	74,81%
0,0009	68,52%	73,33%
0,0010	69,63%	74,07%
0,0011	67,78%	72,96%
0,0012	70,74%	74,44%
0,0013	67,78%	75,56%

Experimento-10: baseando-se nos resultados do Experimento-8 e Experimento-9 utilizou-se o detector de regiões de interesse *dense sampling* para a base Caltech 101, e comparou-se o descritor M-LMP com os descritores CS-LMP, SIFT e LIOP. Neste experimento foi utilizada a Metodologia-2.

O descritor CS-LMP usa a mesma curva sigmóide do descritor M-LMP, e portanto, para achar o melhor valor para o parâmetro β , foi ajustado o tamanho do cluster para $k = 400$, o número de elementos do descritor para 80 ($b = 4$), e variou-se o parâmetro β para obter a mais alta sensibilidade. A Tabela 41 mostra o parâmetro β versus sensibilidade levando em consideração todas as 2160 imagens (80 imagens $\times$ 27 classes). A melhor sensibilidade foi alcançada com $S = 77,65\%$ e $\beta = 0,0012$.

Tabela 41: Desempenho do descritor CS-LMP durante o ajuste do parâmetro β para 80 imagens por classe (Caltech 101)

β	Sensibilidade (S)
0,0007	76,91%
0,0008	75,93%
0,0009	76,17%
0,0010	76,79%
0,0011	76,17%
0,0012	77,65%
0,0013	77,28%

Para comparar o descritor M-LMP com os descritores CS-LMP, LIOP e SIFT, foram consideradas as 80 imagens para cada uma das 27 classes baseado nos resultados do Experimento-8 e Experimento-9, variando o tamanho do cluster k . Foi mantido o tamanho dos descritores M-LMP e CS-LMP em 80 elementos e foram usados os seguintes parâmetros: $\beta = 0,0008$ (M-LMP) e $\beta = 0,0012$ (CS-LMP). Para cada descritor e para cada classe, foram utilizadas 50 imagens de treinamento (geração do vocabulário) e 30 imagens para teste (classificação). Os resultados são apresentados na Tabela 42.

Tabela 42: Sensibilidade (27 classes da base de imagens Caltech 101) usando *dense sampling*

Tamanho do <i>cluster</i>	M-LMP 80 elementos	SIFT 128 elementos	LIOP 144 elementos	CS-LMP 80 elementos
$k = 200$	76,91%	76,54%	67,41%	75,56%
$k = 250$	76,17%	76,79%	67,41%	76,91%
$k = 300$	77,53%	77,16%	67,90%	77,28%
$k = 350$	76,79%	77,41%	67,41%	75,56%
$k = 400$	78,15%	77,28%	68,52%	77,65%
$k = 450$	77,53%	78,64%	68,15%	77,04%
$k = 500$	77,78%	77,90%	67,53%	77,65%

Com a análise da Tabela 42, a melhor sensibilidade para o descritor M-LMP ($S = 78,15\%$) foi alcançada quando $k = 400$, enquanto para o SIFT, foi necessário um número maior de *clusters* para obter uma melhor sensibilidade ($S = 78,64\%$ para $k = 450$). Para o descritor LIOP, a melhor sensibilidade ($S = 68,52\%$) foi alcançada também quando $k = 400$. Para o descritor CS-LMP a melhor sensibilidade foi $S = 77,65\%$ para os tamanhos de cluster $k = 400$ e $k = 500$. Neste caso foi considerado o menor tamanho de *cluster* ($k = 400$) como o melhor resultado.

As Figuras 46 a 49 mostram as matrizes de confusão com as corretas classificações colocadas na diagonal, para cada uma das 30 imagens analisadas das 27 classes da base de dados Caltech 101. As linhas referem-se às classes de entrada e as colunas são as classes atribuídas.

Figura 46: Matriz de Confusão para o descritor M-LMP usando *dense sampling* (Caltech 101)

Analisando os resultados das matrizes de confusão das Figuras 46 até 49, observa-se

Figura 47: Matriz de confusão para o descritor SIFT usando *dense sampling* (Caltech 101)

Figura 48: Matriz de confusão para o descritor LIOP usando *dense sampling* (Caltech 101)

que o descritor M-LMP teve um excelente desempenho (definido como 30 acertos, ou 100% de classificações corretas) em 5 classes: *faces*, *leopards*, *airplanes*, *car_side* e *tri-*

Figura 49: Matriz de confusão para o descritor CS-LMP usando *dense sampling* (Caltech 101)

lobite (veja Figura 50), enquanto LIOP alcançou 30 acertos somente em 3 classes (veja Figura 51): *faces*, *leopards* e *trilobite*. O descritor M-LMP proposto teve um baixo desempenho (abaixo de 50% de classificações incorretas) nas classes *background_google* (13 acertos), *butterfly* (13 acertos), *menorah* (14 acertos) e *scorpion* (12 acertos). SIFT teve um baixo desempenho nas classes *background_google* (10 acertos), *butterfly* (14 acertos) e *scorpion* (14 acertos), e o descritor LIOP teve um baixo desempenho nas classes *background_google* (10 acertos), *butterfly* (11 acertos), *chandelier* (6 acertos), *helicopter* (12 acertos), *menorah* (13 acertos) e *scorpion* (3 acertos). O descritor proposto CS-LMP teve excelente desempenho (30 acertos) em 6 classes: *faces*, *faces_easy*, *leopards*, *airplanes*, *car_side* e *trilobite* (veja Figura 52), e o pior desempenho nas classes *background_google* (11 acertos), *butterfly* (14 acertos), *menorah* (12 acertos) e *scorpion* (12 acertos). O descritor SIFT teve um excelente desempenho (30 acertos) em 5 classes: *faces*, *faces_easy*, *leopards*, *ketch* e *trilobite* (veja Figura 53), mas superou M-LMP em somente 9 das 27 classes. É importante notar que a classe *background_google* tem muitas imagens diferentes umas das outras; consequentemente, todos os descritores comparados obtiveram um desempenho ruim (veja Figura 54). M-LMP e CS-LMP confundiram muitas imagens da classe *ibis* com a classe *kangaroo* (veja Figura 55), enquanto SIFT e LIOP confundiram muitas imagens da classe *scorpion* com a classe *kangaroo* (veja Figura 56).

A Tabela 43 sumariza o pior desempenho dos quatro descritores comparados (M-LMP, SIFT, LIOP and CS-LMP). Pode-se observar que as classes *background_google*, *butterfly*

Figura 50: Exemplos de imagens de 5 classes mostrando a melhor performance do descritor M-LMP na base de imagens Caltech 101 (30 acertos = 100% classificações corretas)

Figura 51: Exemplos de imagens de 5 classes mostrando a melhor performance do descritor LIOP na base de imagens Caltech 101 (30 acertos = 100% de classificações corretas)

Figura 52: Exemplos de imagens de 6 classes mostrando a melhor performance do descritor CS-LMP na base de imagens Caltech 101 (30 acertos = 100% classificações corretas)

e *scorpion* são classes de difícil reconhecimento devido à maneira como se apresentam. Normalmente são imagens com bastante variações intra-classe. Mesmo assim, o número de acertos dos descritores propostos é ligeiramente superior ao dos descritores comparados, dentro dessas classes.

A Tabela 44 mostra o tempo de processamento requerido para a construção do vocabulário visual (50 imagens de treinamento) usando os descritores: M-LMP ($k = 400$), SIFT ($k = 450$), LIOP ($k = 400$) e CS-LMP ($k = 400$). Conforme o tamanho do descritor

Figura 53: Exemplo de imagens de 5 classes mostrando a melhor performance do descritor SIFT na base de imagens Caltech 101 (30 acertos = 100% classificações corretas)

Figura 54: A pior performance para todos os descritores: exemplos da classe *background_google* (Caltech 101)

Figura 55: Classes com um número grande de imagens classificadas incorretamente pelos descritores M-LMP e CS-LMP para a base de imagens Caltech 101 (confusões interclasses)

Figura 56: Classes com um grande número de imagens classificadas incorretamente pelos descritores LIOP e SIFT para a base de imagens Caltech 101 (confusões interclasses)

Tabela 43: O pior desempenho (*hit-rate*) dos métodos comparados para a base de imagens Caltech 101 (*dense sampling*)

	M-LMP	SIFT	LIOP	CS-LMP
<i>background_google</i>	43,34% (13 acertos)	33,34% (10 acertos)	33,34% (10 acertos)	36,67% (11 acertos)
<i>butterfly</i>	43,34% (13 acertos)	46,67% (14 acertos)	36,67% (11 acertos)	46,67% (14 acertos)
<i>menorah</i>	46,67% (14 acertos)	-	43,34% (13 acertos)	40,00% (12 acertos)
<i>scorpion</i>	40,00% (12 acertos)	46,67% (14 acertos)	10,00% (3 acertos)	40,00% (12 acertos)
<i>helicopter</i>	-	-	40,00% (12 acertos)	-
<i>chandelier</i>	-	-	20,00% (6 acertos)	-

aumenta, o tempo de processamento aumenta. Como os descritores M-LMP e CS-LMP tem os menores tamanhos e os menores valores de k , o tempo de processamento são os menores. O computador usado para o teste foi configurado com um Intel (R) Core (TM)

i7-4770k CPU com 8GB de RAM rodando o sistema operacional Windows 8.0.

Tabela 44: Tempo de processamento (em segundos) para a construção do vocabulário visual para os descritores M-LMP, SIFT, LIOP e CS-LMP usando *dense sampling* como detector de regiões de interesse

M-LMP: 80 elementos and $k = 400$	SIFT: 128 elementos and $k = 450$	LIOP: 144 elementos and $k = 400$	CS-LMP: 80 elementos and $k = 400$
4,71s	6,18s	6,31s	4,80s

A Tabela 45 mostra o tempo de processamento para a tarefa de classificação usando os descritores: M-LMP ($k = 400$), SIFT ($k = 450$), LIOP ($k = 400$) e CS-LMP ($k = 400$).

Tabela 45: Tempo de processamento (em segundos) para a tarefa de classificação dos descritores M-LMP, SIFT, LIOP and CS-LMP usando *dense sampling* como detector de regiões de interesse

M-LMP 80 elementos ($k = 400$)	SIFT 128 elementos ($k = 450$)	LIOP 144 elementos ($k=400$) ($k = 400$)	CS-LMP 80 elementos ($k = 400$)
0,0406s	0,0464s	0,0414s	0,0417s

Experimento-11: foi realizada a mesma comparação do Experimento-10, porém utilizando o detector de região *Hessian Multiscale* ao invés de *dense sampling*. A base de imagens Caltech 101 foi mantida.

Para o ajuste do parâmetro β dos descritores M-LMP e CS-LMP foi fixado o tamanho do *cluster* em $k = 400$. Os resultados são apresentados na Tabela 46.

Tabela 46: Desempenho dos descritores M-LMP e CS-LMP durante o ajuste do parâmetro β usando *Hessian Multiscale* e um descritor de 80 elementos (Caltech 101)

	M-LMP	CS-LMP
β	Sensibilidade (S)	Sensibilidade (S)
0,0007	67,04%	65,93%
0,0008	67,16%	67,53%
0,0009	76,54%	67,41%
0,0010	66,79%	68,02%
0,0011	67,53%	66,91%
0,0012	67,78%	67,04%
0,0013	67,41%	66,42%

Analisando a Tabela 46, observa-se que o melhor desempenho do descritor proposto M-LMP ocorreu em $S=76,54\%$ com $\beta = 0,0009$. Para o descritor proposto CS-LMP o melhor desempenho ocorreu em $S = 68,02\%$ com $\beta = 0,0010$.

Após o ajuste do parâmetro β , foram utilizadas 50 imagens para treinamento (geração do vocabulário) e 30 imagens para teste (classificação) para cada classe. Foi comparado o descritor M-LMP com os descritores SIFT, LIOP e CS-LMP, variando o tamanho do *cluster* de 200 até 600. Os resultados são mostrados na Tabela 47. Os valores em negrito

indicam a melhor sensibilidade pelos métodos. O descritor M-LMP alcançou a melhor sensibilidade ($S = 77,78\%$ para $k = 500$). O descritor SIFT obteve $S = 72,59\%$ de sensibilidade para $k = 550$, o descritor LIOP atingiu $S=63,46\%$ de acertos para $k = 350$, e o descritor CS-LMP obteve $S= 68,15\%$ de sensibilidade em $k = 600$.

