



Universidade Federal de Campina Grande

Centro de Engenharia Elétrica e Informática

Curso de Graduação em Engenharia Elétrica

ARTHUR CHAVES COSTA

**AVALIAÇÃO DE EXTRATORES DE CARACTERÍSTICA EM
DIVERSAS APLICAÇÕES PARA RECONHECIMENTO DE
PADRÕES**

Campina Grande, Paraíba.

Maio de 2016

ARTHUR CHAVES COSTA

AVALIAÇÃO DE EXTRATORES DE CARACTERÍSTICA EM
DIVERSAS APLICAÇÕES PARA RECONHECIMENTO DE
PADRÕES

*Trabalho de Conclusão de Curso submetido
à Coordenação do Curso de Graduação em
Engenharia Elétrica da Universidade
Federal de Campina Grande como parte dos
requisitos necessários para a obtenção do
grau de Engenheiro Eletricista.*

Área de Concentração: Processamento Digital de Imagem

Orientadora:

Profa. Dra. Luciana Ribeiro Veloso

Campina Grande, Paraíba.

Maio de 2016

ARTHUR CHAVES COSTA

AVALIAÇÃO DE EXTRATORES DE CARACTERÍSTICA EM
DIVERSAS APLICAÇÕES PARA RECONHECIMENTO DE
PADRÕES

*Trabalho de Conclusão de Curso submetido
à Coordenação do Curso de Graduação em
Engenharia Elétrica da Universidade
Federal de Campina Grande como parte dos
requisitos necessários para a obtenção do
grau de Engenheiro Eletricista.*

Área de Concentração: Processamento Digital de Imagem

Aprovado em ____ / ____ / _____

Professor Avaliador

Universidade Federal de Campina Grande

Avaliador, UFCG

Profa. Dra. Luciana Ribeiro Veloso

Universidade Federal de Campina Grande

Orientadora, UFCG

A todos que me acompanharam durante esse ano de tantas mudanças e amadurecimento, e pelo suporte durante todos esses anos.

AGRADECIMENTOS

Agradeço a Deus, por todas as oportunidades, pela força para continuar tentando, corrigindo erros e reconstruindo meu caminho.

Agradeço a minha família, por tudo que fazem por mim, pelo sustento, apoio e por compartilharem suas alegrias e tristezas. É em vocês que se encontram as minhas raízes para poder ir em frente e sempre perceber mil motivos para agradecer.

A todos os meus amigos, que tornam o dia a dia mais leve e descontraído, fazendo com que todas as dificuldades pareçam tão pequenas e os sonhos muito maiores. Muito obrigado.

A professora Luciana Veloso, pela orientação durante todo esse tempo de pesquisa.

A todos da coordenação e secretaria de graduação de engenharia elétrica da UFCG, pela sua simpatia e pelo seu apoio.

SUMÁRIO

LISTA DE FIGURAS	v
LISTA DE TABELAS	ix
RESUMO	xi
1 INTRODUÇÃO	1
1.1 OBJETIVO	2
1.2 ESTRUTURA DO TRABALHO	2
2 REVISÃO BIBLIOGRÁFICA E FUNDAMENTAÇÃO TEÓRICA.....	3
2.1 PROCESSO DE CLASSIFICAÇÃO DE IMAGENS	3
2.2 DETECÇÃO E EXTRAÇÃO DOS DESCRITORES	6
2.2.1 HOG	8
2.2.2 SIFT.....	11
2.2.3 SURF.....	13
2.3 CLASSIFICADORES	14
3 METODOLOGIA.....	15
3.1 BASE DE IMAGENS.....	15
3.2 SISTEMA DE RECONHECIMENTO IMPLEMENTADO	17
4 RESULTADOS EXPERIMENTAIS	20
4.1 BASE DE IMAGENS DE MANUSCRITOS DOS MESES.....	20
4.1.1 DETECTOR E EXTRATOR SURF	20
4.1.2 EXTRAÇÃO POR GRADE E DESCRITOR SURF.....	23
4.1.3 HOG	29
4.1.4 SIFT.....	34
4.2 BASE DE IMAGENS CALTECH 101	36
4.2.1 DETECTOR E EXTRATOR SURF	36
4.2.2 EXTRAÇÃO POR GRADE E DESCRITOR SURF.....	40
4.2.3 HOG	44

4.2.4	SIFT.....	54
4.3	ANÁLISE COMPARATIVA	55
5	CONCLUSÃO.....	58
6	REFERÊNCIAS	59

LISTA DE FIGURAS

Figura 1. Representação do fluxo do algoritmo de classificação de imagens com pacote de <i>visual words</i> (MATHWORKS).....	4
Figura 2. Esquema de geração de histograma das ocorrências dos vocabulários em uma imagem (MATHWORKS)	5
Figura 3. Esquema do agrupamento de histogramas para cada classe do classificador (MATHWORKS)	5
Figura 4. Dois pares de imagens para identificação de correspondências (SZELISKI, 2010)...	6
Figura 5. Imagem de entrada em escala de cinza para cálculo de descritor HOG	9
Figura 6. Na esquerda: Derivada em x da imagem inicial; Na direita: Derivada em y da imagem inicial	9
Figura 7. Magnitude do gradiente calculado para a imagem inicial com a divisão em células .9	
Figura 8. Histogramas de orientação de gradiente (HOG) em cada célula da imagem.....	10
Figura 9. Esquema do vetor HOG para descrição da imagem (MATHWORKS).....	10
Figura 10. Esquema de cálculo de cada Diferença de Gaussiana por meio das Gaussianas calculadas em diferentes escalas (LOWE, 2004)	11
Figura 11. Esquema de comparação de <i>pixel</i> (marcado com 'x') com sua vizinhança para detectar se é mínimo ou máximo da Diferença de Gaussiana (LOWE, 2004).....	12
Figura 12. Esquema de descrição de ponto de interesse com histogramas de orientação do gradiente (LOWE, 2004)	12
Figura 13. Representação da aproximação da derivada parcial Gaussiana de segunda ordem por filtros em caixa (BAY <i>et al.</i> , 2006).....	13
Figura 14. Amostras da base de imagens dos meses do ano manuscritos	16
Figura 15. Amostras da base de imagens do Caltech 101 (FEI-FEI <i>et al.</i> , 2007)	17
Figura 16. Fluxograma do sistema de reconhecimento com <i>bag of words</i>	18
Figura 17. Fluxograma do sistema de reconhecimento para uso de extrator global	18
Figura 18. Pontos de interesse detectados em diferentes escalas pelo método SURF	20
Figura 19. Gráfico da maior taxa de reconhecimento dos classificadores para cada número de <i>clusters</i> (detecção e descritor SURF)	21
Figura 20. Tempo em segundos gasto para classificar uma imagem em função do número de <i>clusters</i> (detecção e descritor SURF)	22

Figura 21. Exemplo de imagem de manuscrito de janeiro com os pontos de interesse selecionados pelo método de grade 16x16 e descritor SURF	24
Figura 22. Gráfico da maior taxa de reconhecimento do classificador para cada número de <i>clusters</i> (detecção por grade 16x16 e descritor SURF)	25
Figura 23. Tempo em segundos gasto para classificar uma imagem em função do número de <i>clusters</i> (detecção por grade 16x16 e descritor SURF)	26
Figura 24. Exemplo de imagem de manuscrito de janeiro com os pontos de interesse selecionados pelo método de grade 8x8 e descritor SURF	26
Figura 25. Gráfico da maior taxa de reconhecimento do classificador para cada número de <i>clusters</i> (detecção por grade 8x8 e descritor SURF)	27
Figura 26. Tempo em segundos gasto para classificar uma imagem em função do número de <i>clusters</i> (detecção por grade 8x8 e descritor SURF)	28
Figura 27. Visualização do HOG utilizando bloco de dimensão 2x2 com diversos tamanhos de célula.....	30
Figura 28. Visualização do HOG utilizando bloco de dimensão 4x4 com diversos tamanhos de célula.....	31
Figura 29. Gráfico da taxa de reconhecimento em porcentagem do classificador em função do tamanho de célula e de bloco.....	32
Figura 30. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste	33
Figura 31. Pontos de interesse detectados pelo método SIFT	34
Figura 32. Descritores de cinco pontos de interesse detectados pelo método SIFT.....	35
Figura 33. Dez pontos de interesse detectados pelo método SURF	37
Figura 34. Gráfico da maior taxa de reconhecimento do classificador para cada número de <i>clusters</i> (detector e descritor SURF)	38
Figura 35. Tempo em segundos gasto para classificar uma imagem em função do número de <i>clusters</i> (detecção e descritor SURF)	39
Figura 36. Exemplo de histograma de frequência relativa de ocorrências de palavras do vocabulário	40
Figura 37. Exemplo de imagem com os pontos de interesse selecionados pelo método de grade 16x16 e descritor SURF	41
Figura 38. Exemplo de imagem com os pontos de interesse selecionados pelo método de grade 24x24 e descritor SURF	41

Figura 39. Exemplo de imagem com os pontos de interesse seleccionados pelo método de grade 32x32 e descritor SURF	42
Figura 40. Gráfico da maior taxa de reconhecimento do classificador para cada número de <i>clusters</i> (detecção por grade 16x16 e descritor SURF)	43
Figura 41. Tempo em segundos gasto para classificar uma imagem em função do número de <i>clusters</i> (detecção por grade 16x16 e descritor SURF)	44
Figura 42. Gráfico da taxa de reconhecimento em percentagem do classificador em função do tamanho de célula e de bloco.....	45
Figura 43. Exemplo de imagem com os dez pontos de interesse mais significativos detectados pelo SURF descritos por HOG com célula 14x14 e bloco 4x4.....	46
Figura 44. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste	47
Figura 45. Exemplo de imagem com pontos de interesse seleccionados por grade 8x8 e descritos por HOG	47
Figura 46. Gráfico da taxa de reconhecimento em percentagem do classificador em função do tamanho de célula e de bloco.....	48
Figura 47. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste	49
Figura 48. Visualização do HOG utilizando bloco de dimensão 2x2 com diversos tamanhos de célula.....	50
Figura 49. Visualização do HOG utilizando bloco de dimensão 4x4 com diversos tamanhos de célula.....	51
Figura 50. Gráfico da taxa de reconhecimento em percentagem do classificador em função do tamanho de célula e de bloco.....	52
Figura 51. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste	53
Figura 52. Pontos de interesse detectados pela função SIFT do VLFeat	54
Figura 53. Pontos de interesse detectados pela função demonstrativa disponibilizada por David Lowe	55
Figura 54. Maior taxa de reconhecimento do classificador de cada método aplicado a base de imagens dos manuscritos	56

Figura 55. Tempo em segundos para classificar uma imagem correspondente às condições de maior reconhecimento de cada método com 1050 *clusters* aplicado à base de imagens dos manuscritos.....56

Figura 56. Maior taxa de reconhecimento do classificador de cada método aplicado a base de imagens do Caltech.....57

Figura 57. Tempo em segundos para classificar uma imagem correspondente às condições de maior reconhecimento de cada método com 1050 *clusters* aplicado à base de imagens do Caltech.....57

LISTA DE TABELAS

Tabela 1. Combinações de métodos para detecção e extração de características utilizadas nos experimentos.....	19
Tabela 2. Taxas de reconhecimento do classificador gerado com detector e descritor SURF para diferentes quantidades de elementos para treino e número de clusters	21
Tabela 3. Tempos de processamento em segundos para cada etapa da geração do classificador utilizando detector e descritor SURF.....	22
Tabela 4. Exemplo de tabela de confusão para extrator e descritor SURF com 80% de partição e 650 <i>clusters</i>	23
Tabela 5. Taxas de reconhecimento do classificador gerado com detecção por grade 16x16 e descritor SURF para diferentes quantidades de elementos para treino e número de clusters ..	24
Tabela 6. Tempos de processamento em segundos para cada etapa de geração do classificador utilizando detecção por grade 16x16 e descritor SURF	25
Tabela 7. Taxas de reconhecimento do classificador gerado com detecção por grade 8x8 e descritor SURF para diferentes números de elementos para treino e número de <i>clusters</i>	27
Tabela 8. Tempos de processamento em segundos para cada etapa de geração do classificador utilizando detecção por grade 8x8 e descritor SURF	28
Tabela 9. Taxa de reconhecimento em porcentagem do classificador em relação ao tamanho de célula e bloco para o método HOG.....	32
Tabela 10. Tempo em segundos para extrair descritor e classificar uma imagem de teste em relação ao tamanho de célula e bloco para o método HOG	33
Tabela 11. Tabela confusão em porcentagem do classificador gerado com extrator HOG com tamanho de bloco 4x4 e tamanho de célula 14x14.....	34
Tabela 12. Taxas de reconhecimento do classificador gerado com extração de características pelo método SIFT para diferentes números de <i>clusters</i>	35
Tabela 13. Tempos de processamento em segundos para cada etapa de geração do classificador e teste utilizando extrator de características SIFT	36
Tabela 14. Taxas de reconhecimento do classificador gerado com extração de características pelo método SURF	37
Tabela 15. Tempos de processamento em segundos para cada etapa de geração do classificador e teste utilizando extrator de características SURF	38

Tabela 16. Tabela confusão da taxa de reconhecimento para 3050 <i>clusters</i> e 90% das imagens utilizadas para treino para método com extrator SURF	39
Tabela 17. Taxas de reconhecimento do classificador gerado com detecção de pontos de interesse por grade 16x16 e descritor SURF	42
Tabela 18. Tempos de processamento em segundos para cada etapa de geração do classificador e teste utilizando detecção por grade 16x16 e descritor SURF	43
Tabela 19. Taxa de reconhecimento do classificador em relação ao tamanho de célula e bloco para o método com detecção SURF e descrição HOG, para 1050 <i>clusters</i> e 80% das imagens para treino	45
Tabela 20. Tempo para classificar uma imagem em segundos para o método de detecção por SURF e descrição por HOG, para 1050 <i>clusters</i> e 80% das imagens para treino	46
Tabela 21. Taxa de reconhecimento do classificador em relação ao tamanho de célula e bloco para o método de detecção por grade 8x8 e HOG, para 1050 <i>clusters</i> e 80% das imagens para treino	48
Tabela 22. Tempo para classificar uma imagem em segundos para o método de detecção por grade 8x8 e descrição por HOG, para 1050 <i>clusters</i> e 80% das imagens para treino	49
Tabela 23. Taxa de reconhecimento do classificador pelo método de extração global por HOG, para 1050 <i>clusters</i> e 80% das imagens para treino	52
Tabela 24. Tempo de extração do descritor e classificação de imagem de teste pelo método de extração global por HOG, para 1050 <i>clusters</i> e 80% das imagens para treino	53
Tabela 25. Taxa de reconhecimento do classificador pelo método de extração com SIFT	55

RESUMO

Em uma aplicação de reconhecimento de padrões é necessário o treinamento de um classificador virtual com informações do que se deseja reconhecer. As ferramentas utilizadas para fornecer tais informações são conhecidas como extratores de características. Diferentes extratores de características foram desenvolvidos visando neutralizar variações na percepção da imagem, como luz, posição e tamanho. Visando testar alguns dos extratores existentes em diferentes aplicações de reconhecimento, foram desenvolvidos dois sistemas, um com treinamento do classificador por meio da criação de um vocabulário, que é utilizado para representar cada imagem, e outro sistema com o treinamento do classificador através de descritores globais. Os experimentos foram realizados com duas bases de dados, uma de manuscritos dos meses do ano e outro de objetos. O classificador foi treinado com as diferentes características extraídas pelos extratores SIFT(*Scale Invariant Feature Transform*), SURF(*Speeded Up Robust Features*) e HOG(*Histogram of Oriented Gradients*) e seu desempenho foi comparado nas duas aplicações.

Palavras-chave: Extratores de características; reconhecimento de padrões; pacote de palavras visuais.

1 INTRODUÇÃO

Atualmente a disseminação e o desenvolvimento de novas tecnologias de visão computacional e reconhecimento de objetos, da computação móvel e na Internet, assim como de novas interfaces humano computador, estão criando oportunidades nunca antes vistas para novas classes de aplicações e serviços que conectem objetos do mundo real e humanos ao mundo digital. Mais especificamente, o reconhecimento de objetos em imagens e vídeos é um tópico emergente neste cenário multidisciplinar.

Dentre as tarefas de visão computacional que o computador pode desempenhar, a análise de uma cena e o reconhecimento de todos os seus objetos constituintes é uma das mais desafiadoras. Esse processo, que é realizado pelos seres humanos com relativa facilidade, é uma tarefa computacionalmente complexa, pois um objeto tridimensional real pode ser representado através de uma grande quantidade de imagens bidimensionais conforme a sua variação de posição, orientação, tamanho, luz e contexto. O próprio sistema de lentes para a captura da imagem pode provocar distorções e também a posição que o observador se encontra. Condições adicionais como variações intraclasses, oclusões e deformações geram um número ainda maior de possibilidades, dificultando o reconhecimento (KIM; LÓPEZ, 2010), (SILVA JUNIOR, 2007).

Os trabalhos para reconhecimento e classificação de objetos em uma imagem utilizam descritores ou filtros para a extração de características, seguida de uma etapa de quantização e agrupamento para redução da dimensionalidade e, por último, uma etapa de classificação. Os descritores de característica têm como função operar sobre a imagem buscando gerar uma representação compacta das informações que são pertinentes e úteis para a classificação, ao mesmo tempo eliminando as informações irrelevantes e desnecessárias. Esses descritores devem ser preferencialmente insensíveis à iluminação e variações comuns (rotação, translação, tamanho do objeto na imagem) do objeto. Os classificadores são modelos matemáticos que buscam identificar a classe a qual a imagem analisada pertence com base nas informações fornecidas pelos vetores de característica (VASCONCELOS; GIBSON, 2014).

1.1 OBJETIVO

A finalidade deste trabalho de conclusão de curso foi o estudo e a avaliação de extratores de características no reconhecimento de padrões em imagens digitais. Inicialmente, os extratores de característica foram avaliados em um sistema de reconhecimento de palavras manuscritas dos meses do ano. Em seguida, avaliou-se o sistema em aplicação de reconhecimento de objetos utilizando a base do Caltech 101 (FEI-FEI *et al.*, 2007). Por fim, foi elaborado por meio da validação e dos testes, uma análise dos resultados obtidos.

1.2 ESTRUTURA DO TRABALHO

Este trabalho está organizado em cinco seções, além desta seção introdutória. Na Seção 2 é faz-se uma revisão bibliográfica de trabalhos significativos na área de classificação de imagens e extratores de características com a apresentação da fundamentação teórica sobre os extratores e o sistema de reconhecimento utilizado. As ferramentas e recursos utilizados para a realização dos testes, incluindo os bancos de imagem utilizados são descritos juntamente com a metodologia na Seção 3. Na Seção 4 são relatados os resultados obtidos variando-se parâmetros do processo e com o uso de diferentes extratores. Por último, a Seção 5 traz as conclusões seguidas das referências bibliográficas.

2 REVISÃO BIBLIOGRÁFICA E FUNDAMENTAÇÃO TEÓRICA

2.1 PROCESSO DE CLASSIFICAÇÃO DE IMAGENS

O processo de reconhecimento de imagens é uma tarefa importante em diversas aplicações de visão computacional tais como: vigilância, segurança, sensores inteligentes, entre outros.

As técnicas de classificação podem se basear puramente em características locais (*bag of words models*), suas posições relativas (*part-based models*), ou podem envolver a segmentação da imagem em regiões semanticamente significativas (SZELISKI, 2010). Um tema presente em todas estas técnicas é a aprendizagem.

Um dos modelos de algoritmos mais simples para o reconhecimento de objetos é o *bag of words* (*bag of features* ou *bag of keypoints*). Nesse modelo, o algoritmo calcula a distribuição (histograma) dos descritores de cada palavra do vocabulário encontradas na imagem analisada (*query image*) e compara essa distribuição com as encontradas nas imagens de treinamento. Palavras visuais são descritores (características) extraídos de vários pequenos recortes (*patches*) da imagem (SZELISKI 2010). Muitos sistemas (SIVIC, 2005; CHUM; ZISSERMAN, 2007; PHILBIN *et al.*, 2007; JÉGOU *et al.*, 2009) utilizam a abordagem *bag of words* com diferentes descritores quantizados.

O modelo de algoritmo *bag of words* é estruturado basicamente com as seguintes etapas:

- Aquisição de uma base de dados representativa dos objetos que se deseja classificar;
- Detecção e extração dos descritores das imagens pertencentes ao conjunto de treino;
- Formação das palavras visuais por meio da quantização vetorial dos descritores;
- Treinamento de um classificador;
- Avaliação do classificador com as imagens teste.

Esses passos são elaborados para maximizar a precisão do classificador enquanto se minimiza o esforço computacional. Dessa forma, os descritores extraídos não devem conter informações que são irrelevantes ao processo de classificação (transformações na imagem, variação de iluminação e oclusão), mas conter informações que são pertinentes e úteis para a classificação. Os agrupamentos de característica, também denominados de vocabulários, devem ser extensos o suficiente para distinguir mudanças relevantes nas partes de imagem, mas não tão extensos de forma a distinguir variações irrelevantes como ruído (MATHWORKS) (CSURKA *et al.*, 2004).

A Figura 1 ilustra o esquema de um sistema de reconhecimento de objetos baseado no método *bag of words*. Nessa figura são ilustradas as etapas de geração de um classificador por meio da criação de um pacote de *visual words*, ou seja, representações visuais do vocabulário gerado pelo agrupamento de informações extraídas das imagens de treino.

Figura 1. Representação do fluxo do algoritmo de classificação de imagens com pacote de *visual words* (MATHWORKS)



Esse método assume que os objetos podem ser descritos por um vetor (histograma) contendo a quantidade de cada uma das palavras visuais que a imagem possui. (SAVELONAS, 2015).

O algoritmo extrai os descritores (vetores de características) dos pontos mais discriminativos da imagem, chamados de pontos de interesse ou dos pontos de interseção de uma grade aplicada na imagem. Em seguida, essas informações são separadas em grupos de características representativas, também conhecidos por palavras visuais ou *clusters*, por meio de um algoritmo de agrupamento, que tem a finalidade de agrupar iterativamente os descritores em conjuntos mutuamente excludentes no espaço. Esse processo geralmente é realizado utilizando o algoritmo de agrupamento de dados *k-means*, que consiste em associar iterativamente cada ponto ao centro de *cluster* mais próximo e depois recomputar os centros

de *clusters*. Os grupos resultantes são compactos e separados por características similares e cada centro de um grupo corresponde às coordenadas de uma palavra do vocabulário (MATHWORKS).

Na Figura 2 é possível visualizar um esquema gráfico da construção de um vetor de características pelo histograma de ocorrências de vocabulário para uma imagem. No processo, cada descritor da imagem é rotulado com a palavra mais próxima. Em seguida, o histograma é formado com a contagem da frequência de ocorrência das palavras visuais na imagem analisada. O vetor de característica de cada imagem é o seu histograma de palavras visuais (Figura 3). O conjunto dos vetores de característica juntamente com a etiqueta que designa a classe da imagem é utilizado para treinar o classificador.

Figura 2. Esquema de geração de histograma das ocorrências dos vocabulários em uma imagem (MATHWORKS)

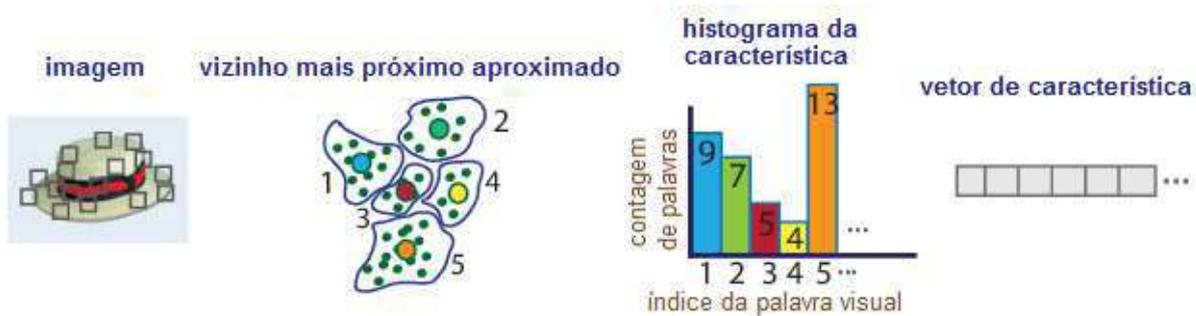


Figura 3. Esquema do agrupamento de histogramas para cada classe do classificador (MATHWORKS)



Ao término desta etapa, cada imagem passa a ser representada por um vetor contendo a quantidade de cada uma das palavras que a imagem possui. A etapa final consiste no treinamento de um classificador, para que esta aprenda a distinguir as diversas classes de imagens com base no vetor de palavras.

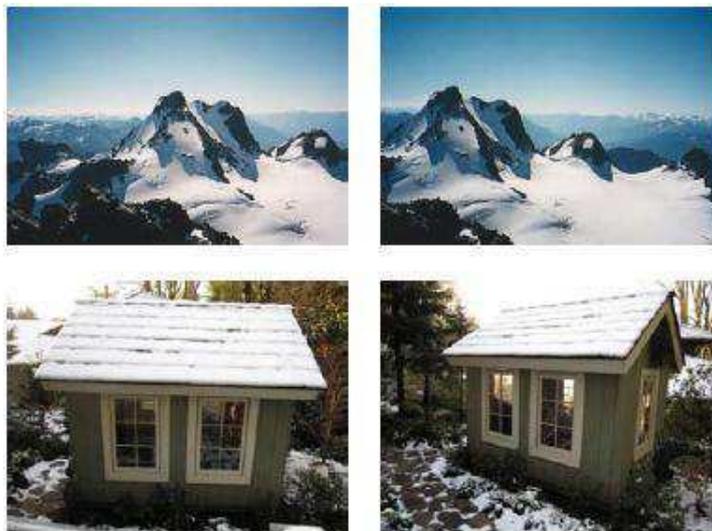
2.2 DETECÇÃO E EXTRAÇÃO DOS DESCRITORES

A maneira como se descreve uma imagem pode levar em conta uma análise geral da imagem ou de pontos importantes com grande informação discriminativa. Tais condições separam os descritores de características em globais, que analisam a imagem como um todo, e locais que descrevem regiões de interesse, com características importantes da imagem.