Tabela 47: Desempenho do método (taxa de acerto) usando o detector *Hessian Multiscale* (para 27 classes de imagens da base de dados Caltech 101)

Tamanho do <i>cluster</i>	M-LMP: 80 elementos	SIFT: 128 elementos	LIOP: 144 elementos	CS-LMP: 80 elementos
$k = 200$	75,80%	71,98%	62,47%	67,04%
$k = 250$	76,54%	72,47%	63,33%	65,31%
$k = 300$	77,53%	72,35%	62,84%	66,05%
$k = 350$	77,41%	71,36%	63,46%	67,90%
$k = 400$	76,54%	70,62%	61,98%	68,02%
$k = 450$	76,91%	70,86%	62,59%	66,30%
$k = 500$	77,78%	70,74%	61,48%	67,28%
$k = 550$	77,16%	72,59%	63,21%	67,04%
$k = 600$	77,28%	71,85%	61,48%	68,15%

Nas Figuras 57 até 60 são mostradas as matrizes de confusão para as 27 classes de imagens do banco de dados Caltech 101 (conjunto de teste = 30 elementos por classe), onde as linhas se referem-se às classes de entrada e as colunas às classes atribuídas.

Com a análise das Figuras 57 até 60 conclui-se que o M-LMP obteve melhor desempenho (30 acertos=100% classificações corretas) em 5 classes: *faces*, *faces_easy*, *leopards*, *airplanes* e *trilobite*, enquanto o descritor SIFT proporcionou o reconhecimento correto de 30 imagens (excelente desempenho) em somente 2 classes (*faces_easy* e *airplanes*), e o descritor LIOP não obteve 30 acertos em nenhuma classe. O descritor CS-LMP proporcionou 30 acertos em somente uma classe: *airplane*. O descritor M-LMP teve um desempenho baixo (abaixo de 50% de classificações corretas) nas classes *background_google* (12 acertos), *butterfly* (13 acertos), *menorah* (14 acertos) e *scorpion* (12 acertos). Como pode ser notado na Figura 31 do Capítulo 4, a classe *background_google* é a classe com mais dissimilaridade intraclasses, sendo difícil, mesmo para seres humanos, classificar este tipo de imagem. Como a metodologia proposta é baseada nas “características globais”, todos os descritores comparados tiveram seus desempenhos diminuídos em classes de alta dissimilaridade de objetos e fundos na mesma classe, tal como as classes *background_google* e *butterfly*. Outra característica que diminui o desempenho para todos os métodos é o tamanho do “objeto principal”. Conforme descrito por Hoiem et al. (HOIEM; CHOD-PATHUMWAN; DAI, 2012), pequenos objetos tendem a ser mais difíceis de reconhecer e localizar. O descritor SIFT teve um desempenho baixo nas classes *background_google* (9 acertos), *butterfly* (6 acertos) e *ibis* (14 acertos). O descritor LIOP teve um desempenho baixo nas classes *background_google* (9 acertos), *buddha* (13 acertos), *butterfly* (4 acertos),

Figura 57: Matriz de confusão para o descritor M-LMP usando *Hessian Multiscale* como detector de regiões de interesse (Caltech 101)

chandelier (13 acertos), *helicopter* (13 acertos), *ibis* (13 acertos) e *scorpion* (11 acertos), enquanto o descritor CS-LMP teve um desempenho baixo nas classes *background_google* (9 acertos), *buddha* (13 acertos), *butterfly* (6 acertos), *ibis* (13 acertos) e *starfish* (9 acertos). Pode-se observar que os descritores M-LMP, CS-LMP e LIOP confundiram muitas imagens da classe *ibis* com imagens da classe *kangaroo*, enquanto SIFT e CS-LMP confundiram muitas imagens da classe face com imagens da classe *faces_easy*.

A Tabela 48 mostra o pior desempenho dos quatro métodos comparados. Nota-se mais uma vez que a classe *background_google* é a classe mais difícil para classificação. Mesmo quando se utiliza uma malha mais esparsa (*Hessian Multiscale*), a alta dissimilaridade intraclasse, incluindo as imagens da classe *background_google*, diminui drasticamente o desempenho de todos os métodos.

A Tabela 49 mostra o tempo de processamento em segundos para classificação de objetos usando os descritores M-LMP ($k = 500$), SIFT ($k = 550$), LIOP ($k = 350$) e CS-LMP ($k = 600$). O descritor LIOP consome menos tempo de processamento, mas tem o pior desempenho entre os quatro métodos comparados (Tabela 47).

A Tabela 50 mostra um resumo comparando os quatro descritores, levando em consideração os dois métodos utilizados para calcular as regiões de interesse (*dense sampling* e *Hessian Multiscale*). Em geral, os descritores M-LMP e SIFT atingiram o melhor desempe-

Figura 58: Matriz de confusão para o descritor SIFT usando *Hessian Multiscale* como detector de regiões de interesse (Caltech 101)

Tabela 48: O pior desempenho (taxa de acerto) dos métodos comparados (*Hessian Multiscale*)

	M-LMP	SIFT	LIOP	CS-LMP
<i>background_google</i>	40,00% (12 acertos)	30,00% (9 acertos)	30,00% (9 acertos)	30,00% (9 acertos)
<i>butterfly</i>	43,34% (13 acertos)	20,00% (6 acertos)	13,34% (4 acertos)	20,00% (6 acertos)
<i>menorah</i>	46,67% (14 acertos)	-	-	-
<i>scorpion</i>	40,00% (12 acertos)	-	36,67% (11 acertos)	-
<i>helicopter</i>	-	-	43,34% (13 acertos)	-
<i>ibis</i>	-	46,67% (14 acertos)	43,34% (13 acertos)	43,34% (13 acertos)
<i>buddha</i>	-	-	43,34% (13 acertos)	43,34% (13 acertos)
<i>starfish</i>	-	-	-	30,00% (9 acertos)
<i>chandelier</i>	-	-	43,34% (13 acertos)	-

inho. Ambos os métodos alcançaram uma sensibilidade perto de 78% em 40 milissegundos de tempo de execução usando a técnica *dense sampling*, e cada um atingiu 100% (30 acertos) em cinco classes. Contudo, usando o detector *Hessian Multiscale*, nosso método utilizando o descritor M-LMP ultrapassou o descritor SIFT tanto em sensibilidade como em número de classes classificadas com 100% de acerto (30 acertos).

Experimento-12: foi utilizada a base de imagens ImageNet, com 27 classes e 80 imagens por classes. Subdividiu-se as imagens em 50 imagens para treinamento e 30 imagens para teste e utilizou *dense sampling* para detectar as regiões de interesse nas imagens. Foi ajustado o parâmetro β e o descritor M-LMP foi comparado com outros três descritores: CS-LMP, SIFT e LIOP.

Foi fixado o tamanho do *cluster* em $k = 250$ e o tamanho do descritor em 80 elementos

Figura 59: Matriz de confusão para o descritor LIOP usando *Hessian Multiscale* como detector de regiões de interesse (Caltech 101)

Tabela 49: Tempo de processamento (em segundos) para classificação de objetos da base Caltech 101 usando os descritores M-LMP, SIFT, LIOP e CS-LMP (*Hessian Multiscale*)

M-LMP 80 elementos (k=500)	SIFT 128 elementos (k=550)	LIOP 144 elementos (k=350)	CS-LMP 80 elementos (k=600)
0,0505s	0,0559s	0,0379s	0,0620s

Tabela 50: Desempenho dos métodos

	<i>Dense Sampling</i>				<i>Hessian Multiscale</i>			
	M-LMP	SIFT	LIOP	CS-LMP	M-LMP	SIFT	LIOP	CS-LMP
Sensibilidade	78,15%	78,64%	68,52%	77,65%	77,78%	72,59%	63,46%	68,15%
Tamanho do cluster	400	450	400	400	500	550	350	600
Tempo de execução	0,0406s	0,0464s	0,0414s	0,0417s	0,0505s	0,0559s	0,0379s	0,0620s
30 acertos = 100%	5 classes	5 classes	3 classes	6 classes	5 classes	2 classes	-	1 classe
<50%	4 classes	3 classes	6 classes	4 classes	4 classes	3 classes	7 classes	5 classes

tanto para o M-LMP como para o CS-LMP, baseado nos experimentos anteriores.

A Tabela 51 apresenta o ajuste do parâmetro β para o descritor M-LMP utilizando o detector *dense sampling*. O parâmetro β que será utilizado nos próximos experimentos foi o que obteve a maior sensibilidade ($\beta = 0.0039$). Para o descritor CS-LMP foi utilizado o mesmo parâmetro $\beta = 0.0012$ encontrado na Tabela 41.

Para os resultados apresentados na Tabela 52, foram utilizadas 50 imagens para treinamento (geração do vocabulário) e 30 imagens para teste (classificação) para cada classe

Figura 60: Matriz de confusão para o descritor CS-LMP usando *Hessian Multiscale* como detector de regiões de interesse (Caltech 101)

Tabela 51: Ajuste do parâmetro β para o descritor M-LMP usando 80 imagens por classe (ImageNet)

β	Sensibilidade (S)
0.000010	70,00%
0.000110	71,98%
0.003900	75,68%
0.003910	75,56%
0.000951	73,33%
0.001100	73,58%
0.001200	73,70%
0.001300	74,44%
0.000100	70,86%
0.000020	71,85%
0.000200	71,11%
0.006100	73,83%
0.007100	73,95%
0.001110	73,83%

variando-se o tamanho do *cluster* (k). Os valores em negrito indicam a melhor sensibilidade alcançada para cada descritor. O descritor M-LMP obteve a melhor sensibilidade

de $S=77,65\%$ para $k = 500$. O descritor SIFT obteve $S=77,53\%$ para $k = 550$, o descritor LIOP atingiu $S=67,41\%$ para $k = 500$, e CS-LMP alcançou $S= 73,09\%$ quando $k = 300$. Pode-se observar que o descritor M-LMP obteve uma sensibilidade um pouco maior que o descritor SIFT com um valor de k menor.

Tabela 52: Desempenho dos descritores comparados (sensibilidade) usando o detector *dense sampling* (para 27 classes da base de imagens ImageNet)

Tamanho do <i>cluster</i>	M-LMP: 80 elementos	SIFT: 128 elementos	LIOP: 144 elementos	CS-LMP: 80 elementos
$k = 250$	75,68%	75,56%	65,06%	71,98%
$k = 300$	76,17%	75,93%	66,42%	73,09%
$k = 500$	77,65%	76,67%	67,41%	72,96%
$k = 550$	76,67%	77,53%	66,79%	72,72%

As Figuras 61 até 64 mostram as matrizes de confusão com as classificações corretas colocadas na diagonal, para cada uma das 30 imagens nas 27 classes da base de imagens ImageNet. As linhas referem-se às classes de entrada e as colunas às classes atribuídas. A sensibilidade (S), foi calculada usando a Equação 28.

Figura 61: Matriz de confusão para o descritor M-LMP usando *dense sampling* (ImageNet)

Figura 62: Matriz de confusão para o descritor SIFT usando *dense sampling* (ImageNet)

Analisando os resultados das matrizes de confusão (Figuras 61 até 64), pode-se observar que os descritores M-LMP e CS-LMP tiveram um excelente desempenho (definido com 30 acertos, ou 100% de classificações corretas) nas classes *bonsai*, enquanto os descritores LIOP e SIFT não tiveram 30 acertos em nenhuma classe. Dos quatro descritores analisados, somente o LIOP exibiu baixo desempenho (abaixo de 50% de classificações corretas) nas classes *butterfly* (14 acertos), *coffecup* (12 acertos), *fish* (12 acertos) e *menorah* (13 acertos).

Experimento-13: foi realizado o mesmo teste do Experimento-12, porém o detector de regiões de interesse utilizado foi o *Hessian-Multiscale* ao invés do *dense sampling*.

A Tabela 53 apresenta os resultados do ajuste do parâmetro β para o descritor M-LMP usando o detector de regiões de interesse *Multiscale Hessian*, e a Tabela 54 apresenta o ajuste do parâmetro β para o descritor CS-LMP. Nas duas Tabelas foram escolhidos o parâmetro β que obteve a maior sensibilidade.