Detectores e descritores locais são blocos de construção de vários algoritmos em visão computacional. As aplicações incluem registro de imagem, detecção de objetos e classificação, rastreamento em vídeo e estimativa de movimento. O uso dessas informações pelos algoritmos permite que estes sejam insensíveis a escala e a rotação dos objetos, por exemplo.

De maneira a exemplificar a detecção de pontos de interesse, é possível visualizar na Figura 4 dois pares de imagem em que se deseja encontrar pontos de correspondência. Tais pontos podem ser previstos intuitivamente como os picos de montanha, ângulos da construção, portas ou pedaços com neve de formato peculiar. Esses pontos são considerados como pontos de interesse e são frequentemente descritos pela aparência dos pacotes de *pixels* que os circundam. Uma outra classe de pontos de interesse são os contornos, como a silhueta da montanha sobre o céu. Este tipo de característica pode ser usado para reconhecimento baseando-se em sua orientação e são bons indicadores de contornos de objetos e para reconhecimento em caso de oclusão (SZELISKI, 2010).

Figura 4. Dois pares de imagens para identificação de correspondências (SZELISKI, 2010).



Há duas principais abordagens para extração de pontos de característica e sua correspondência. A primeira é encontrar pontos de característica em uma imagem que podem ser precisamente rastreados utilizando uma técnica de busca local, como correlação e mínimos quadrados. A segunda é detectar características independentemente em todas as imagens em consideração e depois encontrar as correspondências baseando-se em seu aparecimento local (SZELISKI, 2010).

Regiões de imagem que não possuem textura e onde não existe grande variação de contraste são regiões que possuem uma derivada pequena e são difíceis de serem localizadas. As derivadas no sentido horizontal e vertical da imagem são obtidas convoluindo-se a imagem com uma função derivativa Gaussiana. O gradiente da imagem representa a intensidade e orientação da variação que ocorre na imagem. Quanto mais orientações de variação existirem em um ponto, mais fácil torna-se sua localização e menor é a probabilidade de ser confundido com outro ponto (SZELISKI, 2010).

De maneira geral, em um algoritmo de detecção de características, com as derivadas da imagem computadas, é realizada uma análise sobre cada ponto verificando se ele representa um mínimo ou máximo local, ou seja, se dentro de certo limite e comparando com a sua vizinhança, ele possui o maior ou o menor valor de gradiente (SZELISKI, 2010).

Após a detecção dos pontos é necessário realizar uma descrição de sua vizinhança, ou seja, extrair o descritor. Muitas técnicas diferentes descrevendo regiões de imagem local foram desenvolvidas. O descritor mais simples corresponde a um vetor de pixels da imagem, o que resulta em um vetor com alta dimensão. Com objetivo de reduzir a quantidade de informação e aprimorar a percepção de informações relevantes foram pesquisados vários métodos para simplificação desse vetor. Dentre as operações para reduzir o tamanho do vetor, foi utilizada a subamostragem, a análise do gradiente da imagem, entre outros (MIKOLAJCZYK; SCHMID, 2005).

Na maioria dos casos, porém, a aparência local das características sofre mudanças com a orientação e escala devido a transformações afins na imagem. Dessa forma, tem-se investido em descritores que compensem essas variações e tornem-se invariantes a muitas das variações da imagem (SZELISKI, 2010).

Mikolajczyk e Schmid avaliaram o desempenho de descritores locais (SIFT, GLOH, filtros direcionáveis, PCA-SIFT, correlação cruzada, gradientes de momento e contexto de

formatos) aplicados em imagens com transformações geométricas e de iluminação. Eles concluíram que para os descritores de maior dimensão, SIFT e GLOH obtiveram melhor desempenho. Para os descritores de baixa dimensão, os melhores resultados foram obtidos com os filtros direcionáveis e gradientes de momento. (MIKOLAJCZYK; SCHMID, 2005).

A seguir serão descritos alguns dos principais detectores e descritores encontrados na literatura.

2.2.1 HOG

O descritor HOG (*Histogram of Oriented Gradients*) de característica foi introduzido por Navneet Dalal e Bill Triggs (DALAL; TRIGGS, 2005), os quais desenvolveram e testaram diversas variantes do descritor, com diferentes organizações espaciais, computação de gradiente e métodos de normalização.

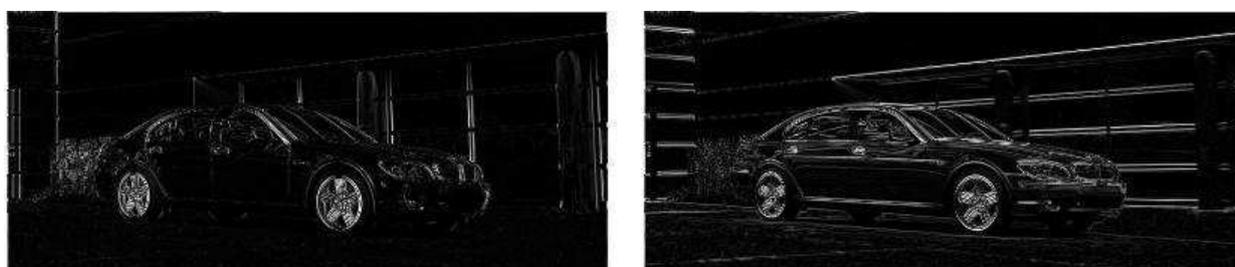
Em essência, o método descreve a aparência local do objeto e seu formato por meio da distribuição dos gradientes de intensidade ou direção de contornos. A implementação desse descritor é realizada dividindo-se a imagem em pequenas regiões conectadas, chamadas células, e para cada célula gera-se um histograma de direções do gradiente ou orientação dos contornos sobre os *pixels* contidos na região. O descritor é formado pela concatenação desses histogramas. Para um melhor desempenho, os histogramas locais podem ser normalizados em contraste através do cálculo da medida de intensidade sobre uma região maior da imagem, conhecida como bloco. Essa normalização garante melhor invariância à iluminação e sombreamento (DALAL; TRIGGS, 2005),

O gradiente é calculado com a aplicação de dois filtros derivativos, um para a direção horizontal (x) e outro para a vertical (y). Após a convolução de tais filtros, a magnitude do gradiente é obtida pela composição das duas imagens derivadas e sua orientação é determinada pelo arco tangente da componente da imagem derivada x sobre a imagem derivada y. Na Figura 5 tem-se um exemplo de imagem a ser tratada e na Figura 6 é possível visualizar a derivada no eixo x e a derivada no eixo y respectivamente.

Figura 5. Imagem de entrada em escala de cinza para cálculo de descritor HOG



Figura 6. Na esquerda: Derivada em x da imagem inicial; Na direita: Derivada em y da imagem inicial



O passo seguinte é criar os histogramas das células. Cada *pixel* da célula contribui para a definição do histograma de orientação de acordo com o gradiente computado. Os valores para orientação vão de 0° até 180° e são divididos em 9 faixas, conforme definidos pelos autores como melhores parâmetros para se conseguir melhor desempenho. Dessa forma, as orientações do gradiente de cada *pixel* são contabilizadas e armazenadas na faixa de ângulo correspondente (DALAL; TRIGGS, 2005). Na Figura 7 é apresentada a magnitude do gradiente da imagem inicial com as divisões das células realizadas e na Figura 8 é possível visualizar o histograma correspondente de cada célula dessa imagem.

Figura 7. Magnitude do gradiente calculado para a imagem inicial com a divisão em células

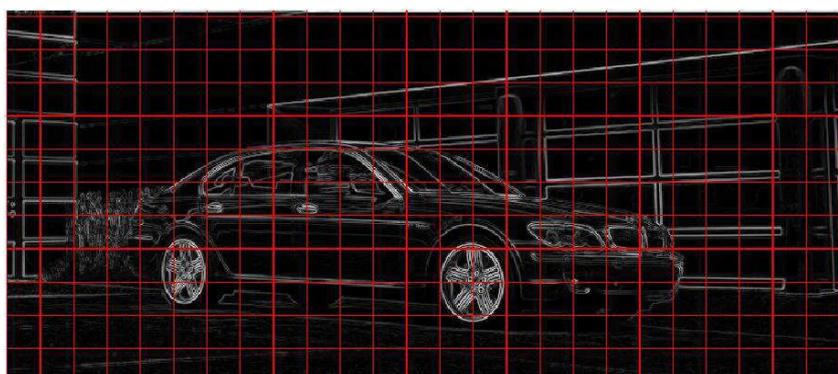
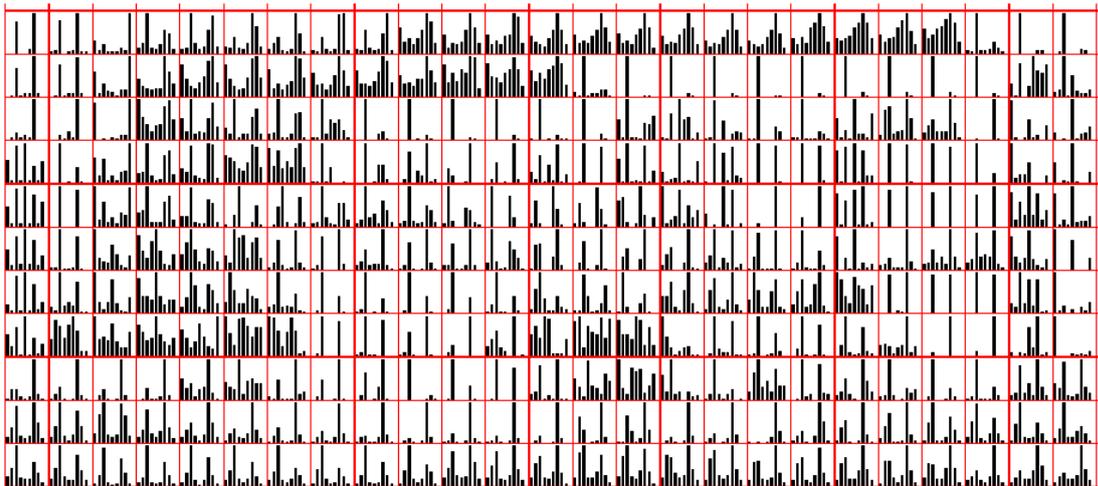
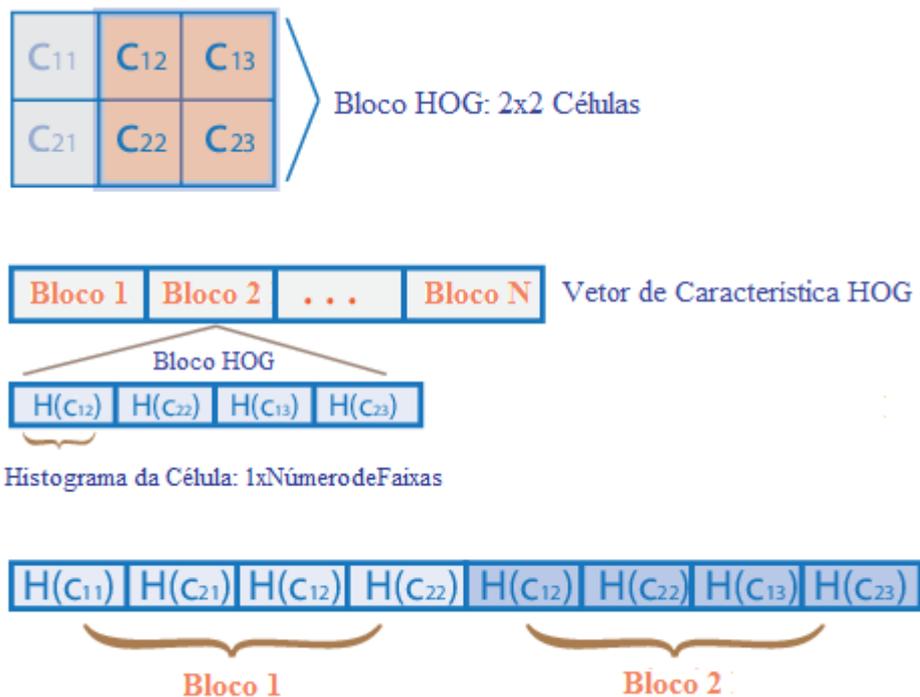


Figura 8. Histogramas de orientação de gradiente (HOG) em cada célula da imagem



Para a realização da normalização da intensidade dos gradientes é necessário agrupar as células em blocos maiores e conectados entre si, de forma a balancear a importância de cada gradiente no descritor. O descritor HOG é então o vetor das componentes dos histogramas normalizados de cada célula de todas as regiões de bloco. Uma sobreposição de blocos também é realizada para aprimorar a normalização e dessa forma as células acabam contribuindo mais de uma vez para o descritor final (MATHWORKS). Um esquema de formação do vetor é apresentado na Figura 9.