Foi comparado o descritor M-LMP com outros três descritores (CS-LMP, SIFT e LIOP) para o reconhecimento de objetos usando o detector *Hessian Multiscale*. Escolheu-se o parâmetro $\beta = 0,0009110$ da Tabela 53 e $\beta = 0.0029$ da Tabela 54 onde foi obtida a melhor sensibilidade. Na Tabela 55, foram utilizadas 50 imagens para o treinamento (geração do vocabulário) e 30 imagens para teste (classificação) para cada classe, e variou-se o tamanho do *cluster* (k). Os valores em negrito indicam a melhor sensibilidade alcan-

Figura 63: Matriz de confusão para o descritor LIOP usando *dense sampling* (ImageNet)

Tabela 53: Sintonização do parâmetro β na base de imagens ImageNet para o descritor M-LMP

β	Sensibilidade (S)
0,0009110	67,41%
0,0000001	66,79%
0,0010000	66,42%
0,0005000	66,79%
0,0008000	66,91%
0,0009100	66,91%
0,0009200	67,28%
0,0009300	67,28%
0,0009600	64,81%
0,0009490	67,16%
0,0009510	67,04%
0,0011180	67,16%
0,0011190	67,04%
0,0012220	66,17%

çada para cada método. Para o descritor M-LMP a melhor sensibilidade foi $S = 71,60\%$ para $k = 550$, SIFT obteve $S = 70,99\%$ quando $k = 550$, LIOP teve $S = 64,07\%$ para $k = 250$, e por fim, CS-LMP alcançou $70,12\%$ para $k = 350$.

Figura 64: Matriz de confusão para o descritor CS-LMP usando *dense sampling* (ImageNet)

Tabela 54: Sintonização do parâmetro β da base de imagens ImageNet para o descritor CS-LMP

β	Sensibilidade (S)
0,0029	68,64%
0,0001	67,16%
0,0100	68,52%
0,0040	68,02%
0,0050	67,78%
0,0060	67,78%
0,0070	67,78%
0,0900	62,10%
0,0021	67,41%
0,0031	68,02%

As Figuras 65 até 68 mostram as matrizes de confusão para as 27 classes da base de imagens ImageNet, onde as linhas referem-se as classes de entrada e as colunas as classes atribuídas.

Analisando os resultados das matrizes de confusão, observa-se que nenhum método obteve 30 acertos. O descritor M-LMP teve o melhor desempenho (29 acertos) em duas classes: *leopard* e *wine*, enquanto o descritor SIFT teve 27 acertos nas mesmas classes.

Tabela 55: Desempenho dos descritores comparados (Sensibilidade) usando o detector de regiões *Hessian Multiscale* (para 27 classes de imagens da base de imagens ImageNet)

Tamanho do <i>cluster</i>	M-LMP: 80 elementos	SIFT: 128 elementos	LIOP: 144 elementos	CS-LMP: 80 elementos
$k = 250$	67,41%	69,63%	64,07%	68,64%
$k = 350$	69,51%	70,25%	63,09%	70,12%
$k = 450$	67,16%	70,62%	63,95%	68,15%
$k = 550$	71,60%	70,99%	62,84%	68,02%

Figura 65: Matriz de confusão para o descritor M-LMP usando *Hessian Multiscale* (ImageNet)

M-LMP e CS-LMP alcançaram um desempenho baixo (abaixo de 50% de classificações corretas) em *car_elevator_car* (13 acertos e 12 acertos respectivamente). O descritor LIOP teve um baixo desempenho nas classes *butterfly* (13 acertos), *car_elevator_car* (7 acertos), *ewer* (14 acertos) e *menorah* (13 acertos), enquanto o descritor SIFT não teve baixo desempenho em nenhuma classe analisada.

5.3.3 Resultados obtidos com o descritor MCS-LMP

Os resultados apresentados nesta Sub-seção referem-se à base de imagens *Indoor Scene Recognition*.

Experimento-14: O descritor MCS-LMP foi avaliado aplicando a metodologia pro-

Figura 66: Matriz de confusão para o descritor SIFT usando *Hessian Multiscale* (ImageNet)

Figura 67: Matriz de confusão para o descritor LIOP usando *Hessian Multiscale* (ImageNet)

Figura 68: Matriz de Confusão para o descritor CS-LMP usando o detector *Hessian Multiscale* (ImageNet)

posta para classificação de imagens. Foram utilizadas 8 classes, com 200 imagens por classes escolhidas aleatoriamente da base *Indoor Scene Recognition dataset* ⁵.

O descritor MCS-LMP foi comparado com outros três descritores CS-LMP, SIFT e LIOP. Foram selecionadas 30 imagens para o dicionário de palavras visuais, 170 imagens de cada classe para o conjunto de treinamento (1360 imagens) e 30 imagens de cada classe para o conjunto de teste (240 imagens). Para todas as imagens foram aplicadas o detector de regiões de interesse *Hessian-Laplace*.

Os descritores foram agrupados pelo método *Lloyd k-means* (LLOYD, 1982) que utiliza o diagrama de Voronoi ao invés de determinar o centro mais próximo dos conjuntos de pontos. Foi considerado o vocabulário visual de 350 palavras para as 8 classes.

Foi realizado o ajuste do parâmetro β para os descritores MCS-LMP e CS-LMP, de acordo com sua sensibilidade, conforme mostrado nas Tabelas 56 e 57. Os outros parâmetros dos descritores MCS-LMP e CS-LMP foram configurados experimentalmente como $b = 6$, $R = 2$ e $N = 8$.

Foram utilizados os parâmetros encontrados nas Tabelas 56 e 57, nesta Sub-seção são apresentadas as matrizes de confusão dos resultados de classificação obtidos pelos descritores MCS-LMP, CS-LMP, SIFT e LIOP nas Figuras 69 até 72, onde as linhas se referem-se à categoria real e às colunas as categorias atribuídas. Para o parâmetro $b = 6$

⁵<http://web.mit.edu/torralba/www/indoor.html>

Tabela 56: Ajuste do parâmetro β para o descritor MCS-LMP.

β	Sensibilidade
0,00010	58,75%
0,00080	59,17%
0,00090	57,50%
0,00110	60,83%
0,00120	57,50%
0,00130	61,67%
0,00131	63,33%
0,01000	53,75%

Tabela 57: Ajuste do parâmetro β para o descritor CS-LMP.

β	Sensibilidade
0,00010	61,25%
0,00080	59,17%
0,00090	58,33%
0,00110	55,83%
0,00120	60,83%
0,00130	55,42%
0,00131	62,08%
0,01000	60,00%

escolhido empiricamente, baseado em experimentos anteriores, gerou-se um tamanho de 112 posições para os descritores MCS-LMP e CS-LMP.

Figura 69: Matriz de confusão para o descritor MCS-LMP usando *Intersection Kernel* como um núcleo para o SVM aplicado sobre a base *Indoor Scene Recognition*.

Figura 70: Matriz de confusão para o descritor CS-LMP usando *Intersection Kernel* como um núcleo para o SVM aplicado sobre a base *Indoor Scene Recognition*.

Figura 71: Matriz de confusão para o descritor SIFT usando *Intersection Kernel* como um núcleo para o SVM aplicado sobre a base *Indoor Scene Recognition*.

Os descritores MCS-LMP e CS-LMP obtiveram o melhor desempenho nas classes *bakery*, com 23 acertos, enquanto na mesma classe SIFT e LIOP obtiveram 18 e 17 acertos, respectivamente. Por outro lado, o descritor MCS-LMP apresentou o pior desempenho na classe *bedroom* com 14 acertos, enquanto o descritor CS-LMP obteve 14 acertos na classe *toystore*, o descritor SIFT obteve 14 acertos na classe *corridor* e o descritor LIOP

Figura 72: Matriz de confusão para o descritor LIOP usando *Intersection Kernel* como um núcleo para o SVM aplicado sobre a base *Indoor Scene Recognition*.

alcançou 10 acertos na classe *bedroom*. Nosso descritor proposto CS-LMP obteve o melhor desempenho na classe *kitchen* com 24 acertos. O descritor SIFT teve o melhor desempenho na classe *casino* com 23 acertos, e o descritor LIOP 27 acertos na mesma classe.

A Tabela 58 mostra o tempo de processamento (em segundos) para a construção do vocabulário visual com o tamanho dos descritores MCS-LMP e CS-LMP de 112 posições cada, SIFT com 128 posições e LIOP com 144 posições. Como o tamanho do descritor influencia no tempo de processamento, quanto menor seu tamanho mais rápido será seu processamento. Como os descritores propostos MCS-LMP e CS-LMP tem os menores tamanhos, o tempo de processamento é menor. A configuração do computador para o teste foi Intel (R) Core (TM) i7-4770k CPU com 8GB de RAM e sistema operacional Windows 8.0 Enterprise.

Tabela 58: Tempo de processamento em segundos para a construção do vocabulário visual usando os descritores MCS-LMP, CS-LMP, SIFT e LIOP.

MCS-LMP	CS-LMP	SIFT	LIOP
4,53s	4,83s	5,19s	5,20s

Experimento-15: foram usadas as mesmas condições do Experimento-14 apenas trocando o núcleo do classificador SVM pelo *Chi-Square Kernel*.

Foi utilizado o mesmo parâmetro β encontrado nas Tabelas 56 e 57. As matrizes de confusão dos descritores MCS-LMP, CS-LMP, SIFT e LIOP são apresentadas nas Figuras 73, 74, 75 e 76 usando o núcleo qui-quadrado para o classificador SVM.

Figura 73: Matriz de confusão para o descritor MCS-LMP usando *Chi-Square kernel* para o classificador SVM aplicado sobre a base *Indoor Scene Recognition*.

Figura 74: Matriz de confusão para o descritor CS-LMP usando *Chi-Square kernel* para o classificador SVM aplicado sobre a base *Indoor Scene Recognition*.

O descritor proposto MCS-LMP obteve melhor desempenho nas classe *bakery* com 25 acertos, enquanto na mesma classe, os descritores CS-LMP, SIFT e LIOP obtiveram 22, 11 e 22 acertos, respectivamente. Nosso descritor MCS-LMP apresentou desempenho ruim na classe *bedroom* com 12 acertos, enquanto o descritor CS-LMP teve 13 acertos em ambas as classes *bedroom* e *toystore*. O descritor SIFT acertou apenas 11 imagens

Figura 75: Matriz de confusão para o descritor SIFT usando *Chi-Square kernel* para o classificador SVM aplicado sobre a base *Indoor Scene Recognition*.

Figura 76: Matriz de confusão para o descritor LIOP usando *Chi-Square kernel* para o classificador SVM aplicado sobre a base *Indoor Scene Recognition*.

na classe *bakery* e o descritor LIOP obteve 6 acertos na classe *bedroom*. Nosso descritor CS-LMP atingiu o melhor desempenho na classe *kitchen* com 23 acertos e o descritor SIFT foi melhor na classe *casino* com 24 acertos, enquanto LIOP obteve 26 acertos na mesma classe.

5.4 Considerações Finais

Neste Capítulo foram apresentados os resultados obtidos na utilização dos descritores propostos.

Os primeiros resultados obtidos com os descritores CS-LMP e M-LMP para correspondência de imagens forneceram subsídios valiosos para verificação de número de pontos correspondentes gerados corretamente pela metodologia proposta. Os descritores foram comparados com os descritores SIFT e CS-LBP e analisados via curvas *recall* $\times$ *precision* e AUC. A análise dos resultados foi vantajosa para os descritores CS-LMP e M-LMP mostrando uma maior correspondência correta de pontos entre duas imagens com variação de escala, rotação e iluminação do que os gerados pelos descritores comparados. Após isso, os *bins* do histograma foram reduzidos para ambos os descritores, e consequentemente o tamanho dos descritores propostos apresentaram melhores resultados comparado ao descritores SIFT e CS-LBP, e com a redução, o tempo de processamento também diminuiu, ficando mais rápido que o SIFT e o CS-LBP.

Os resultados seguintes utilizaram a metodologia BoF para reconhecimento de objetos, cenas e faces. Foram utilizadas as bases de imagens Caltech 2004, Pascal VOC 2006, Caltech 101, *Indoor Scene Recognition*, Feret, ORL e *Yale Face*. Neste caso, os descritores propostos foram comparados aos descritores SIFT e LIOP. Na maioria dos resultados obtidos, a sensibilidade do descritor M-LMP ficou próxima da do descritor SIFT, porém o descritor SIFT precisa de um número maior de *clusters* para a construção do vocabulário visual. Além disso, o tamanho dos descritores M-LMP e CS-LMP são menores do que para os outros descritores comparados, melhorando o tempo de processamento.