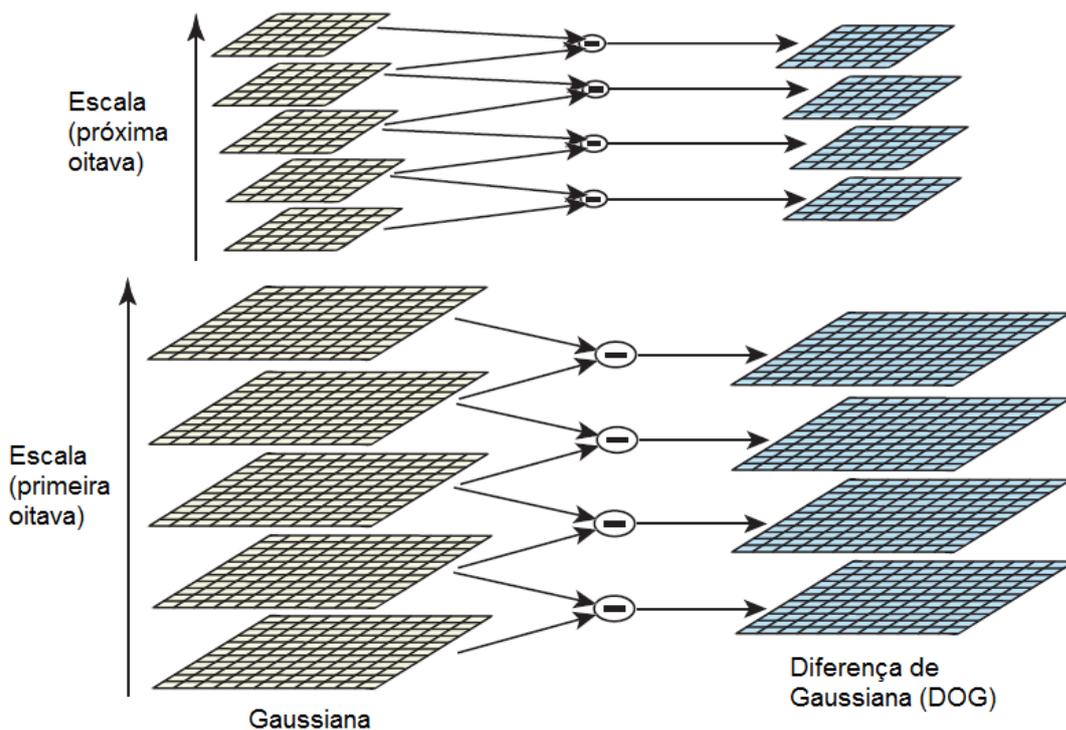
Figura 9. Esquema do vetor HOG para descrição da imagem (MATHWORKS)



2.2.2 SIFT

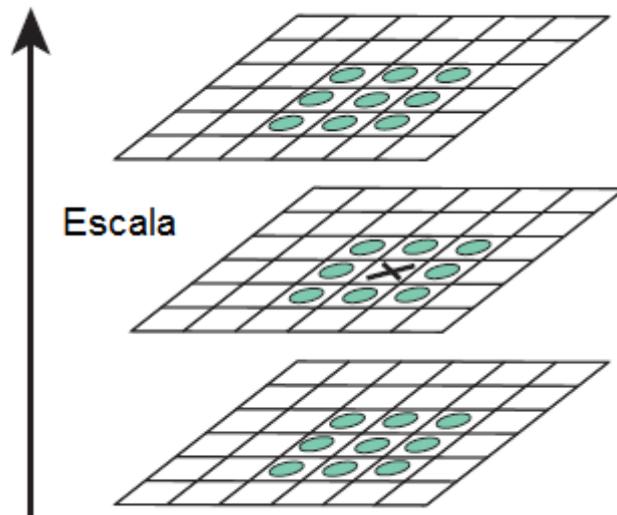
SIFT (*Scale-Invariant Feature Transform*) é uma técnica popular e eficiente de extração de características, concebida para calcular as correspondências entre pontos-chave de duas imagens independente da escala e da rotação. O processo utiliza a DoG (*Difference-of-Gaussian*) para encontrar os pontos de interesse. Constrói uma pirâmide de imagens, filtrando cada camada com Gaussianas com diferentes σ , e calculando a diferença entre cada par resultante (MORDVINTSEV, 2013). Essa diferença entra as Gaussianas reproduz uma aproximação do Laplaciano de Gaussiana. Um esquema do processo é ilustrado na Figura 10.

Figura 10. Esquema de cálculo de cada Diferença de Gaussiana por meio das Gaussianas calculadas em diferentes escalas (LOWE, 2004)



Cada candidato a ponto de interesse é analisado em relação aos seus *pixels* vizinhos da mesma escala e das escalas adjacentes, para detectar se o ponto corresponde a um máximo ou um mínimo local, conforme é visualizado no esquema da Figura 11 (LOWE, 2004).

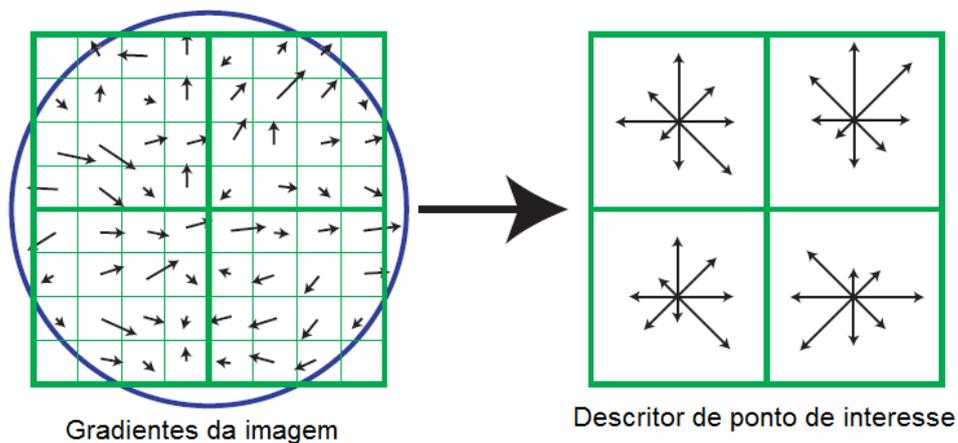
Figura 11. Esquema de comparação de *pixel* (marcado com 'x') com sua vizinhança para detectar se é mínimo ou máximo da Diferença de Gaussiana (LOWE, 2004)



Uma orientação é associada a cada ponto de interesse baseada na orientação do gradiente na região. Essa orientação garante que o descritor seja representado sempre na mesma orientação e torne o método invariante à rotação da imagem.

O descritor é criado computando-se a magnitude e a orientação do gradiente em cada ponto de amostra da imagem que se encontra na região ao redor do ponto de interesse. Essas informações são ponderadas por uma janela Gaussiana que é representada pelo círculo sobreposto na região apresentada na Figura 12. Essas amostras são acumuladas em histogramas de orientação que resumem o conteúdo de sub-regiões de 4x4 *pixels* (LOWE, 2004). Na Figura 12 o comprimento de cada seta indicando a soma das magnitudes do gradiente contidas na faixa correspondente àquela direção.

Figura 12. Esquema de descrição de ponto de interesse com histogramas de orientação do gradiente (LOWE, 2004)

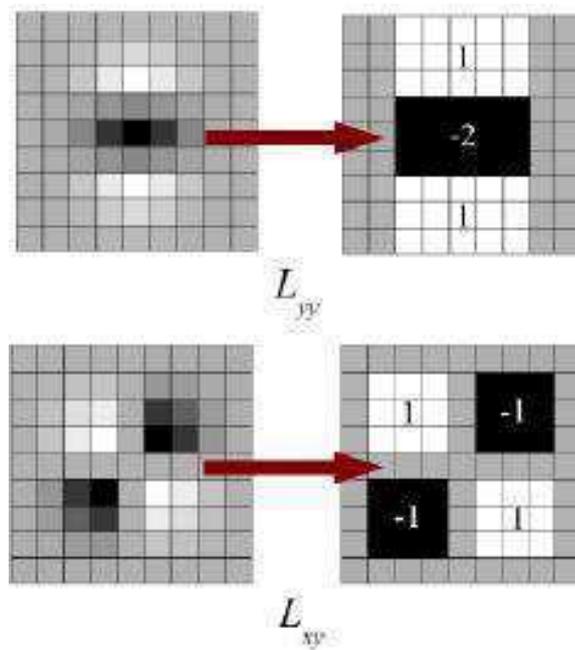


O descritor contém os valores de todos os histogramas de orientação contidos na região de interesse. Diferente do esquema apresentado na Figura 12, os histogramas são calculados de conjuntos de 4x4 sub-regiões, pois segundo experimentos realizados pelo autor do método, David Lowe, o melhor desempenho ocorre para esse caso. Como são utilizadas 8 faixas de valores para a quantização das orientações, o descritor resultante possui 4x4x8 elementos, ou seja, 128 elementos para cada ponto de interesse (LOWE, 2004).

2.2.3 SURF

SURF (*Speeded-Up Robust Features*) é uma ferramenta para extração de características que foi realizada como sendo uma versão mais veloz da SIFT. Esse método baseia-se em aproximações da matriz de Hessian, matriz de derivadas parciais de segunda ordem que pode ser decomposta em derivadas Gaussianas de segunda ordem (MORDVINTSEV, 2013). A aproximação é realizada com filtros em caixa (*box filters*) como é apresentado na Figura 13.

Figura 13. Representação da aproximação da derivada parcial Gaussiana de segunda ordem por filtros em caixa (BAY *et al.*, 2006)



Assim como no método SIFT, diferentes escalas são analisadas para a posterior identificação dos pontos como mínimos e máximos locais.

Para a associação da orientação do ponto de interesse, SURF utiliza a resposta de transformadas *wavelet* em ambas as direções para uma vizinhança de raio $6s$, onde s corresponde à escala na qual o ponto foi detectado. Uma ponderação por gaussiana é aplicada e a orientação dominante é estimada calculando-se a soma de todas as respostas para cada faixa de 60° (BAY *et al.*, 2006).

Com relação ao descritor, SURF utiliza as respostas da *wavelet* em ambas as direções. Uma vizinhança de $20s$ por $20s$ é tomada ao redor do ponto de interesse e dividida em sub-regiões de 4×4 . Cada sub-região tem as respostas da transformada *wavelet* horizontal e vertical somadas normalmente e em módulo e representadas no descritor com o total de 64 elementos (BAY *et al.*, 2006).

2.3 CLASSIFICADORES

Os classificadores são algoritmos de aprendizado de máquina que são capazes de analisar uma entrada de dados e fornecer uma classificação para ela de acordo com a etapa de treinamento.

O processo de desenvolvimento de um classificador supervisionado é realizado em duas etapas: treinamento e teste. Durante o treinamento, o vetor de característica e o rótulo do conjunto de dados de treinamento são utilizados pelo algoritmo de treinamento para adaptar o procedimento de decisão para distinção de classes (CSURKA *et al.*, 2004).

O classificador utilizado para a realização dos experimentos propostos foi o SVM (*Support Vector Machines*). Essa escolha é baseada no fato desse classificador produzir os melhores resultados em problemas com elevada dimensão (CSURKA *et al.*, 2004).

SVM's são classificadores bastante universais. Em sua forma mais básica, SVM's aprendem uma função linear para estabelecer o limiar entre as regiões que delimitam cada classe. Entretanto, pode ser aprimorado para estabelecer esse limiar de diferentes formas, usando para aprender uma fronteira de decisão polinomial, por exemplo. Uma propriedade importante é que sua habilidade de aprendizado independe da dimensionalidade do espaço onde estão distribuídas as características extraídas. SVM's calculam a complexidade de uma hipótese em uma avaliação baseando-se na margem com a qual as informações estão

separadas, e não com a quantidade de características extraídas (JOACHIMS, 1998). O algoritmo busca achar um limiar de divisão entre as classes que possua a maior margem em relação a esse limiar. Os pontos mais próximos do limiar, e que estabelecem o tamanho da margem são conhecidos como *support vectors* (vetores de suporte) (CSURKA *et al.*, 2004).