Conclusão

Nesta tese foram propostos três novos descritores de regiões de interesse com o objetivo de extrair características locais para classificação e detecção de objetos, faces e cenas. Além dos descritores propostos, foi também utilizado um *framework* moderno para reconhecimento e classificação das imagens analisadas, baseado na abordagem *Bag-of-Features* (BoF), que descreve cenas ou objetos pertencentes à mesma por meio de um dicionário de palavras visuais. Com isso, foi possível avaliar os descritores em torno dos pontos-chave sem realizar a segmentação das imagens e comparar os resultados obtidos com descritores amplamente conhecidos na literatura científica.

Para tal, uma metodologia de experimentos foi aplicada visando avaliar em cada etapa o grau de eficiência de cada descritor dentro de bases de imagens de alta complexidade para o reconhecimento.

Antes da sequência de experimentos para reconhecimento das imagens, foi avaliado se os descritores propostos discriminam corretamente os pontos-chave correspondentes em pares de imagens que sofreram transformações, de maneira coerente, em número suficiente para o reconhecimento e com menos *outliers* que os descritores existentes.

Utilizando o detector de regiões de interesse *Harris-Affine*, os descritores propostos M-LMP, e CS-LMP foram comparados com dois descritores da literatura, o CS-LBP e o SIFT. Por meio de curvas *recall* $\times$ *precision*, foram medidas as *Area Under Curve* (AUC) para cada par de imagens analisadas. Os descritores propostos CS-LMP e M-LMP apresentaram valores de AUC maiores do que os gerados pelos descritores SIFT e CS-LBP, com maior número de pares de pontos correspondentes corretos. Nas imagens avaliadas, o desempenho dos descritores propostos ultrapassou o desempenho dos outros descritores. Este experimento foi conclusivo para sua aplicação no reconhecimento de imagens usando o *framework* BoF.

Na literatura, existem diversas bases de imagens disponíveis para testes de algoritmos de visão computacional. Dada às complexidades das cenas e a consequente dificuldade de interpretação e classificação, diversos algoritmos publicados na área buscam o reconhecimento dos objetos presentes nas imagens realizando um recorte manual (*cropping*)

e centralizando os objetos a serem reconhecidos, o que não pode ser equiparado ao método proposto nesta tese. Este é o caso dos trabalhos publicados no ILSVRC-ImageNet *Large Scale Visual Recognition Challenge* (RUSSAKOVSKY et al., 2015) cujas imagens são pré-processadas, recortadas e padronizadas em uma resolução pré-definida. Além disso, para o desafio ILSVRC, o erro de classificação é definido em *Top-5*, ou seja, os algoritmos retornam as cinco imagens mais similares, o que aumenta significativamente os índices de acurácia. Nos experimentos realizados neste trabalho foi utilizada a metodologia *Top-1* que procura pela imagem mais similar na base de consulta, o que deixa o tempo de execução do algoritmo mais rápido.

No Experimento-1 foram utilizadas quatro classes de imagens da base Caltech 2004 que incluem 1074 aviões, 1155 carros, 450 faces e 826 motos. O erro de classificação obtido foi de apenas 15,0% para o descritor proposto M-LMP contra 23,0% de erro para o SIFT e 24,0% para o CS-LBP. Como o número de classes utilizadas foi pequeno neste experimento, foi realizado o Experimento-2 aumentando-se o número de classes para seis. Com o aumento do número de classes o desempenho de todos os métodos caiu, mas o descritor M-LMP continuou com maior taxa de reconhecimento. Neste experimento, as taxas de erro foram de 27,0% para o M-LMP, 32,0% para o SIFT e 33,7% para o CS-LBP.

Para o Experimento-3 foram utilizadas quatro classes da base de dados Pascal VOC 2006 com 180 imagens de ônibus, 197 imagens de vacas, 247 imagens de cavalos e 251 imagens de ovelhas. Dada a diversidade de fundos existentes nas imagens de uma mesma classe desta base, o que a torna mais difícil para a tarefa de reconhecimento dos objetos principais localizados na cena, o desempenho dos três descritores caiu. As taxas de erro obtidas foram de 45% para o M-LMP e para o SIFT e de 49% para o CS-LBP.

Para o Experimento-4 foram escolhidas quatro classes da base de dados Caltech 101: *bonsai*, *faces_easy*, *leopards* e *watch*, com imagens com variação de rotação, iluminação, escala, compressão e aliasing. O objetivo deste experimento foi a redução do tamanho do descritor variando o parâmetro b = número de *bins* do descritor M-LMP. Como o descritor proposto é paramétrico, este é um teste necessário para que se possa verificar a eficiência do mesmo com diferentes tamanhos de histograma. A constante troca de bases de imagens visa a comparação das técnicas em diferentes situações. Neste caso, apesar das transformações existentes nas imagens desta base, as mesmas possuem seus fundos “mais comportados” sem grandes variações dentro de uma mesma classe. As taxas de erro foram de 14,16% para o M-LMP com 256 *bins* e de 12,50% para o M-LMP com 80 *bins*. Já os descritores comparados tiveram maior taxa de erro, sendo de 19,17% para o SIFT e de 21,67% para o CS-LBP. Também neste experimento o descritor proposto mostrou-se robusto a transformações nas imagens e que a redução do número de *bins* pode ser utilizada quando a aplicação necessitar de menor tempo de execução, com ligeiro aumento de acurácia. Aumentando-se o número de classes de quatro para cinco e mantendo-se o número reduzido de *bins* em 80, o Experimento-5 mostra mais

uma vez uma redução de desempenho da metodologia, mas com taxa de erro inferior para o descritor proposto M-LMP (22,00%) em comparação com o SIFT (27,00%) e CS-LBP(32,67%). É de se observar, no entanto, que a classe que gerou o maior número de erros de reconhecimento foi a classe introduzida neste experimento (*airplanes*) o que permitiu o acerto de apenas dezessete dos trinta elementos de classe para o descritor M-LMP. Analisando-se visualmente as imagens desta classe, mais uma vez constata-se que o fundo (*background*) dos elementos de classe são os mais variados, o que contribui significativamente com a queda de desempenho.

Em todos os experimentos anteriores, o classificador utilizado foi do tipo K-NN, com retorno do elemento mais semelhante. Como o classificador SVM tem a tendência de trabalhar melhor do que o K-NN com outliers, pois usa pontos relevantes para encontrar a separação das classes, decidiu-se a partir deste experimento substituir o classificador.

O objetivo do Experimento-6 foi aplicar a metodologia proposta no reconhecimento de faces humanas. Normalmente, as bases de faces humanas disponíveis são bem comportadas, com fundos padronizados e quase sempre com imagens frontais. Neste experimento foi escolhida a base Feret e o descritor proposto M-LMP foi comparado com dois descritores comumente utilizados para reconhecimento de faces humanas, o SIFT e o LIOP. Além disso, foi escolhido um método da literatura com resultados sobre a base Feret e denominado de *Face Fuzzy*. O M-LMP (com 95,41% de acertos) e o LIOP (com 95,56% de acertos) empataram neste quesito com ligeira vantagem para o M-LMP no número de agrupamentos necessários para a geração do dicionário, o que dá menor custo computacional para o descritor proposto. O SIFT, neste caso teve o pior desempenho entre os três descritores comparados (92,00% de acertos e maior número de agrupamentos necessários).

Considerando-se o bom desempenho obtido com a base Feret, no Experimento-7 foi utilizada a base de imagens ORL e a base Yale Face. Neste caso, a motivação foi comparar a abordagem proposta com vários métodos de reconhecimento facial, já previamente avaliados e publicados na literatura (SOLDERA; BEHAINE; SCHARCANSKI, 2015). O número de acertos para o descritor proposto usando a abordagem proposta foi de 97,50% para a base ORL e de 100% para a base Yale Face. Nenhum dos métodos comparados atingiu tal desempenho, demonstrando mais uma vez um grande potencial do método proposto com o descritor M-LMP para reconhecimento de imagens com fundo sem grandes variações, mesmo que sofram transformações de rotação, escala e iluminação.

Para os próximos experimentos, foram extraídas regiões de interesse utilizando duas abordagens: *Hessian-Multiscale* (MIKOLAJCZYK; SCHMID, 2002) e *dense sampling* (ISCEN et al., 2015). *Hessian Multiscale* usa o determinante da matriz Hessiana em multi-escala para detectar características em múltiplas escalas. No detector *dense sampling*, o centro da região de interesse é estabelecido fazendo um deslocamento a cada 2 pixels e assim é construído uma malha regular de tamanho e forma fixa. O objetivo desse experimento foi verificar diferentes detectores de Regiões de Interesse e avaliar suas influências dentro da

metodologia proposta baseada em BoF. Também o método de agrupamento foi alterado para o Elkan k-means (ELKAN, 2003). Para a avaliação da metodologia proposta foram utilizadas duas bases de imagens: Caltech 101 e ImageNet, das quais foram escolhidas aleatoriamente 27 classes contendo 80 imagens por classe. Estas são as bases mais utilizadas no *ILSVRC-ImageNet Large Scale Visual Recognition Challenge*, mas com as ressalvas já discutidas anteriormente. O grande desafio aqui foi colocar à prova dois descritores propostos nesta tese, o M-LMP extensivamente explorado nos experimentos anteriores e o CS-LMP, contra os tradicionais descritores da literatura SIFT e LIOP em bases com maior número de imagens e de difícil classificação.

No Experimento-8 foi utilizada a base Caltech 101 e *dense sampling* para definir o menor tamanho para os descritores propostos M-LMP e CS-LMP. Foi verificado que 80 elementos para os descritores mantinham um desempenho ótimo de reconhecimento. O objetivo do Experimento-9 foi verificar se todas as imagens da base devem ser usadas para o ajuste do parâmetro β tanto para o M-LMP como para o CS-LMP. Conclui-se que, para ajustar devidamente o parâmetro β para a mais alta sensibilidade, é necessário usar todas as imagens para cada classe.

No Experimento-10, utilizando-se os parâmetros sintonizados pelo Experimento-8 e Experimento-9 na base Caltech 101, fez-se a comparação entre os métodos propostos com M-LMP, CS-LMP e os descritores SIFT e LIOP usando *dense sampling*. O M-LMP com 80 elementos teve desempenho similar ao do SIFT com 128 elementos. O M-LMP acertou 78,15% das imagens da base com 78,64% de acertos para o SIFT. É importante ressaltar, no entanto, que além do tamanho menor do descritor, o número de agrupamentos necessários para o dicionário visual é inferior para o M-LMP do que para o SIFT o que dá ligeira vantagem para o descritor proposto em tempo de computação. Já o CS-LMP também proposto e o LIOP tiveram desempenhos bem menores. Analisando-se visualmente as imagens desta base observa-se que quando o fundo da mesma é mais comportado, caso nas classes *faces* e *face_easy*, os dois descritores classificam corretamente todos os elementos da classe. Em compensação, quando o fundo é muito diverso como na classe denominada de *background_google* todos os descritores apresentam o pior desempenho.

Para o Experimento-11 foi utilizado o mesmo procedimento do Experimento-10, mas, utilizando o detector de região *Hessian Multiscale* ao invés de *dense sampling*. A taxa de acertos do descritor M-LMP foi superior aos outros comparados, além de ter menor tamanho e utilizar número de agrupamentos menor. É importante notar também que todos os descritores tiveram uma pequena redução na taxa de acerto quando se utilizou o detector de regiões *Hessian Multiscale* comparativamente ao *dense sampling*. Conclui-se pela utilização de *dense sampling* tanto devido a uma maior taxa de acerto como menor tempo computacional. E dentre os descritores comparados, conclui-se favoravelmente ao M-LMP com taxas de acerto similar ao SIFT, mas de menor tamanho e com número de agrupamentos menor.

No Experimento-12 foi utilizada a base de imagens ImageNet, com 27 classes e 80 imagens por classe e dense sampling, e no Experimento-13 foi utilizada a mesma base e *Hessian Multiscale*. Os resultados foram similares aos obtidos com a base Caltech 101 realizados no Experimento-10 e Experimento-11. Isto demonstra a robustez do método proposto como um todo e a vantagem do descritor proposto M-LMP sobre os outros comparados, tanto no desempenho de número de acertos como em menor tempo computacional devido ao menor tamanho do descritor e também a um número menor de agrupamentos necessários para geração do dicionário visual.

O objetivo do Experimento-14 foi avaliar uma variação do descritor CS-LMP denominada de MCS-LMP. Para isso foram utilizadas imagens mais complexas que as testadas anteriormente pertencentes à base *Indoor Scene Recognition Dataset*. Esta base é de difícil classificação, pois não possuem objetos principais sobre fundo complexo, mas mostram cenas de interiores de padarias, quartos de dormir, corredores, cassinos, mercearias, cozinhas, restaurantes e lojas de brinquedos. Seres humanos necessitam de “conhecimento anterior” para classificarem corretamente tais imagens e a metodologia clássica de visão computacional baseada em segmentação de objetos provavelmente logrará êxito. O descritor MCS-LMP foi comparado com os descritores CS-LMP, SIFT e LIOP. Foram utilizadas 30 imagens de cada classe para o conjunto de teste (240 imagens). Para todas as imagens foram aplicadas o detector de regiões de interesse *Hessian-Laplace* e o classificador SVM com *Intersection Kernel*. Apesar da baixa taxa de acertos de todos os descritores, o que era de se esperar dada a complexidade das cenas de ambientes internos, os descritores propostos MCS-LMP e CS-LMP tiveram maior desempenho (63,33% e 62,08% respectivamente) em menor tempo de computação do que os dois descritores comparados SIFT (60,83%) e LIOP (59,17%). E, finalmente, no Experimento-15 foi alterado o núcleo do classificador SVM para *Chi-Square kernel*. O que se observou nos resultados é que as taxas de acertos dos descritores comparados sofreram quase nenhuma alteração em seus valores. Mas, os descritores propostos MCS-LMP e CS-LMP foram superiores aos descritores da literatura.

Após todos os exaustivos experimentos realizados nesta tese, conclui-se que os descritores propostos (CS-LMP, MCS-LMP e M-LMP), todos derivados da abordagem *Local Mapped Pattern* (LMP), são mais rápidos e mais eficientes que descritores existentes tais como SIFT, CS-LBP e LIOP. O uso de pontos-chave e dicionários visuais permite o reconhecimento de objetos em imagens com fundos complexos sem a necessidade de segmentação ou recorte (*cropping*) dos objetos principais, tornando o processo de classificação mais automático. Imagens com variação no *background* dentro da mesma classe prejudicam o desempenho de todos os descritores, pois, as palavras visuais geradas terão componentes do fundo reduzindo a taxa de acertos. Em bases mais comportadas, como faces humanas, nas quais os fundos são geralmente padronizados, mesmo que as faces sofram rotação, variação de escala e de iluminação, todos os descritores melhoram o seu

desempenho usando a metodologia proposta, com sensível ganho para os descritores propostos. Em imagens complexas, como as cenas de interiores, os resultados são bastante prejudicados, e mesmo utilizando-se um classificador mais robusto a *outliers* como o SVM, com diferentes núcleos, pouco ganho é conseguido na taxa de acertos. Mas, mesmo neste caso, os descritores propostos nesta tese alcançaram maior desempenho.

Todos os resultados avaliados aqui indicam que novos descritores devem ser investigados, pois a abordagem *Local Mapped Pattern* permite variações não somente na maneira como as diferenças de níveis de cinza podem ser utilizadas, mas também no formato da função de mapeamento. No caso deste trabalho a função de mapeamento escolhida foi a sigmoidal devido ao fato de estar sendo amplamente investigada pelo grupo de pesquisa do LAVI - Laboratório de Visão Computacional da EESC-USP. Mas, a continuidade deste trabalho conduz para o uso de outras funções de mapeamento tais como arco tangente já utilizada no descritor denominado de WLD (*Weber Law Descriptor*) (CHEN et al., 2009), função Gaussiana, função triangular que privilegia as bordas da imagem e é o núcleo no descritor denominado de FUNED (*Fuzzy Number Edge Detector*) e outras ainda não avaliadas.

6.1 Contribuições

Nesta tese foram propostos novos descritores locais de imagens utilizando a metodologia BoF. Resumidamente, as principais contribuições para o presente trabalho foram:

- ❑ O desenvolvimento do descritor *Center-Symmetric Local Mapped Pattern* (CS-LMP), que usa uma função sigmóide para mapear os níveis de cinza da imagem. Tal descritor usa os pares simétricos ao centro no cálculo do mapeamento dos pixels de uma vizinhança. O descritor passou por um processo de redução de dimensionalidade e seu desempenho foi comparado com outros descritores amplamente utilizados na literatura.
- ❑ Foi desenvolvido uma extensão do descritor *Center-Symmetric Local Mapped Pattern* (CS-LMP), chamado *Modified Center-Symmetric Local Mapped Pattern* (MCS-LMP), que introduziu a informação do pixel central na modelagem do descritor para melhor caracterizar o conteúdo da imagem. O descritor foi testado em sua forma reduzida em uma base de dados de cenas complexas e comparado com o CS-LMP e outros descritores já publicados na literatura.
- ❑ O terceiro e último descritor desenvolvido foi o *Mean-Local Mapped Pattern* (M-LMP), que calcula a diferença de todos os vizinhos com a média da vizinhança. Tal descritor foi construído com a observação que o LBP original compara os pixels da vizinhança com o pixel central, gerando a possibilidade de ruídos no resultado

final. Para evitar os ruídos e deixar as características mais robustas, a média das sub-regiões foram utilizadas ao invés de comparar o nível de cinza de apenas um pixel.

6.2 Trabalhos Futuros

Nesta tese foram apresentados descritores locais parametrizados, gerando uma necessidade de estudo tanto dos parâmetros envolvidos, quanto da extensão dos métodos propostos.

Estas são algumas sugestões que podem ser utilizadas para trabalhos futuros:

- ❑ Estudo de novas funções de mapeamento;
- ❑ Avaliar o impacto do parâmetro b (número de bins do histograma) na redução do seu tamanho no menor valor possível e sem perda de eficiência;
- ❑ Avaliar os descritores propostos em outros bancos de imagens;
- ❑ Utilizar outros descritores locais para comparar a metodologia proposta;
- ❑ Modificar e testar a função de peso do modelo matemático;
- ❑ Avaliar o desempenho de reconhecimento de objetos com outros classificadores.

6.3 Artigos aceitos e submetidos

As publicações de artigos científicos apresentando o desenvolvimento do projeto foram realizadas durante todo o processo de doutoramento. A seguir são apresentadas as publicações geradas até o momento, juntamente com artigos submetidos.

[1] Vieira, R. T., Ferraz, C. T., Langoni, V. M., Gonzaga, A. Uma nova proposta para análise de textura local por conjuntos fuzzy. *In: VIII Workshop de Visão Computacional, 2012, Goiânia. Anais WVC 2012, 2012.*

[2] Ferraz, C. T., Vieira, R. T., Langoni, V. M., Gonzaga, A. Avaliação de desempenho de detectores de bordas em imagens ruidosas. *In: VIII Workshop de Visão Computacional, 2012, Goiânia. Anais WVC 2012, 2012.*

[3] Vieira, R. T., Chierici, C. E. O., Ferraz, C. T., Gonzaga, A. Local Fuzzy Pattern: A New Way for Micro-pattern Analysis. *Lecture Notes in Computer Science. 1ed.: Springer Berlin Heidelberg, 2012, p. 602-611.*

- [4] Vieira, R. T. ; Chierici, C. E. O., Ferraz, C. T. ; Gonzaga, A. Image micro-pattern analysis using Fuzzy Numbers. *In: SIBGRAPI 2012 - Conference on Graphics, Patterns and Images, 2012, Ouro Preto. Proceedings of SIBGRAPI 2012 - Conference on Graphics, Patterns and Images, 2012.*
- [5] Chierici, C. E. O., Vieira, R. T., Ferraz, C. T., Travaini, J. N., Gonzaga, A. . A new approach for analyzing rotated textures. *In: IX Workshop de Visão Computacional, 2013, Rio de Janeiro. Anais do IX Workshop de Visão Computacional, 2013.*
- [6] Ferraz, C. T., Pereira, O., Gonzaga, A. A new Region of Interest descriptor for image correspondence. *In: IX Workshop de Visão Computacional, 2013, Rio de Janeiro. Anais do IX Workshop de Visão Computacional, 2013.*
- [7] Ferraz, C. T., Pereira, O., Gonzaga, A. Feature Description based on Center-Symmetric Local Mapped Patterns. *In: 29th Symposium on Applied Computing, 2014, Gyeongju, South Korea. Proceedings of the 29th Symposium on Applied Computing. New York: ACM, 2014. p. 39-44.*
- [8] Ferraz, C. T., Pereira, O., Verdini, M. V. R., Gonzaga, A. Object Recognition Based On Bag Of Features and A New Local Pattern Descriptor. *International Journal of Pattern Recognition and Artificial Intelligence, 2014.*
- [9] Ferraz, C. T., Oliveira, O., Gonzaga, A. Feature description based on Mean Local Mapped Pattern. *In: X Workshop de Visão Computacional, 2014, Uberlândia. Anais WVC 2014, 2014.*
- [10] Ferraz, C. T., Gonzaga, A. Improving image classification performance by descriptor size reduction and Bag-of-Features. *In XI Workshop de Visão Computacional, 2015, São Carlos. Anais WVC 2015, 2015.*
- [11] Ferraz, C. T., Manzato, M., Gonzaga A. Complex indoor scene classification based on a new feature descriptor. *International Conference on Pattern Recognition Systems (ICPRS-16), Talca, Chile. p. 3 (6 .), 2016.*
- [12] Ferraz, C. T., Gonzaga A. Object classification using a local texture descriptor and a Support Vector Machine. *Multimedia Tools and Applications, 2016. Submetido.*

Referências

- ABDEL-HAKIM, A. E.; FARAG, A. A. Csift: A sift descriptor with color invariant characteristics. **Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)**, 2006. v. 2, p. 1978–1983, 2006.
- ABID, M. R.; MELO, L. B. S.; PETRIU, E. M. Dynamic sign language and voice recognition for smart home interactive application. In: **Medical Measurements and Applications Proceedings (MeMeA), 2013 IEEE International Symposium on**. [S.l.: s.n.], 2013.
- ALITAPPEH, R. J.; MAHMOUDI, F. Mgs-sift: A new illumination invariant feature based on sift descriptor. **International Journal of Computer Theory and Engineering**, 2013. v. 5, n. 1, p. 99–103, 2013.
- ALITAPPEH, R. J.; SARAVI, K. J.; MAHMOUDI, F. A new illumination invariant feature based on sift descriptor in color space. **Procedia Engineering**, 2012. v. 41, p. 305–311, 2012.
- ALWANI, A. A.; CHAHIR, Y. Neonatal events recognition using lbp descriptor and wavelet thresholding technique. **Multimedia Computing and Systems (ICMCS)**, 2014. p. 427–432, 2014.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. [S.l.]: New York, Addison-Wesley, 1999.
- BAI, G.; ZHU, Y.; DING, Z. A hierarchical face recognition method based on local binary pattern. **CISP08 Proceedings of the 2008 Congress on Image and Signal Processing**, 2008. v. 2, p. 610–614, 2008.
- BARATA, C.; MARQUES, J. S.; MENDONÇA, T. Bag-of-features classification model for the diagnose of melanoma in dermoscopy images using color and texture descriptors. **10th International Conference, ICIAR 2013, Póvoa do Varzim, Portugal, June 26-28, 2013. Proceedings**, 2013. v. 7950, p. 547–555, 2013.
- BATISTA, N. C.; LOPES, A. P. B.; ARAÚJO, A. de A. Vocabulários visuais aplicados à detecção de edifícios em fotografias históricas. In: **XXXV Conferencia Latinoamericana de Informática (Latin-American Conference on Informatics), CLEI, Pelotas, RS, Brazil**. [S.l.: s.n.], 2009.