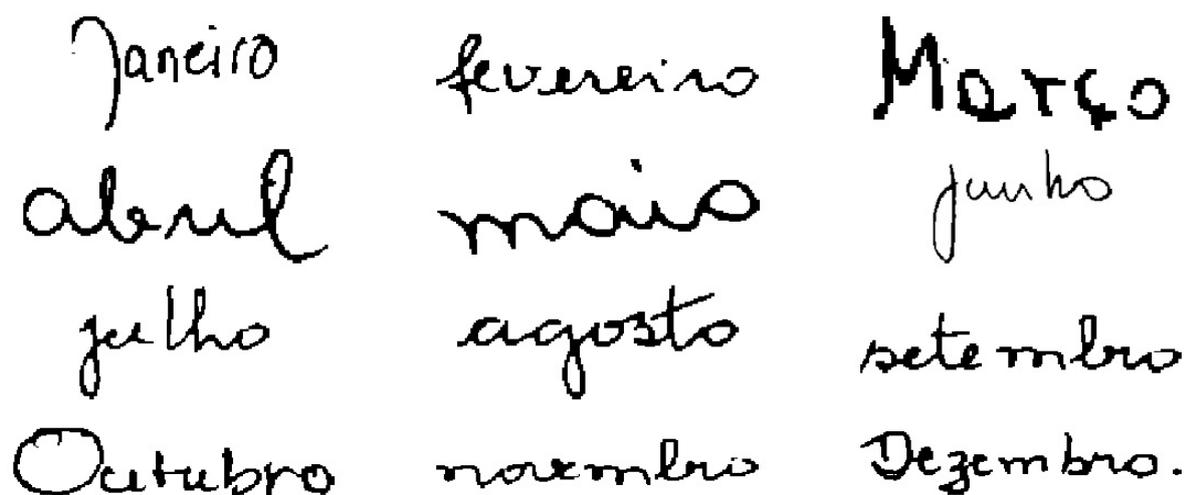
O trabalho desenvolvido por Gabriela Csurka *et al.* (CSURKA *et al.*, 2004) apresenta um método de reconhecimento de padrões utilizando quantização vetorial de descritores invariantes a transformações afins. São testados e comparados dois classificadores: Naïve Bayes e SVM (*Support Vector Machines*), para sete classes distintas. O descritor de características utilizado foi o SIFT (*Scale Invariant Feature Transform*) e os resultados foram comparados para diferentes números de *clusters* (palavras visuais). Como conclusão, seus resultados apresentaram o classificador SVM com maior taxa de reconhecimento em relação ao Naïve Bayes, e verificou-se que embora as imagens não contivessem planos de fundo uniformes, se obteve uma boa taxa de reconhecimento, atestando a robustez do sistema a essas interferências.

3 METODOLOGIA

3.1 BASE DE IMAGENS

De forma a propiciar aos extratores uma grande quantidade de informação para ser trabalhada no processo de classificação, foi utilizada uma base de imagens dos meses do ano com o total de 6000 imagens. Esta base de dados está disponível no LAPSI/DEE/UFCG, contendo 500 imagens binárias de palavras manuscritas de cada mês do ano, com resolução de 200 dpi (*dots per inch*), obtidas de 500 escritores (OLIVEIRA, 2006). Na Figura 14 é possível visualizar alguns exemplos de palavras pertencentes a essa base de imagens.

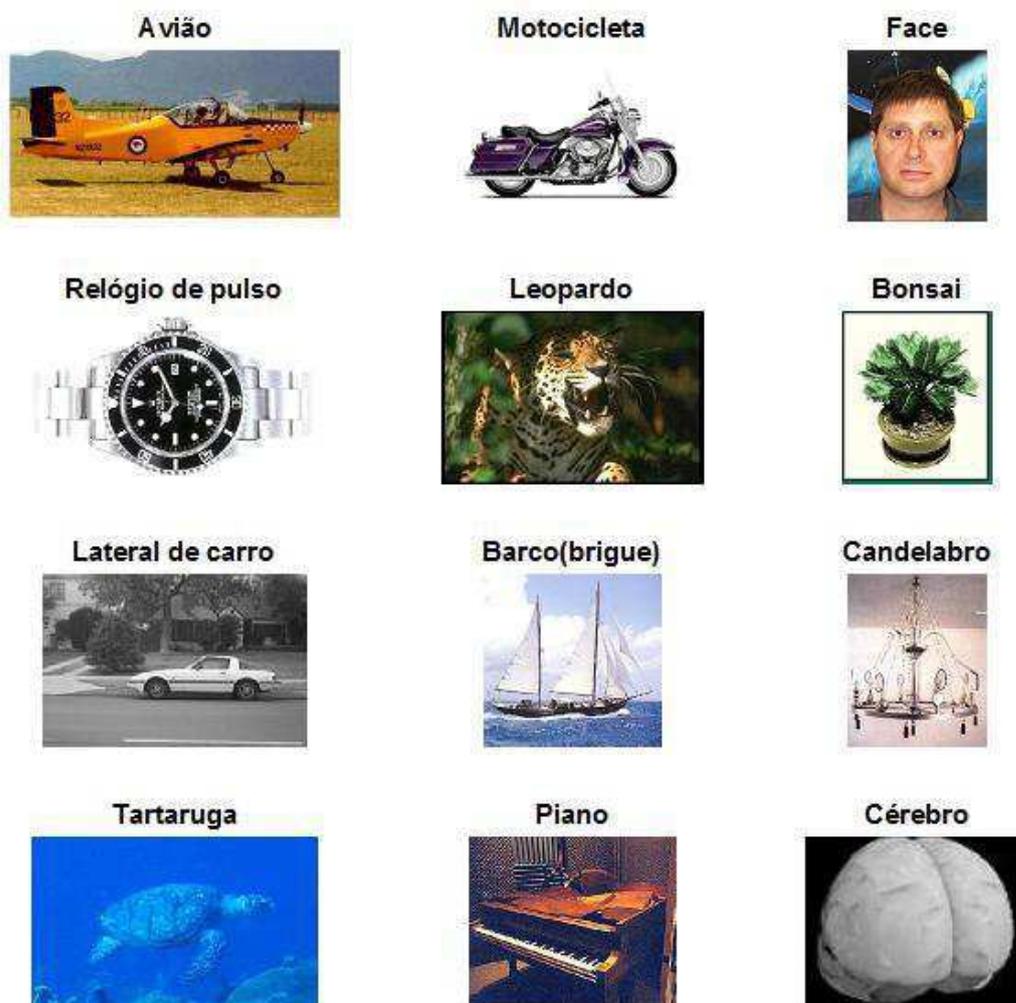
Figura 14. Amostras da base de imagens dos meses do ano manuscritos



Handwritten samples of the twelve months of the year in Portuguese, arranged in three columns. The first column contains: Janeiro, abril, julho, and Outubro. The second column contains: fevereiro, maio, agosto, and novembro. The third column contains: Março, junho, setembro, and Dezembro.

O outro banco de imagens utilizado foi o conjunto de 101 categorias de imagens disponível no portal do Instituto de Tecnologia da Califórnia, mais conhecido como Caltech (*California Institute of Technology*). Cada categoria possui aproximadamente entre 40 a 800 imagens e em média suas dimensões são de 300 x 200 *pixels*. Essas imagens foram coletadas em setembro de 2003 por Fei-Fei Li, Marco Andreetto e Marc 'Aurelio Ranzato (FEI-FEI *et al.*, 2007). Uma amostra de algumas dessas imagens que constituem o banco de dados é apresentada na Figura 15.

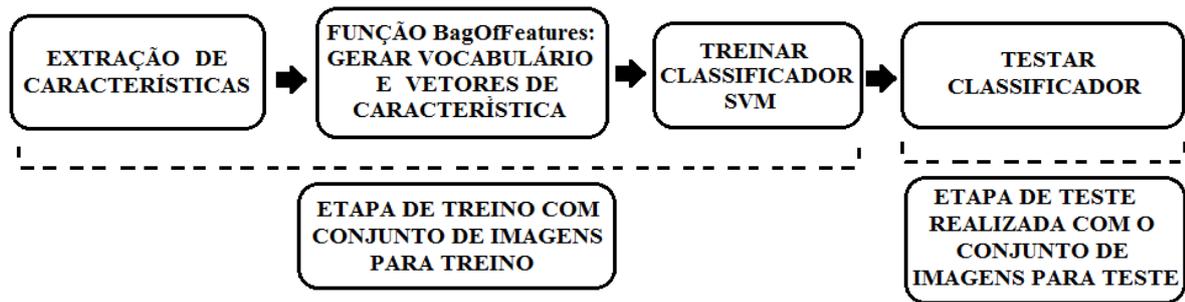
Figura 15. Amostras da base de imagens do Caltech 101 (FEI-FEI *et al.*, 2007)



3.2 SISTEMA DE RECONHECIMENTO IMPLEMENTADO

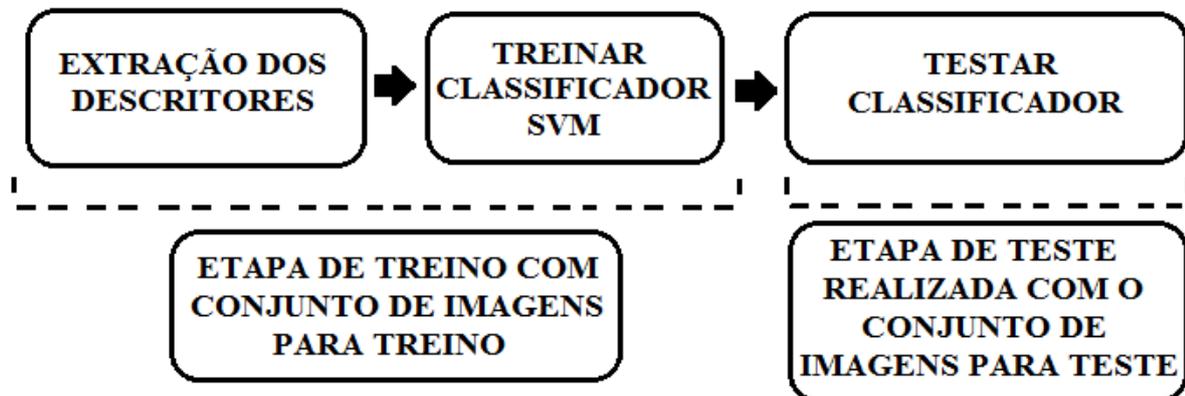
Dois sistemas de treinamento diferentes foram implementados para a realização dos testes. O primeiro método, utilizando *bag of words*, recebe as imagens do conjunto de treino para extrair suas características com descritores locais, balanceia suas quantidades para cada classe rotulada e estabelece os *clusters* no espaço que representam cada vocabulário por meio da função *bagOfFeatures*. O classificador multiclasse treinado é a SVM, com o auxílio de códigos de correção de erros: ECOC (*Error Correcting Output Codes*). O classificador treinado é então testado no conjunto de imagens teste para validar sua capacidade de reconhecimento. Na Figura 16 é apresentado um fluxograma desse método.

Figura 16. Fluxograma do sistema de reconhecimento com *bag of words*



O segundo método utilizado foi implementado para utilização de extratores globais. Em primeiro lugar é extraído um descritor de cada imagem com o seu devido rótulo. Devido à necessidade de possuir tamanho de descritores padronizados, e como nesse caso, esse tamanho varia com as dimensões da imagem, todas as imagens foram redimensionadas para um valor médio. Em seguida, todos os descritores são encaminhados para a realização do treinamento de um classificador SVM multiclasse. E por último, o classificador é avaliado em um conjunto de imagens teste. O fluxograma das etapas citadas é visualizado na Figura 17.

Figura 17. Fluxograma do sistema de reconhecimento para uso de extrator global



As bases de imagens utilizadas para os testes foram os doze conjunto de manuscritos de meses e as doze classes de objetos com mais imagens da base de dados Caltech 101. Os conjuntos de objetos utilizados foram: *avião, motocicleta, face, relógio de pulso, leopardo, bonsai, lateral de carro, barco (brigue), candelabro, tartaruga, piano e cérebro*. Todas as imagens da base Caltech 101 foram pré-processadas para níveis de cinza

Como os dois bancos de imagens diferem em suas características principais, por representarem textos e objetos, esperou-se que os diferentes tipos de extratores fornecessem

resultados distintos. A separação das imagens que foram utilizadas para o treino do classificador e para o teste foi realizada de modo sequencial para garantir que todos os testes fossem realizados com as mesmas imagens.

As funções utilizadas para detecção dos pontos de interesse, geração de descritor, formação de vocabulário estão presentes no pacote de visão computacional *Computer Vision System Toolbox*TM do MATLAB®.

Para detecção e extração de características foram utilizados diferentes algoritmos e combinações que podem ser visualizadas na Tabela 1.

Tabela 1. Combinações de métodos para detecção e extração de características utilizadas nos experimentos

Detector	Descritor
SURF	SURF
Grade	SURF
SURF	HOG
Grade	HOG
-	HOG(Global)
SIFT	SIFT

Para o método SIFT, como é um método patenteado pela *University of British Columbia* por David Lowe (LOWE, 1999), foram utilizados os códigos demonstrativos fornecidos pelo próprio David Lowe, e o código disponibilizado na VLFeat¹, uma coleção em plataforma cruzada de códigos abertos para visão computacional, com destaque para características visuais e *clusters*.