- BAY, H.; TUYTELAARS, T.; GOOL, L. V. Surf: speeded up robust features. **European Conference on Computer Vision**, 2006. v. 1, p. 404–417, 2006.
- BAZI, Y.; MELGANI, F. Toward an optimal svm classification system for hyperspectral remote sensing images. **IEEE Transactions On Geoscience and Remote Sensing**, 2006. v. 44, n. 11, p. 3374–3385, 2006.
- BEHAINE, C. A. R.; SCHARCANSKI, J. Enhancing the performance of active shape models in face recognition applications. **IEEE Transactions on Instrumentation and Measurement**, 2012. v. 61, n. 8, p. 2330–2333, 2012.
- BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 1997. v. 19, n. 7, p. 711–720, 1997.
- BERKHIN, P. **Survey of Clustering Data Mining Techniques**. [S.l.: s.n.], 2002.
- BOAVENTURA, I. A. G.; GONZAGA, A. Edge detection in digital images using fuzzy numbers. **International Journal of Innovative Computing and Applications**, 2009. v. 2, p. 1–12, 2009.
- CAI, D.; HE, X.; HAN, J.; ZHANG, H. J. Orthogonal laplacianfaces for face recognition. **IEEE Transactions on Image Processing**, 2006. v. 15, n. 11, p. 3608–3614, 2006.
- CALINSKI, T.; HARABASZ, J. A dendrite method for cluster analysis. **Communications in Statistics**, 1974. v. 3, n. 1, p. 1–27, 1974.
- CHANG, C.-C.; LIN, C.-J. Libsvm: A library for support vector machines. **ACM Transactions on Intelligent Systems and Technology**, 2011. v. 2, n. 3, 2011.
- CHEN, J.; SHAN, S.; HE, C.; ZHAO, G.; PIETIKAINEN, M.; CHEN, X.; GAO, W. Wld: A robust local image descriptor. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2009. v. 32, n. 9, p. 1705–1720, 2009.
- CHIERICI, C.; VIEIRA, R. T.; FERRAZ, C. T.; AL. et. A generalized method for texture classification by local pattern analysis. **IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics (submitted)**, 2013. 2013.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, 1995. v. 20, n. 3, p. 273–297, 1995.
- CRÓSTA, A. P. **Processamento digital de imagem de sensoriamento remoto**. [S.l.]: Instituto de Geociências, 1993.
- CUI, J.; CUI, M.; XIAO, B.; LI, G. Compact and discriminative representation of bag-of-features. **Neurocomputing**, 2015. v. 169, p. 55–67, 2015.
- DARDAS, N.; CHEN, Q.; GEORGANAS, N. D. Hand gesture recognition using bag-of-features and multi-class support vector machine. **Proceedings of IEEE International Symposium on Haptic Audio-Visual Environments and Games**, 2010. p. 1 – 5, 2010.
- DUDA, O. R.; HART, P. E.; STORK, D. G. **Pattern Classification**. [S.l.]: A Wiley-Interscience Publication, 2000.

- ELKAN, C. Using the triangle inequality to accelerate k-means. **Proceedings of the Twentieth International Conference on Machine Learning**, 2003. 2003.
- ELSHOURBAGY, M.; HEMAYED, E.; FAYEK, M. Enhanced bag of words using multilevel k-means for human activity recognition. **Egyptian Informatics Journal**, 2016. 2016.
- FAN, R.-E.; CHEN, P.-H.; LIN, C.-J. Working set selection using second order information for training support vector machines. **The Journal of Machine Learning Research**, 2005. p. 1889–1918, 2005.
- FAUDZI, S. A. A. M.; YAHYA, N. Evaluation of lbp-based face recognition techniques. **Intelligent and Advanced Systems (ICIAS)**, 2014. p. 1–6, 2014.
- FEI, L. F.; FERGUS, R.; PERONA, P. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. **IEEE. CVPR 2004, Workshop on Generative-Model Based Vision**, 2004. p. 178, 2004.
- FERNANDEZ, A.; ALVAREZ, M. X.; BIANCONI, F. Texture description through histograms of equivalent patterns. **Journal of Mathematical Imaging and Vision**, 2013. v. 45, p. 76–102, 2013.
- FERRAZ, C. T.; JUNIOR, O. P.; GONZAGA, A. Feature description based on center-symmetric local mapped patterns. **29th Symposium On Applied Computing**, 2014. p. 39–44, 2014.
- FONTANA, A.; NALDI, M. C. **Estudo e Comparação de Métodos para Estimação de Números de Grupos em Problemas de Agrupamento de Dados**. [S.l.], 2009.
- FOODY, G. M.; MATHUR, A. Multiclass and binary svm classification: Implications for training and classification users. **IEEE Geoscience and Remote Sensing Letters**, 2008. v. 5, n. 2, p. 241–245, 2008.
- GUO, B.; GUNN, S. R.; DAMPER, R. I.; NELSON, J. D. B. Customizing kernels functions for svm-based hyperspectral image classification. **IEEE Transactions on Image Processing**, 2008. v. 17, n. 4, p. 622–629, 2008.
- HE, D.-C.; WANG, L. Texture unit, texture spectrum, and texture analysis. **IEEE Transactions On Geoscience and Remote Sensing**, 1990. v. 28, p. 509–512, 1990.
- HE, X.; YAN, S.; HU, Y.; NIYOGI, P.; ZHANG, H.-J. Face recognition using laplacianfaces. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2005. v. 27, n. 3, p. 328–340, 2005.
- HEIKKILÄ, M.; PIETIKÄINEN, M.; SCHMID, C. Description of interest regions with center-symmetric local binary patterns. **Springer-Verlag Berlin Heidelberg**, 2006. 2006.
- _____. Description of interest regions with local binary patterns. **Pattern Recognition**, 2009. v. 42, p. 425–436, 2009.
- HOIEM, D.; CHODPATHUMWAN, Y.; DAI, Q. Diagnosing error in object detectors. **Computer Vision (ECCV)**, 2012. v. 7574, p. 340–353, 2012.

- HUANG, C.; DAVIS, L. S.; TOWNSHEND, J. R. G. An assessment of support vector machines for land cover classification. **International Journal of Remote Sensing**, 2002. v. 23, n. 4, p. 725–749, 2002.
- HUANG, W.; WEI, Y.; XIE, Y.; JIN, H. Survey of local invariant feature description. **Chinese Automation Congress (CAC)**, 2013. p. 353–358, 2013.
- ISCEN, A.; TOLIAS, G.; GOSSELIN, P. H.; JÉGOU, H. A comparison of dense region detectors for image search and fine-grained classification. **IEEE Transactions on Image Processing**, 2015. v. 24, n. 8, p. 2369–2381, 2015.
- ITTI, L.; KOCH, C. Computational modeling of visual attention. **Nature Reviews Neuroscience**, 2001. v. 2, p. 194–203, 2001.
- JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. **ACM Comput. Surv.**, 1999. v. 31, n. 3, p. 264–323, 1999.
- JANSEN, S.; SHANTIA, A.; WIERING, M. A. The neural-sift feature descriptor for visual vocabulary object recognition. **2015 International Joint Conference on Neural Networks (IJCNN)**, 2015. p. 1–8, 2015.
- JIANG, Y.-G.; NGO, C.-W.; YANG, J. Towards optimal bag-of-features for object categorization and semantic video retrieval. In: **Proceedings of the 6th ACM international conference on Image and video retrieval, CIVR '07**. [S.l.: s.n.], 2007. p. 494–501.
- JIN, H.; LIU, Q.; LU, H.; TONG, X. Face detection using improved lbp under bayesian framework. **Proceedings of the Third International Conference on Image and Graphics**, 2004. p. 306–309, 2004.
- JURIE, F.; TRIGGS, B. Creating efficient codebooks for visual recognition. In: **ICCV**. [S.l.: s.n.], 2005. v. 1, p. 604–610.
- KAN, M.; SHAN, S.; ZHANG, H.; LAO, S.; CHEN, X. Multi-view discriminant analysis. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2015. v. 38, n. 1, p. 188–194, 2015.
- KE, Y.; SUKTHANKAR, R. A more distinctive representation for local image descriptors. **Computer Vision and Pattern Recognition**, 2004. v. 2, p. 506–513, 2004.
- K.MIKOLAJCZYK; SCHMID, C. A performance evaluation of local descriptors. **IEEE Trans. Pattern Anal. Mach. Intell.**, 2005. v. 27, n. 10, p. 1615–1630, 2005.
- KOCH, C.; ULLMAN, S. Shifts in selective visual attention: Towards the underlying neural circuitry. **Conceptual Structures in Cognitive Neuroscience**, 1987. v. 188, p. 115–141, 1987.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: PEREIRA, F.; BURGESS, C. J. C.; BOTTOU, L.; WEINBERGER, K. Q. (Ed.). **Advances in Neural Information Processing Systems 25**. Curran Associates, Inc., 2012. p. 1097–1105. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf>>.

- LACHI, R. L.; ROCHA, H. V. da. **Aspectos básicos de clustering: conceitos e técnicas**. [S.l.], 2005.
- LANDGREBE, D. A. **Signal Theory Methods In Multispectral Remote Sensing**. [S.l.]: Wiley Interscience, 2003.
- LATEGAHN, H.; GROSS, S.; STEHLE, T.; AACH, T. Texture classification by modeling joint distributions of local patterns with gaussian mixtures. **IEEE Transactions on Image Processing**, 2010. v. 19, p. 1548–1557, 2010.
- LAZEBNIK, S.; SCHMID, C.; PONCE, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. **IEEE Conference on Computer Vision and Pattern Recognition**, 2006. v. 2, p. 2169–2178, 2006.
- LEI, Z.; YI, D.; LI, S. Z. Local gradient order pattern for face representation and recognition. **22nd International Conference on Pattern Recognition**, 2014. p. 387–392, 2014.
- LI, J.; SANG, N.; GAO, C. Ledtd: Local edge direction and texture descriptor for face recognition. **Signal Processing: Image Communication**, 2016. v. 41, p. 40–45, 2016.
- LI, K.; WANG, F.; ZHANG, L. A new algorithm for image recognition and classification based on improved bag of features algorithm. **Optik**, 2016. v. 127, p. 4736–4740, 2016.
- LI, Z.; IMAI, J. ichi; KANEKO, M. Robust face recognition using block-based bag of words. In: **2010 International Conference on Pattern Recognition**. [S.l.: s.n.], 2010.
- LIAO, S.; ZHU, X.; LEI, Z.; ZHANG, L.; LI, S. Z. Learning multi-scale block local binary patterns for face recognition. **Lecture Notes in Computer Science**, 2007. v. 4642, p. 828–837, 2007.
- LINDEBERG, T. Feature detection with automatic scale selection. **International Journal of Computer Vision**, 1998. v. 30, n. 2, p. 79–116, 1998.
- LIU, H.; LI, S. Target detection using sparse representation with element and construction combination feature. **IEEE Transactions on Instrumentation and Measurement**, 2015. v. 64, n. 2, p. 290–298, 2015.
- LLOYD, S. P. Least squares quantization in pcm. **IEEE Transactions on Information Theory**, 1982. v. 28, n. 2, p. 129–137, 1982.
- LOPES, A. P. B.; AVILA, S. E. F. de; PEIXOTO, A. N. A.; OLIVEIRA, R. S.; COELHO, M. de M.; ARAÚJO, A. de A. Nude detection in video using bag-of-visual-features. In: **SIBGRAPI '09 Proceedings of the 2009 XXII Brazilian Symposium on Computer Graphics and Image Processing**. [S.l.: s.n.], 2009.
- LOPES, A. P. B.; OLIVEIRA, R. S.; ALMEIDA, J. M. de; ARAÚJO, A. de A. Comparing alternatives for capturing dynamic information in bag-of-visual-features approaches applied to human actions recognition. In: **Proceedings of the IEEE International Workshop on Multimedia Signal Processing, MMSP, IEEE Computer Society Press, Rio de Janeiro, RJ, Brazil**. [S.l.: s.n.], 2009.

- LOWE, D. G. Object recognition from local scale-invariant features. **International Conference on Computer Vision**, 1999. p. 1150–1157, 1999.
- _____. Distinctive image features from scale-invariant keypoints. **International Journal of Computer Vision**, 2004. v. 60, p. 91–110, 2004.
- MACQUEEN, J. B. Some methods for classification and analysis of multivariate observations. **Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability**, 1967. v. 1, p. 281–297, 1967.
- MARCEL, S.; RODRIGUES, Y.; HEUSCH, G. On the recent use of local binary patterns for face authentication. **International Journal of Image and Video Processing**, 2007. 2007.
- MARÉ, R.; GEURTS, P.; PIATER, J.; WEHENKEL, L. Random subwindows for robust image classification. **Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition**, 2005. v. 1, p. 34 – 40, 2005.
- MARR, D. **A Computational Investigation into the Human Representation and Processing of Visual Information**. New York, NY, USA: Henry Holt and Co., Inc., 1982.
- MELGANI, F.; BRUZZONE, L. Classification of hyperspectral remote sensing images with support vector machines. **IEEE Transactions on Geoscience and Remote Sensing**, 2004. v. 42, n. 8, p. 1778–1790, 2004.
- MENG, F.; ZHAO, Y.; YUE, F. A face recognition method based on fuzzy data fusion. **IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application**, 2008. p. 351–354, 2008.
- MIKOLAJCZYK, K.; SCHMID, C. An affine invariant interest point detector. In: **Proceedings of the 7th European Conference on Computer Vision-Part I**. London, UK: Springer-Verlag, 2002. (ECCV '02), p. 128–142. Disponível em: <<http://dl.acm.org/citation.cfm?id=645315.649184>>.
- _____. Scale & affine invariant interest point detectors. **International Journal of Computer Vision**. Kluwer Academic Publishers, Hingham, MA, USA, 2004. v. 60(1), p. 63–86, 2004.
- MIKOLAJCZYK, K.; TUYTELAARS, T.; SCHMID, C.; ZISSERMAN, A.; MATAS, J.; SCHAFFALITZKY, F.; KADIR, T.; GOOL, L. V. A comparison of affine region detectors. **Int. J. Comput. Vision**, 2005. v. 65(1/2), p. 43–72, 2005.
- MILANESE, R. **Detecting Salient regions in in Image: from biological evidence to computer implementation**. Dissertação (Mestrado) — University of Geneva, 1993.
- MOON, H.-M.; CHOI, D.; KIM, P.; PAN, S. B. Lda-based face recognition using multiple distance training face images with low user cooperation. **Consumer Electronics (ICCE)**, 2015. p. 7–8, 2015.
- N., G. G.; NAIKA, S.; DAS, P. K. Face recognition using mb-lbp and pca: A comparative study. **International Conference on Computer Communication and Informatics**, 2014. p. 1–6, 2014.

- NOWAK, E.; JURIE, F.; TRIGGS, B. Sampling strategies for bag-of-features image classification. In: **Computer Vision**. [S.l.]: 9th European Conference on Computer Vision, 2006. v. 3954, p. 490–503.
- _____. Sampling strategies for bag-of-features image classification. In: **Computer Vision - ECCV 2006**. [S.l.]: Springer Berlin Heidelberg, 2006. (Lecture Notes in Computer Science, v. 3954), p. 490–503.
- OJALA, T.; M.PIETIKÄINEN; D.HARWOOD. A comparative study of texture measures with classification based on feature distributions. **Pattern Recognition**, 1996. v. 29(1), p. 51–59, 1996.
- PAKHIRAA, M. K.; BANDYOPADHYAY, S.; MAULIK, U. Validity index for crisp and fuzzy clusters. **The Journal of the Pattern Recognition Society**, 2004. v. 37, p. 487–501, 2004.
- PEDROSA, G. V.; REZENDE, S. O.; TRAINA, A. J. M. Reducing the dimensionality of the sift descriptor and increasing its effectiveness and efficiency in image retrieval via bag-of-features. In: **WebMedia '12 Proceedings of the 18th Brazilian symposium on Multimedia and the web**. [S.l.: s.n.], 2012. p. 139–142.
- PETPON, A.; SRISU, S. Face recognition with local line binary pattern. **Fifth International Conference on Image and Graphics**, 2009. p. 533–539, 2009.
- PHILBIN, J.; CHUM, O.; ISARD, M.; SIVIC, J.; ZISSERMAN, A. Lost in quantization: Improving particular object retrieval in large scale image databases. In: **CVPR**. [S.l.: s.n.], 2008.
- PHILLIPS, P. J.; RAUSS, P. The face recognition technology (feret) program. **CTAC International Technoogy Symposium**, 1997. p. 8–20, 1997.
- PIRES, M. M. **Agrupamento Incremental e Hierárquico de Documentos**. Dissertação (Mestrado) — Universidade Federal do Rio de Janeiro, 2008.
- QI, X.; LU, Y.; CHEN, S.; LI, C.-G.; GUO, J. Spatial co-occurrence of local intensity order for face recognition. **Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on**, 2013. p. 1–6, 2013.
- RAYNER, K. Eye movements and attention in reading, scene perception, and visual search. **The Quarterly Journal of Experimental Psychology**, 2009. v. 62, n. 8, p. 1457–1506, 2009.
- RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. Imagenet large scale visual recognition challenge. **International Journal of Computer Vision (IJCV)**, 2015. v. 115, n. 3, p. 211–252, 2015.
- SALEEM, S.; BAIS, A.; SABLATNIG, R. A gradient extension of center symmetric local binary patterns for robust rgb-nir image matching. **Pattern Recognition (ICPR)**, 2014. p. 815–820, 2014.
- SCHÖLKOPF, B.; SMOLA, A. J. Learning with kernels: Support vector machines, regularization, optimization, and beyond. **Cambridge, MA, USA: MIT Press**, 2001. 2001.

- SENNA, P.; DRUMMOND, I. N.; BASTOS, G. S. Bmax: a bag of features based method for image classification. **28th SIBGRAPI Conference on Graphics, Patterns and Images**, 2015. p. 49–55, 2015.
- SERMANET, P.; EIGEN, D.; ZHANG, X.; MATHIEU, M.; FERGUS, R.; LECUN, Y. Overfeat: Integrated recognition, localization and detection using convolutional networks. **International Conference on Learning Representations (ICLR2014), CBLS**, 2014. April 2014.
- SIMONYAN, K.; VEDALDI, A.; ZISSERMAN, A. Deep fisher networks for large-scale image classification. **Advances in Neural Information Processing Systems**, 2013. p. 163–171, 2013.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **International Conference on Learning Representations (ICLR)**, 2015. 2015.
- SIVIC, J.; ZISSERMAN, A. Video google: A text retrieval approach to object matching in videos. In: **Proceedings of the Ninth IEEE International Conference on Computer Vision (ICCV'03)**. [S.l.: s.n.], 2003.
- SOLDERA, J.; BEHAINE, C. A. R.; SCHARCANSKI, J. Customized orthogonal locality preserving projections with soft-margin maximization for face recognition. **IEEE Transactions on Instrumentation and Measurement**, 2015. v. 64, n. 9, p. 2417–2426, 2015.
- STREICHER, A.; BURKHARDT, H.; FEHR, J. A bag of features approach for 3d shape retrieval. **International Symposium on Visual Computing**, 2009. 2009.
- SUN, W.; KISE, K.; CHAMPEIL, Y. Similar fragment retrieval of animations by a bag-of-features approach. In: **DAS '12 Proceedings of the 2012 10th IAPR International Workshop on Document Analysis Systems**. [S.l.: s.n.], 2012. p. 140–144.
- SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. **IEEE Conference on Computer Vision and Pattern Recognition (CVPR)**, 2015. p. 1–9, 2015.
- TALEB, I.; OUIS, M. E. A.; MAMMAR, M. O. Access control using automated face recognition: Based on the pca & lda algorithms. **ISKO-Maghreb: Concepts and Tools for knowledge Management (ISKO-Maghreb)**, 2014. p. 1–5, 2014.
- TURK, M.; PENTLAND, A. Eigenfaces for recognition. **Journal of Cognitive Neuroscience**, 1991. v. 3, n. 1, 1991.
- ULLAH, M. M.; PARIZI, S. N.; LAPTEV, I. Improving bag-of-features action recognition with non-local cues. In: LABROSSE, F.; ZWIGGELAAR, R.; LIU, Y.; TIDDEMAN, B. (Ed.). **Proceedings of the British Machine Vision Conference**. [S.l.]: BMVA Press, 2010.
- VALENZUELA, R. E. G.; SCHWARTZ, W. R.; PEDRINI, H. Dimensionality reduction through pca over sift and surf descriptors. **Cybernetic Intelligent Systems (CIS)**, 2012. p. 58–63, 2012.

- VAPNIK, V. N. The nature of statistical learning theory. **USA Springer**, 1999. 1999.
- VEDALDI, A.; FULKERSON, B. Vlfeat: an open and portable library of computer vision algorithms. **Proceeding MM'10 Proceedings of the international conference on Multimedia**, 2010. p. 1469–1472, 2010.
- VIEIRA, R. T.; CHIERICI, C. E. de O.; FERRAZ, C. T.; GONZAGA, A. Local fuzzy pattern: A new way for micro-pattern analysis. **Springer-Verlag Berlin Heidelberg. Ideal** **2012**, 2012. v. 7435, p. 602–611, 2012.
- WANG, D.; LI, X.; YANG, G.; LU, H. Robust tracking with spatial pyramid histogram. **Intelligent Visual Surveillance (IVS)**, 2011. p. 9–12, 2011.
- WANG, J.; LI, Y.; ZHANG, Y.; XIE, H.; WANG, C. Bag-of-features based classification of breast parenchymal tissue in the mammogram via jointly selecting and weighting visual words. In: **ICIG '11 Proceedings of the 2011 Sixth International Conference on Image and Graphics**. [S.l.: s.n.], 2011.
- WANG, J.; LI, Y.; ZHANG, Y.; WANG, C.; XIE, H.; CHEN, G.; GAO, X. Bag-of-features based medical image retrieval via multiple assignment and visual words weighting. **IEEE Trans. Med. Imaging**, 2011. v. 30, n. 11, p. 1996–2011, 2011.
- WANG, L.; ZHOU, L.; SHEN, C.; LIU, L.; LIU, H. A hierarchical word-merging algorithm with class separability measure. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 2014. v. 36, n. 3, p. 417–435, 2014.
- WANG, Z.; FAN, B.; WU, F. Local intensity order pattern for feature description. **International Conference on Computer Vision**, 2011. p. 603–610, 2011.
- WU, Y.; ZHENG, N.; LIU, Y.; YUAN, Z. Fine-grained and layered object recognition. **International Journal of Pattern Recognition**, 2012. v. 26, n. 2, p. 32, 2012.
- XIE, X.; BENI, G. A validity measure for fuzzy clustering. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, 1991. v. 13, p. 841–846, 1991.
- YANG, Z.; KURITA, T. Improvements to the descriptor of sift by bof approaches. **2nd IAPR Asian Conference on Pattern Recognition**, 2013. p. 95–99, 2013.
- YU, J.; QIN, Z.; WAN, T.; ZHANG, X. Feature integration analysis of bag-of-features model for image retrieval. **Neurocomputing**, 2013. v. 120, p. 355–364, 2013.
- ZABIH, R.; WOODFILL, J. Non-parametric local transforms for computing visual correspondence. **Third European Conference on Computer Vision. Springer-Verlag New York, Inc., New Jerse**, 1994. p. 151–158, 1994.
- ZEILER, M. D.; FERGUS, R. Visualizing and understanding convolutional networks. **Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland**, 2014. p. 818–833, September 2014.
- ZHANG, J.; DENG, Y.; GUO, Z.; CHEN, Y. Face recognition using part-based dense sampling local features. **Neurocomputing**, 2015. v. 184, p. 176–187, 2015.
- ZHANG, J.; MARSZALEK, M.; LAZEBNIK, S.; SCHMID, C. Local features and kernels for classification of texture and object categories: A comprehensive study. **Int. J. Comput. Vision**, 2007. v. 73, n. 2, p. 213–238, 2007.

- ZHANG, J.; YAN, Y.; LADES, M. Face recognition: Eigenfaces, elastic matching and neuralnets. **Proceedings of the IEEE**, 1997. v. 85, n. 9, p. 1423–1435, 1997.
- ZHANG, M.; SAWCHUK, A. A. A bag-of-features-based framework for human activity representation and recognition. In: **SAGAware '11 Proceedings of the 2011 international workshop on Situation activity & goal awareness**. [S.l.: s.n.], 2011. p. 51–56.
- ZHANG, S.; TIAN, Q.; HUA, G.; HUANG, Q.; LI, S. Descriptive visual words and visual phrases for image applications. In: **ACM Multimedia**. [S.l.: s.n.], 2009. p. 19–24.
- ZHAO, G.; CHEN, L.; CHEN, G.; YUAN, J. Kpb-sift: acompact local feature descriptor. **MM '10 Proceedings of the international conference on Multimedia**, 2010. p. 1175–1178, 2010.
- ZHUO, L.; CHENG, B.; ZHANG, J. A comparative study of dimensionality reduction methods for large-scale image retrieval. **Neurocomputing**, 2014. v. 141, p. 202–210, 2014.

Apêndices

APÊNDICE A

Tabelas

As Tabelas 59, 60 e 61 mostram os respectivos números das imagens da base de imagens Pascal VOC 2006 utilizadas no Experimento-3 para a geração do *Codebook*, conjunto de treinamento e teste, respectivamente.

Tabela 59: Imagens selecionadas para o *Codebook* (Pascal VOC 2006).