O tempo de execução do código para cada etapa do processo foi mensurado, devendo-se considerar que os testes foram realizados em máquina com processador Intel Core i5-2450M CPU 2,50GHz e 4Gb RAM.

¹ VLFeat : <http://www.vlfeat.org/index.html>

4 RESULTADOS EXPERIMENTAIS

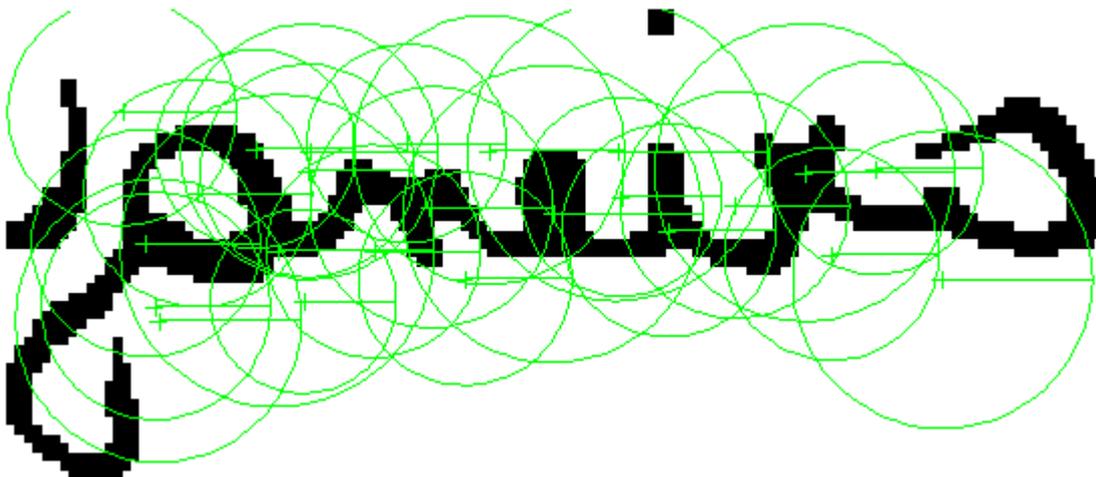
Esta seção apresenta os experimentos e resultados de desempenho dos classificadores e tempo de análise de uma imagem para os métodos citados quando aplicados às duas bases de dados. Cada teste utilizando o método *bag of words* foi realizado levando-se em consideração dois parâmetros, o número de *clusters* (palavras visuais) e a porcentagem de imagens utilizadas para do conjunto de treino. Para os testes com a descrição global com HOG, os parâmetros de bloco e célula da extração do descritor foram utilizados como parâmetros.

4.1 BASE DE IMAGENS DE MANUSCRITOS DOS MESES

4.1.1 DETECTOR E EXTRATOR SURF

Inicialmente, foram realizados testes com o detector e descritor SURF. A Figura 18 ilustra os pontos de interesse detectados em uma imagem da base de dados de manuscritos, esses pontos são indicados por uma cruz verde e a reta partindo do centro até a extremidade da região representa a orientação. Os diferentes tamanhos das regiões de interesse representam as diferentes escalas da detecção dos pontos.

Figura 18. Pontos de interesse detectados em diferentes escalas pelo método SURF

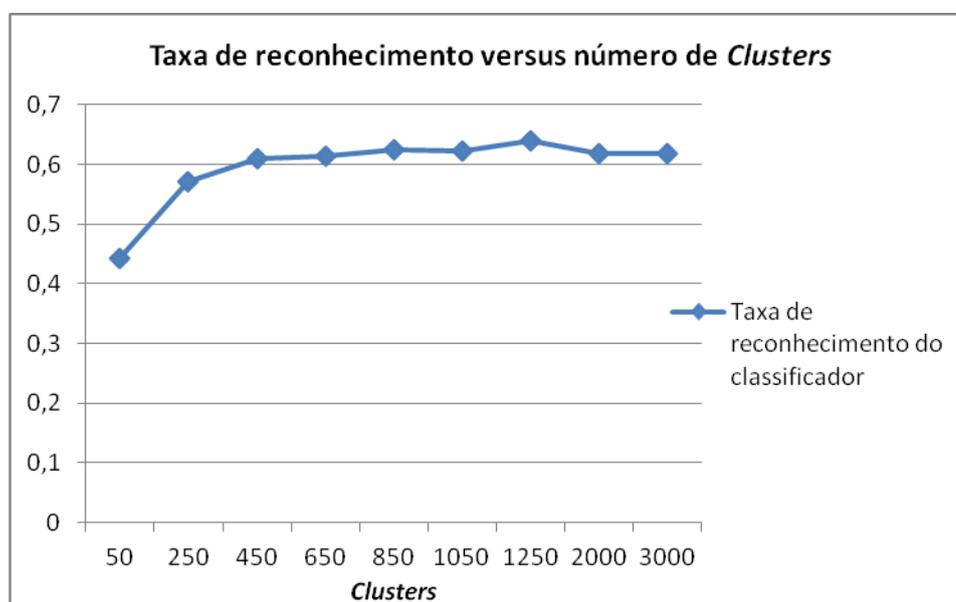


Na Tabela 2 são apresentadas as taxas de reconhecimento do classificador gerado para a identificação dos doze meses variando-se dois parâmetros, a porcentagem de imagens utilizadas para o conjunto de treino e o número de *clusters*. Na Figura 19 é possível visualizar o gráfico da maior taxa de reconhecimento dentre as partições de conjunto em função do número de *clusters*.

Tabela 2. Taxas de reconhecimento do classificador gerado com detector e descritor SURF para diferentes quantidades de elementos para treino e número de clusters

Detector e Descritor SURF		Número de Clusters								
		50	250	450	650	850	1050	1250	2000	3000
Porcentagem de imagens utilizadas para o conjunto de treino	50%	0,4233	0,5673	0,5637	0,562	0,5597	0,5803	0,5653	0,57	0,57
	55%	0,4159	0,5478	0,5848	0,577	0,5856	0,5944	0,5878	0,5652	0,5541
	60%	0,4242	0,5637	0,5888	0,6046	0,5892	0,5842	0,6012	0,5858	0,59
	65%	0,4167	0,5686	0,6024	0,5952	0,5914	0,6029	0,6076	0,6133	0,5948
	70%	0,415	0,5644	0,5839	0,6033	0,6039	0,6022	0,5961	0,5972	0,6022
	75%	0,384	0,5647	0,6107	0,594	0,624	0,5987	0,5987	0,5887	0,6073
	80%	0,4175	0,555	0,5942	0,6133	0,6008	0,6233	0,6233	0,6175	0,6183
	85%	0,4422	0,5667	0,5844	0,6144	0,6078	0,5967	0,64	0,5911	0,6133
	90%	0,4067	0,5717	0,595	0,6067	0,6117	0,6067	0,6117	0,605	0,5967

Figura 19. Gráfico da maior taxa de reconhecimento dos classificadores para cada número de clusters (detecção e descritor SURF)



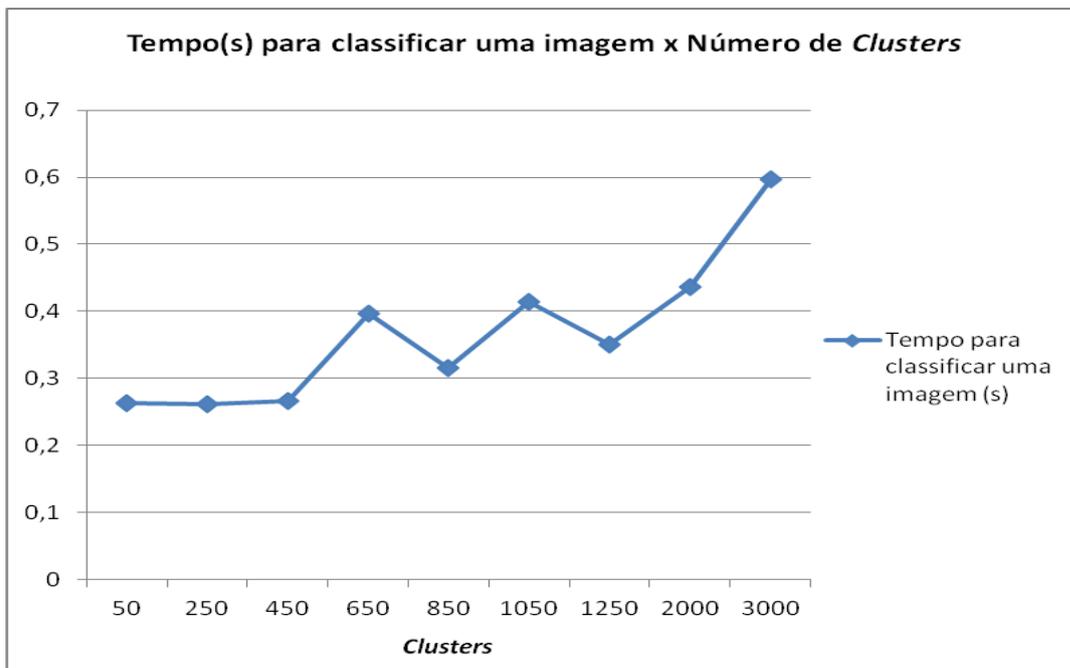
As porcentagens de divisão de imagens entre os conjuntos que obtiveram melhor resultado para a taxa de reconhecimento apresentaram uma média de valor em torno de 80%. Portanto, deste ponto em diante utiliza-se 80% do conjunto de dados para treinar e 20% para

testar, que corresponde a 400 imagens de treino para cada classe, restando 100 imagens para teste de cada classe. A Tabela 3 apresenta os tempos calculados em cada etapa, enfatizando-se o tempo de classificação de uma imagem. Na Figura 20 tem-se o tempo para classificação de uma imagem em função do número de *clusters*.

Tabela 3. Tempos de processamento em segundos para cada etapa da geração do classificador utilizando detector e descritor SURF

100 imagens no conjunto de teste		Número de Clusters								
		50	250	450	650	850	1050	1250	2000	3000
Tempos (s)	Gerar Bag Of Features	115,4	114,24	115,4	121,1	121,14	156,51	125,52	135,2	149,33
	Treinar Classificador	101,23	112,86	122,34	153,74	151,02	194,85	168,86	222,6	323,07
	Avaliar o Conj. Teste	26,339	26,195	26,679	39,715	31,613	41,488	35,019	43,611	59,647
	Tempo médio para avaliar uma imagem de teste	0,2634	0,262	0,2668	0,3972	0,3161	0,4149	0,3502	0,4361	0,5965

Figura 20. Tempo em segundos gasto para classificar uma imagem em função do número de *clusters* (detecção e descritor SURF)



Visto que o gráfico visualizado na Figura 19 se estabiliza em torno de 650 *clusters*, e que o tempo de classificação aumenta com o aumento desse parâmetro, é sensato considerar esse número de *clusters* ideal para a avaliação dos conjuntos, aliando tempo de processamento e reconhecimento como fatores de eficiência. Na Tabela 4 tem-se um exemplo de matriz de confusão para as condições indicadas, onde a taxa de reconhecimento média foi de 0,61.

Tabela 4. Exemplo de tabela de confusão para extrator e descritor SURF com 80% de partição e 650 *clusters*

Classes Verdadeiras ↓	jan.	fev.	mar.	abr.	maio	jun.	jul.	ago.	set.	out.	nov.	dez.
janeiro	0,32	0,17	0,12	0,01	0,03	0,06	0,04	0,07	0,01	0,02	0,04	0,11
fevereiro	0,02	0,71	0,03	0,02	0,03	0,03	0,02	0,01	0,06	0,01	0,03	0,03
março	0,02	0,05	0,67	0,02	0,03	0,06	0,05	0,04	0,01	0,01	0,02	0,02
abril	0	0,01	0,01	0,69	0,05	0,01	0,08	0,02	0,04	0,07	0,01	0,01
maio	0,02	0,07	0,02	0,07	0,58	0,08	0,03	0,03	0,03	0,01	0,06	0
junho	0,07	0,04	0,05	0,03	0,03	0,44	0,13	0,03	0,03	0,03	0,04	0,08
julho	0	0,01	0,02	0,1	0,01	0,07	0,73	0,01	0	0,04	0,01	0
agosto	0,01	0	0	0,02	0	0,02	0	0,83	0,03	0	0,02	0,07
setembro	0	0,05	0,01	0,04	0,04	0	0	0,01	0,55	0,09	0,1	0,11
outubro	0	0,03	0,01	0,14	0,03	0,03	0,03	0,02	0,11	0,5	0,07	0,03
novembro	0,03	0,04	0	0,04	0,06	0,01	0,03	0	0,09	0,02	0,59	0,09
dezembro	0,02	0,04	0,02	0,02	0,01	0,02	0,01	0,03	0,04	0	0,04	0,75

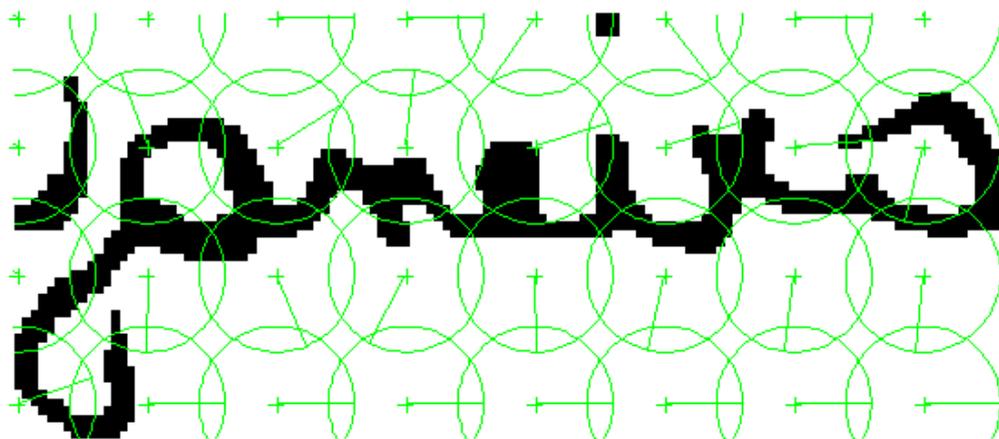
Percebe-se na tabela de confusão que 32% dos manuscritos de janeiro foram realmente classificados como tal, sendo 17% confundidos com fevereiro.