Ônibus: (80 imagens)	Vaca: (80 imagens)
34,120,136,149,161,187,210,245,263,344, 360,366,411,426,453,482,508,541,604,630, 692,754,798,845,867,875,883,946,959,1001, 1021,1034,1047,1052,1071,1112,1165,1184,1195,1206, 1218,1232,1248,1333,1368,1401,1452,1472,1497,1521, 1526,1668,1687,1695,1713,1838,1847,1909,1920,1932, 1942,1955,1991,2005,2070,2091,2127,2131,2184,2255, 2269,2308,2327,2347,2352,2367,2427,2448,2482,2496	19,25,32,37,40,73,74,78,83,87, 94,104,117,137,167,174,202,242,256,257, 308,388,449,557,595,623,633,678,679,750, 764,773,806,846,853,857,899,915,939,951, 975,982,985,989,1007,1027,1044,1065,1095,1128, 1132,1142,1161,1180,1197,1212,1215,1264,1370,1398, 1399,1404,1416,1436,1441,1488,1532,1552,1562,1638, 1641,1654,1676,1714,1731,1741,1757,1798,1813,1844
Cavalo: (80 imagens)	Ovelha: (80 imagens)
8,56,82,110,158,176,188,227,233,260, 271,294,326,341,374,403,424,438,466,488, 493,515,535,540,562,567,579,597,642,691, 723,739,744,789,818,858,880,885,893,925, 973,988,1012,1043,1085,1093,1111,1130,1140,1144, 1178,1179,1222,1249,1260,1266,1279,1286,1296,1352, 1364,1400,1441,1455,1467,1501,1506,1523,1559,1570, 1587,1594,1600,1648,1652,1690,1716,1756,1778,1804	69,92,96,112,113,114,140,150,189,193, 241,255,277,281,000301,309,392,430,440,468, 476,492,519,566,000568,572,583,618,635,645, 751,755,766,777,000797,819,826,828,852,894, 904,0910,916,000927,1000,1011,1019,1064,1082,1091, 1108,1150,1192,1210,1231,1238,1255,1300,1359,1381, 1433,1439,1440,1447,1464,1474,1481,1543,1548,1572, 1625,1639,1645,1662,1680,1745,1769,1771,1835,1836

Tabela 60: Imagens selecionadas para o conjunto de treinamento (Pascal VOC 2006).

Ônibus: (94 imagens)	Vaca: (126 imagens)
2510,2521,2586,2661,2694,2745,2769,4751,4656,4695,5127,5220,2787,2831,2861,2877,2915,2920,2930,5294,4638,4647,5103,5105,3001,3014,3016,3021,3052,3059,3091,5266,5285,5083,5090,3186,3218,3292,3329,3364,3371,3408,5082,3409,3444,3479,3604,3631,3635,3652,3698,3733,3771,3800,3814,3825,3833,3838,3844,3863,3877,3914,3916,3927,4019,4024,4040,4060,4139,4150,4163,4271,4276,4305,4353,4361,4403,4427,4437,4467,4490,4523,4571,4574,4596,4652,4665,4672,4728,4734,4829,4851,4857,4892,4952,4960,4989,5014,5035	1898,1927,1947,1953,1987,2047,2052,5272,5279,5251,4446,4456,2067,2108,2126,2153,2192,2201,2213,5095,5130,5155,4417,4434,2214,2234,2267,2287,2314,2325,2334,5187,5240,5245,4398,4409,2379,2388,2391,2421,2452,2552,2561,5070,5075,5093,4389,2577,2616,2650,2690,2708,2827,2907,5007,5048,5065,2910,2919,2935,2945,2972,2981,3036,4945,4940,4943,3063,3072,3116,3129,3195,3222,3286,4864,4900,4922,3288,3326,3337,3357,3378,3414,3437,4778,4819,4769,3474,3542,3589,3614,3703,3715,3738,4626,4754,4766,3762,3785,3893,3958,4028,4066,4092,4511,4587,4605,4106,4137,4140,4141,4193,4207,4215,4217,4265,4283,4294,4304,4360,4371
Cavalo: (167 imagens)	Ovelha: (171 imagens)
1814,1817,1842,1856,1873,1880,1887,1915,1940,1958,5284,1965,2026,2040,2045,2087,2101,2121,2175,2207,2245,2248,2262,2263,2294,2321,2326,2350,2353,2378,2405,2464,2469,2470,2492,2503,2531,2572,2601,2613,2624,2651,2669,2675,2677,2689,2702,2718,2722,2723,2738,2776,2797,2798,2802,2814,2821,2846,2885,2899,2900,2932,2950,2956,3027,3066,3071,3093,3097,3104,3112,3125,3142,3174,3182,3216,3227,3275,3299,3381,3421,3459,3493,3533,3557,3568,3575,3596,3602,3638,3640,3644,3658,3663,3680,3712,3718,3723,3725,3787,3802,3805,3811,3820,3855,3867,3899,3918,3934,3944,3950,3974,3987,4001,4009,4011,4032,4052,4078,4105,4113,4154,4208,4240,4257,4298,4342,4350,4396,4466,4497,4501,4542,4552,4580,4613,4675,4692,4704,4715,4722,4777,4804,4812,4822,4826,4846,4850,4873,4879,4893,4916,4923,4968,5032,5045,5060,5100,5118,5124,5130,5137,5161,5185,5250,5268,5270,5290	1849,1861,1878,1892,1906,1956,1968,1982,2004,2007,2037,2066,2071,2145,2158,2174,2202,2208,2244,2300,2306,2318,2335,2393,2400,2412,2413,2446,2456,2460,2484,2537,2553,2554,2560,2569,2580,2587,2615,2644,2652,2657,2683,2685,2699,2721,2724,2729,2766,2778,2788,2810,2823,2870,2884,2906,2922,2936,2954,3036,3037,3060,3073,3074,3110,3138,3140,3180,3196,3224,3237,3271,3285,3287,3339,3348,3351,3374,3395,3447,3519,3537,3538,3563,3584,3590,3619,3642,3645,3657,3669,3686,3710,3717,3724,3747,3752,3776,3806,3835,3849,3861,3896,3904,3919,3924,3954,3975,3988,3989,3997,4010,4053,4072,4100,4131,4157,4168,4212,4217,4226,4273,4285,4296,4365,4383,4428,4431,4451,4470,4475,4487,4522,4535,4541,4550,4554,4559,4581,4583,4589,4610,4631,4644,4655,4683,4691,4725,4775,4800,4803,4818,4869,4891,4894,4921,4971,5017,5031,5079,5114,5132,5136,5142,5157,5177,5181,5194,5248,5259

Tabela 61: Imagens selecionadas para o conjunto de teste (Pascal VOC 2006).

Ônibus: (180 imagens)	Vaca: (197 imagens)
21,100,102,143,166,175,178,195,225,286,297,339,357,373,389,413,439,446,478,509,570,594,646,700,756,769,856,872,896,924,929,986,996,1020,1039,1045,1053,1059,1115,1121,1123,1175,1211,1216,1228,1241,1377,1380,1473,1493,1567,1579,1580,1729,1735,1763,1774,1875,1879,1922,1930,1971,2001,2011,2051,2106,2151,2203,2276,2351,2361,2372,2385,2402,2445,2468,2479,2513,4919,4934,2566,2681,2750,2767,2780,2790,2869,2890,2912,2958,4833,4860,2980,2993,2997,3029,3075,3137,3161,3183,3243,3259,4720,4732,3338,3340,3365,3372,3380,3387,3419,3453,3496,3535,4701,4447,3608,3610,3641,3683,3736,3764,3807,3816,3930,3936,4654,4657,3973,4035,4079,4088,4095,4097,4123,4162,4195,4224,4534,4600,4299,4303,4327,4333,4336,4346,4354,4370,4402,4411,4457,4469,5243,5256,5258,5287,505,5167,5197,5239,5094,5110,5150,5160,1087,4974,5033,5052,5080,1316,1339,4918,4972,5211,5226,5267,1659,1682,1711,5133,1952,1959,5068,5092,5111,5112,2219,2258,2274,4979,5005,4962,4880,4752,4676,4622,4512	3,24,72,76,81,85,93,99,106,119,477,4814,4823,4870,148,170,181,183,222,254,273,323,324,363,4642,4680,4686,412,434,447,454,469,473,503,538,553,607,4551,4627,631,651,666,676,716,759,782,862,906,922,969,980,1004,1014,1079,1131,1138,1176,1226,1246,1319,1366,1382,1412,1427,1442,1460,1482,1487,1494,1535,1555,1569,1606,1637,1673,1686,1702,1751,1760,1767,1819,1820,1824,1841,1893,1912,1961,1976,2055,2082,2112,2154,2195,2197,2221,2354,2366,2397,2451,2494,2499,2547,2594,2600,2623,2660,2755,2803,2824,2841,2878,2881,2892,2902,2905,2931,2939,2967,2998,3018,3030,3042,3061,3078,3122,4042,3130,3194,3210,3257,3270,3277,3322,3328,3336,3368,3391,3420,3442,3446,3448,3527,3583,3593,3670,3709,3731,3756,3758,3786,3791,3817,3887,3894,3909,3970,4413,4423,4492,4116,4118,4148,4349,4367,4374,4179,4219,4258,2251,2279,2320,2330,4278,4521,4291,4300,4310,4731,4080,4104
Cavalo: (247 imagens)	Ovelha: (251 imagens)
5,38,59,107,139,157,165,168,184,232,5156,5224,235,239,280,287,300,305,312,362,372,408,5063,5123,5129,409,437,451,480,486,504,531,545,561,576,4890,4938,4990,587,590,603,605,609,683,699,745,748,799,5028,5054,5058,816,820,860,866,881,888,907,956,974,977,4711,4726,978,997,1037,1083,1089,1136,1146,1163,1182,1221,4806,1223,1227,1236,1242,1285,1294,1358,1413,1430,1489,4841,4221,1503,1511,1514,1536,1563,1573,1578,1633,1683,1691,4842,4147,1718,1743,1786,1797,1839,1851,1868,1876,1886,1888,4848,4087,1894,1904,1926,1935,1957,1978,1997,2019,2084,2107,4875,4048,2111,2122,2129,2194,2210,2218,2260,2270,2283,2288,4548,3947,2293,2295,2337,2357,2371,2394,2403,2467,2502,2506,4557,3905,2515,2538,2574,2579,2610,2666,2672,2678,2693,2715,4576,3878,2720,2730,2732,2749,2792,2812,2813,2816,2837,2868,4623,3842,2876,2893,2913,2942,2953,2970,2987,3064,3077,3079,4637,3808,3111,3114,3147,3162,3164,3184,3193,3262,3298,3331,4658,3755,3362,3369,3376,3397,3436,3471,3525,3531,3539,3541,4682,3711,3572,3579,3586,3606,3611,3624,3636,3647,3648,3691,4697,5247,5252,4458,4480,4499,4504,3925,4404,4408,4416,4419,3885,3848,3824,4772,4798,4345,4820,4306,4320,4699,4995,4387,4397,4252,4253,3945,3832,4152,4210,4074,4708,3990,3991,3788,3781,3727,	10,77,144,179,236,244,276,279,290,291,5121,299,302,322,348,17,475,485,510,516,558,5193,5222,5001,5002,569,575,606,643,649,694,711,713,720,730,5254,743,768,791,801,808,817,848,851,863,909,5286,5301,5024,4838,945,960,1018,1026,1035,1040,1061,1088,1094,1118,5056,4982,1141,1159,1199,1230,1268,1274,1275,1292,1312,1320,5081,1336,1363,1411,1421,1449,1470,1475,1490,1558,1571,5085,1601,1631,1667,1706,1742,1759,1784,1818,1826,1837,5089,1852,1891,1905,1924,2006,2010,2020,2034,2059,2086,5099,2097,2110,2136,2137,2147,2150,2178,2185,2206,2239,2278,2303,2329,2333,2339,2340,2396,2406,2408,2411,2415,2455,2480,2498,2516,2252,2551,2585,2588,2595,2671,2686,2697,2706,2707,2740,2770,2775,2793,2858,2862,2882,2909,2964,2985,3019,3034,3038,3069,3108,3123,3139,3171,3190,3200,3212,3226,3241,3255,3263,4344,3267,3281,3306,3342,3349,3356,3375,3389,3439,3491,4241,3528,3546,3609,3625,3672,3708,3722,3729,3783,3789,4238,3821,3847,3860,3889,3957,3964,3966,3983,4004,4005,4012,4086,5148,4996,4685,4709,4723,4376,4378,4937,5106,4473,4484,4515,4166,4176,4213,4426,4430,4436,4533,4601,4629,4876,5283,4735,4780,4811,4668,4118,4125,4144,4379,4380,4381,4284,4292,4114,4099