4.1.2 EXTRAÇÃO POR GRADE E DESCRITOR SURF

Nestes experimentos foi realizada a extração de pontos de interesse pelo método *grid* (grade). Os pontos nesse caso não são detectados sobre características relevantes da imagem, eles são determinados a partir de uma malha, ou grade, de dimensões estabelecidas, onde em cada ponto dessa malha é extraído um ponto de interesse.

Utilizando uma grade de 16x16 *pixels*, os pontos de interesse são recolhidos de maneira espaçada conforme apresentado na Figura 21.

Figura 21. Exemplo de imagem de manuscrito de janeiro com os pontos de interesse seleccionados pelo método de grade 16x16 e descritor SURF



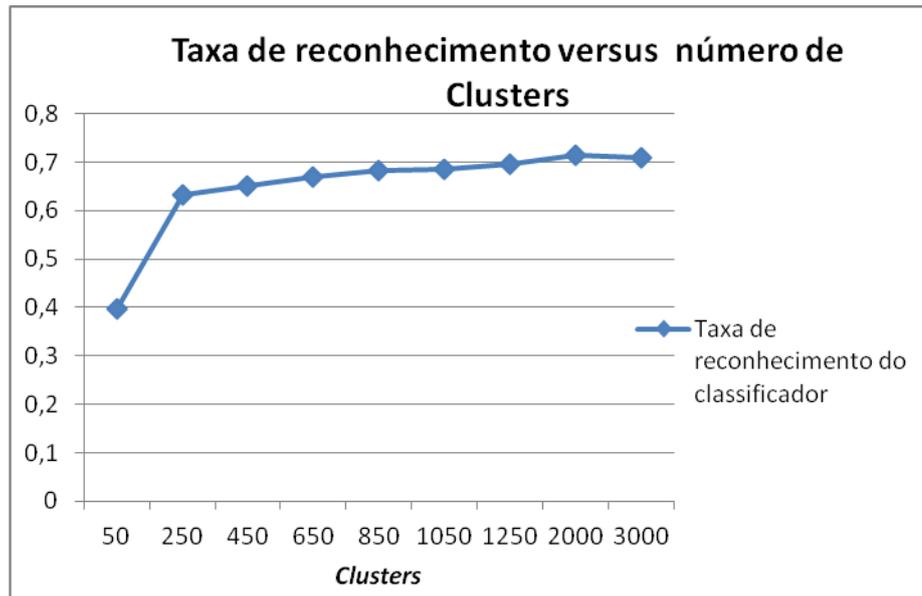
Os círculos envolta dos pontos de interesse representam a região para a extração do descritor SURF e a reta representa a orientação.

Na Tabela 5 são apresentadas as taxas de reconhecimento do classificador obtido com extração de pontos de interesse pelo método de grade com dimensões de 16x16 *pixels* e descritor SURF para diferentes números de imagem no conjunto de treinamento e número de *clusters*. Na Figura 22 é possível visualizar o gráfico da maior taxa de reconhecimento dentre as divisões de imagens para treino em função do número de *clusters*.

Tabela 5. Taxas de reconhecimento do classificador gerado com detecção por grade 16x16 e descritor SURF para diferentes quantidades de elementos para treino e número de clusters

Grade [16 16] e Descritor SURF		Número de Clusters								
		50	250	450	650	850	1050	1250	2000	3000
Porcentagem de imagens utilizadas para o conjunto de treino	50%	0,358	0,5713	0,6107	0,6367	0,6483	0,652	0,643	0,6673	0,674
	55%	0,3881	0,5926	0,6096	0,637	0,6433	0,6548	0,6707	0,6644	0,6767
	60%	0,3863	0,5775	0,6179	0,6621	0,6558	0,6721	0,6725	0,6892	0,6817
	65%	0,381	0,5924	0,6433	0,6705	0,6748	0,6776	0,6762	0,6895	0,7005
	70%	0,3733	0,6044	0,65	0,6656	0,6722	0,6844	0,6794	0,6928	0,7106
	75%	0,3447	0,6013	0,6393	0,658	0,6753	0,6813	0,6953	0,7133	0,6953
	80%	0,375	0,6317	0,645	0,67	0,6825	0,68	0,6825	0,6967	0,7058
	85%	0,3756	0,6044	0,6389	0,6522	0,68	0,6667	0,6867	0,6711	0,7089
	90%	0,3983	0,595	0,625	0,65	0,6617	0,6767	0,6817	0,7083	0,7017

Figura 22. Gráfico da maior taxa de reconhecimento do classificador para cada número de *clusters* (detecção por grade 16x16 e descritor SURF)

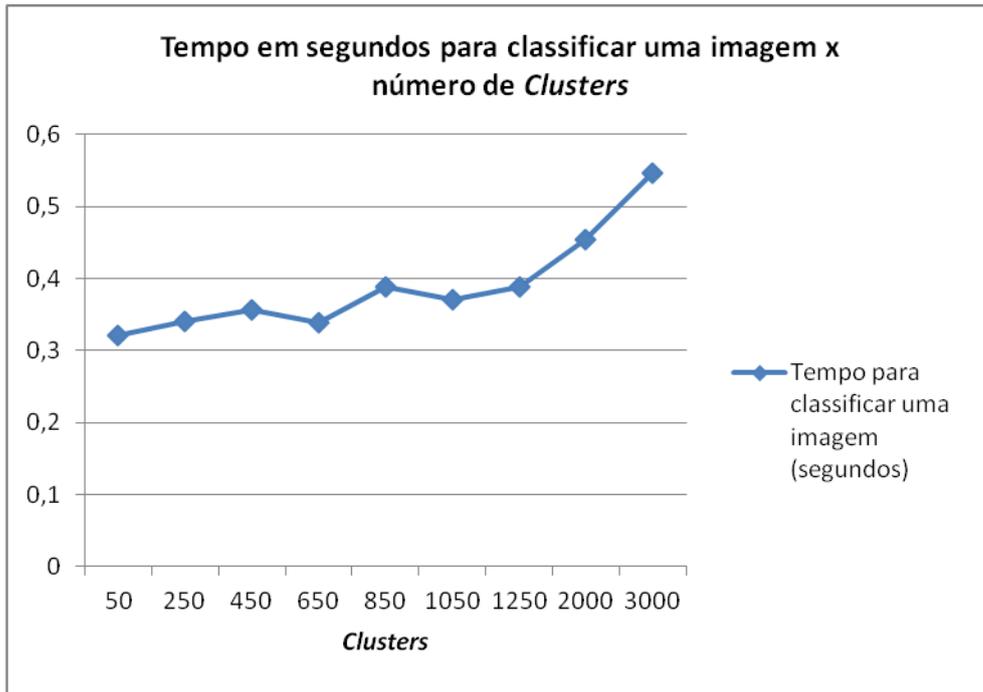


A média do parâmetro de imagens utilizadas para o conjunto de treino foi igual a 75% e seria ideal utilizar esse dado para analisar os outros parâmetros do classificador, entretanto, para fins comparativos, analisaram-se os dados para a partição de 80% utilizada no teste com o detector SURF. A Tabela 6 apresenta os tempos em segundos mensurados em cada etapa, enfatizando-se o tempo de classificação de uma imagem. Na Figura 23 tem-se o tempo para classificação de uma imagem em função do número de *clusters* para a partição das imagens em 80% para treino.

Tabela 6. Tempos de processamento em segundos para cada etapa de geração do classificador utilizando detecção por grade 16x16 e descritor SURF

100 imagens no conjunto de teste		Número de Clusters								
		50	250	450	650	850	1050	1250	2000	3000
Tempos (s)	Gerar Bag Of Features	225,12	237,02	252,75	254,39	253,44	253,14	262,22	321,47	337,38
	Treinar Classificador	123,2	144,61	157,02	156,87	183,96	168,61	180,78	222,59	282,81
	Avaliar o Conj. Teste	32,112	34,034	35,62	33,881	38,884	37,04	38,798	45,34	54,606
	Tempo médio para avaliar uma imagem de teste	0,3211	0,3403	0,3562	0,3388	0,3888	0,3704	0,388	0,4534	0,5461

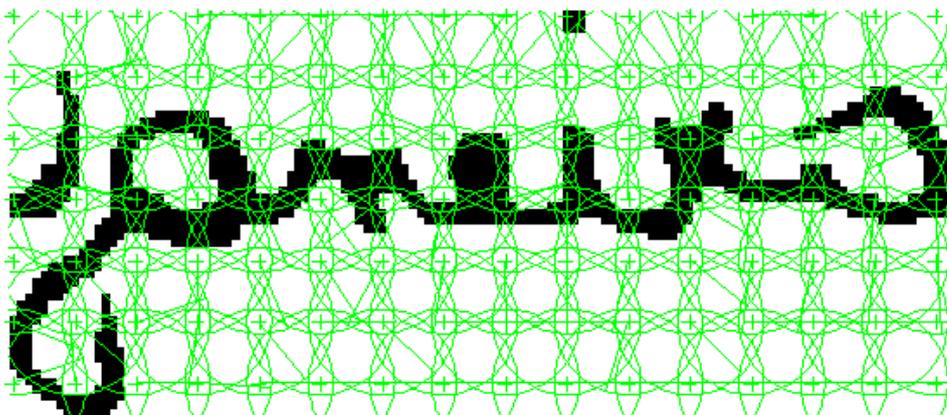
Figura 23. Tempo em segundos gasto para classificar uma imagem em função do número de *clusters* (detecção por grade 16x16 e descritor SURF)



Fixando novamente o número de *clusters* em 650, e percebendo que a curva apresentada na Figura 22 também permanece mais constante nessa faixa de valor, analisou-se os dados nessas condições, obtendo uma taxa de reconhecimento de 0,67 ou 67%.

Uma dimensão de grade menor para a detecção dos pontos de interesse foi investigada. Para tanto, utilizou-se de uma grade de espaçamento de oito *pixels* conforme é exemplificado na Figura 24.

Figura 24. Exemplo de imagem de manuscrito de janeiro com os pontos de interesse seleccionados pelo método de grade 8x8 e descritor SURF



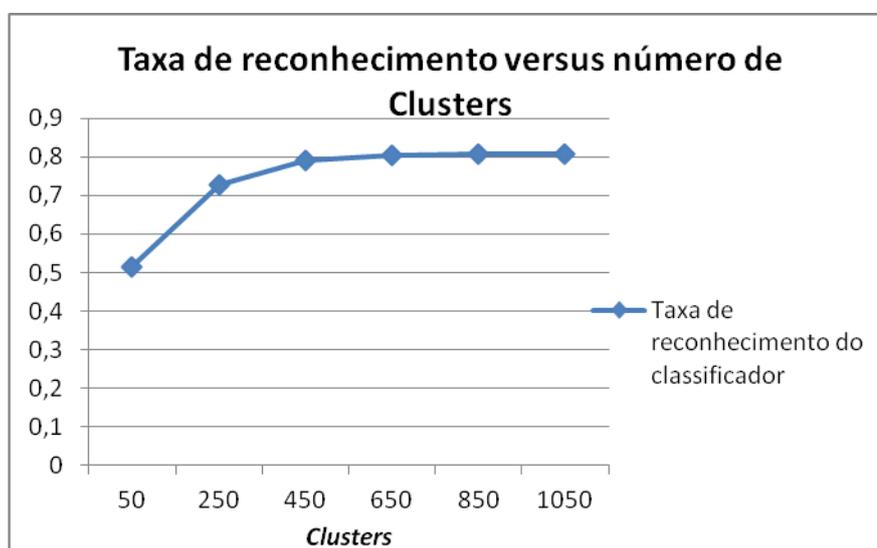
Na Tabela 7 têm-se as taxas de reconhecimento do classificador em função de diferentes números de *cluster* e da proporção das imagens utilizadas para compor o conjunto de treino.

Tabela 7. Taxas de reconhecimento do classificador gerado com detecção por grade 8x8 e descritor SURF para diferentes números de elementos para treino e número de *clusters*

Grade [8 8] e Descritor SURF		Número de Clusters					
		50	250	450	650	850	1050
Porcentagem de imagens utilizadas para o conjunto de treino	50%	0,4973	0,694	0,7377	0,7607	0,7663	0,7693
	55%	0,4926	0,6948	0,7593	0,7737	0,7763	0,7852
	60%	0,5046	0,7154	0,7608	0,7679	0,7829	0,7971
	65%	0,5052	0,7286	0,7643	0,7833	0,8005	0,8033
	70%	0,4917	0,7172	0,765	0,7894	0,7883	0,8056
	75%	0,496	0,7127	0,784	0,7853	0,7913	0,804
	80%	0,5042	0,7192	0,79	0,8025	0,8067	0,8
	85%	0,5156	0,7044	0,7411	0,7689	0,7578	0,7733
	90%	0,47	0,6983	0,735	0,7567	0,7783	0,7783

Na Figura 25 é possível visualizar o gráfico o comportamento da taxa de reconhecimento do sistema para diferentes números de *clusters*.

Figura 25. Gráfico da maior taxa de reconhecimento do classificador para cada número de *clusters* (detecção por grade 8x8 e descritor SURF)



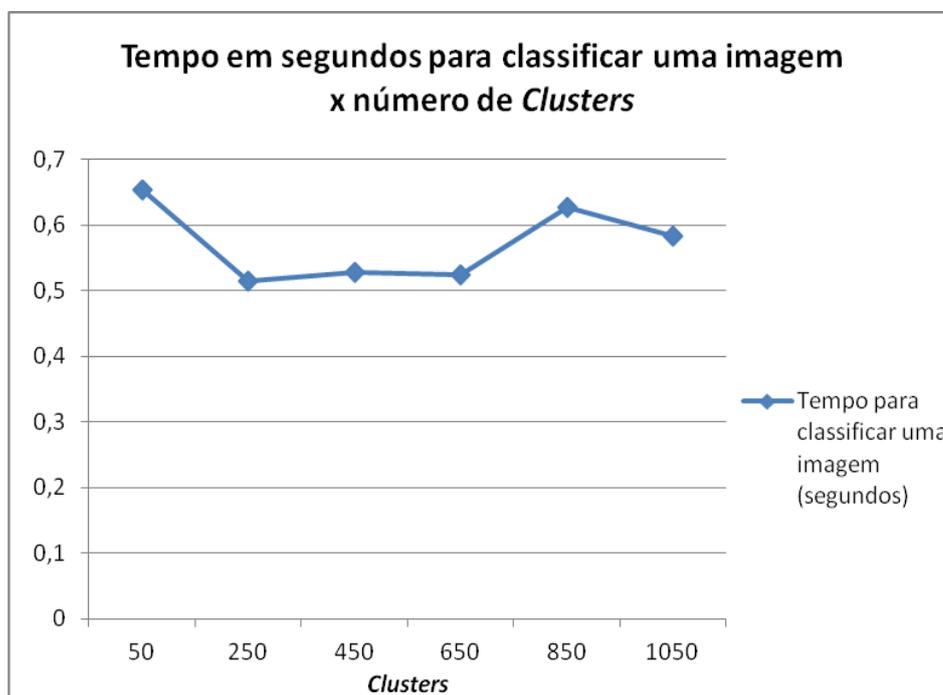
Levando-se novamente em consideração os parâmetros comparativos, na Tabela 8 são apresentados os tempos de processamento em segundos mensurados em cada etapa do classificador, enfatizando-se o tempo de classificação de uma imagem. Na Figura 26 tem-se

o tempo para classificação de uma imagem em função do número de *clusters* para a partição das imagens em 80% para treino.

Tabela 8. Tempos de processamento em segundos para cada etapa de geração do classificador utilizando detecção por grade 8x8 e descritor SURF

100 imagens no conjunto de teste		Número de Clusters					
		50	250	450	650	850	1050
Tempos (s)	Gerar Bag Of Features	691,1955	620,782	542,675	730,181	799,78	707,253
	Treinar Classificador	340,3291	204,334	209,297	219,066	255,442	244,708
	Avaliar o Conj. Teste	65,3018	51,5119	52,8568	52,519	62,7918	58,304
	Tempo médio para avaliar uma imagem de teste	0,653018	0,51512	0,52857	0,52519	0,62792	0,58304

Figura 26. Tempo em segundos gasto para classificar uma imagem em função do número de *clusters* (detecção por grade 8x8 e descritor SURF)



A taxa de reconhecimento obtida com esse método foi de 0,8025 ou 80,25% para 650 *clusters* e 80% de imagens para treino, superando consideravelmente os dois métodos anteriores.

4.1.3 HOG

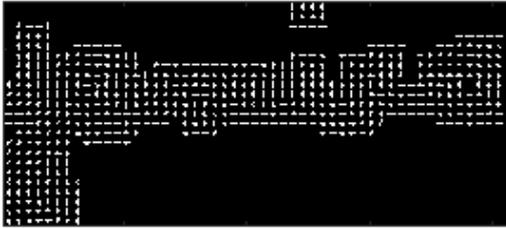
Apenas os resultados para o HOG global são apresentados nessa seção por serem os únicos resultados relevantes com esse extrator.

Como os vetores de características deviam conter um número de elementos padronizados, o tamanho das imagens foi redimensionado para uma largura e altura média, calculadas em relação à dimensão máxima e mínima das imagens. A dimensão resultante foi de 207×92 *pixels*.

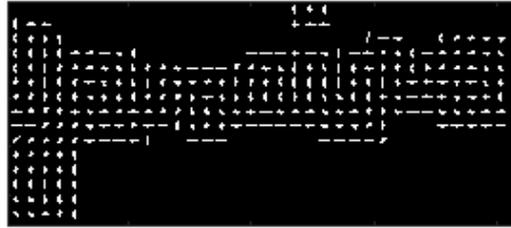
Os parâmetros ajustados para verificar a taxa de reconhecimento foram o tamanho da célula em *pixels* e o tamanho do bloco utilizado para a geração do vetor de característica. Variando esses parâmetros é possível visualizar seu efeito na quantidade de informações codificadas conforme se observa nas Figuras 27 e 28 o efeito dessa variação pode ser percebido para diversos tamanhos de célula do HOG aplicado à primeira palavra do conjunto “Janeiro” utilizando blocos de dimensão 2×2 e 4×4 respectivamente.

Figura 27. Visualização do HOG utilizando bloco de dimensão 2x2 com diversos tamanhos de célula

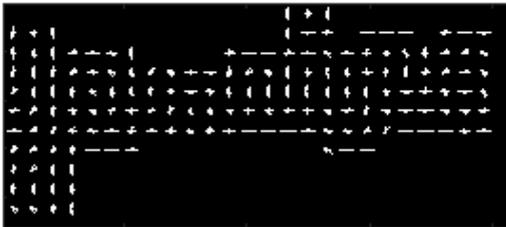
Tamanho de Célula = [4 4]
Tamanho de vetor de característica = 39600



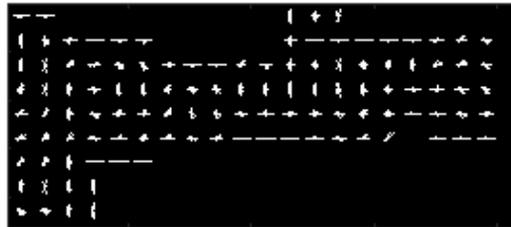
Tamanho de Célula = [6 6]
Tamanho de vetor de característica = 16632



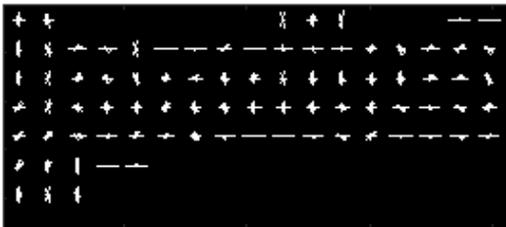
Tamanho de Célula = [8 8]
Tamanho de vetor de característica = 8640



Tamanho de Célula = [10 10]
Tamanho de vetor de característica = 5472



Tamanho de Célula = [12 12]
Tamanho de vetor de característica = 3456



Tamanho de Célula = [14 14]
Tamanho de vetor de característica = 2340

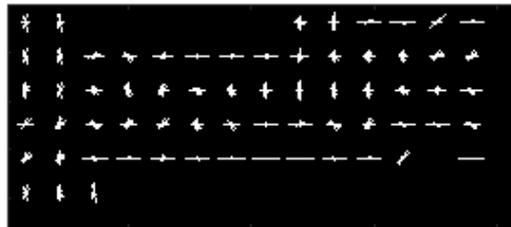
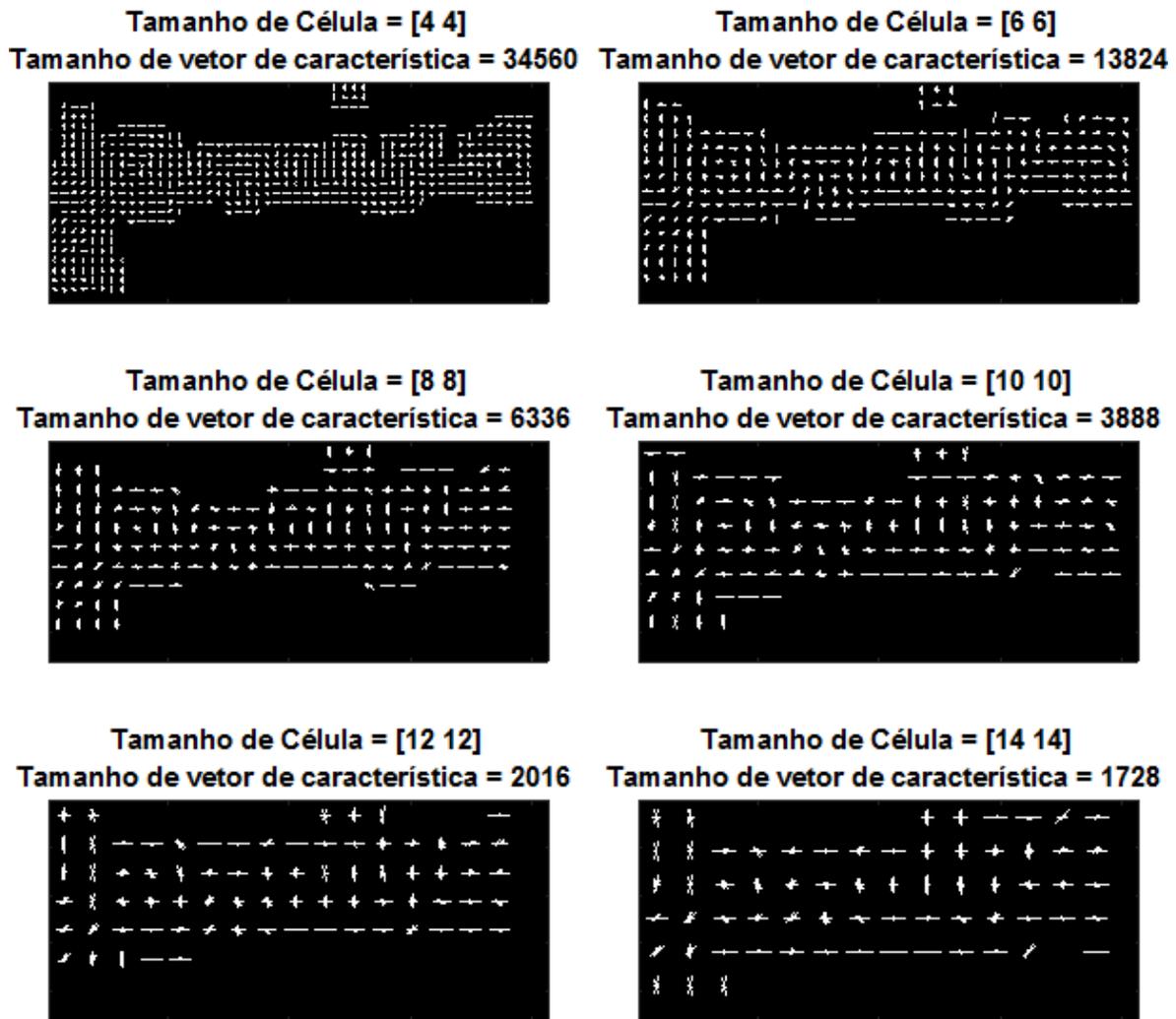


Figura 28. Visualização do HOG utilizando bloco de dimensão 4x4 com diversos tamanhos de célula



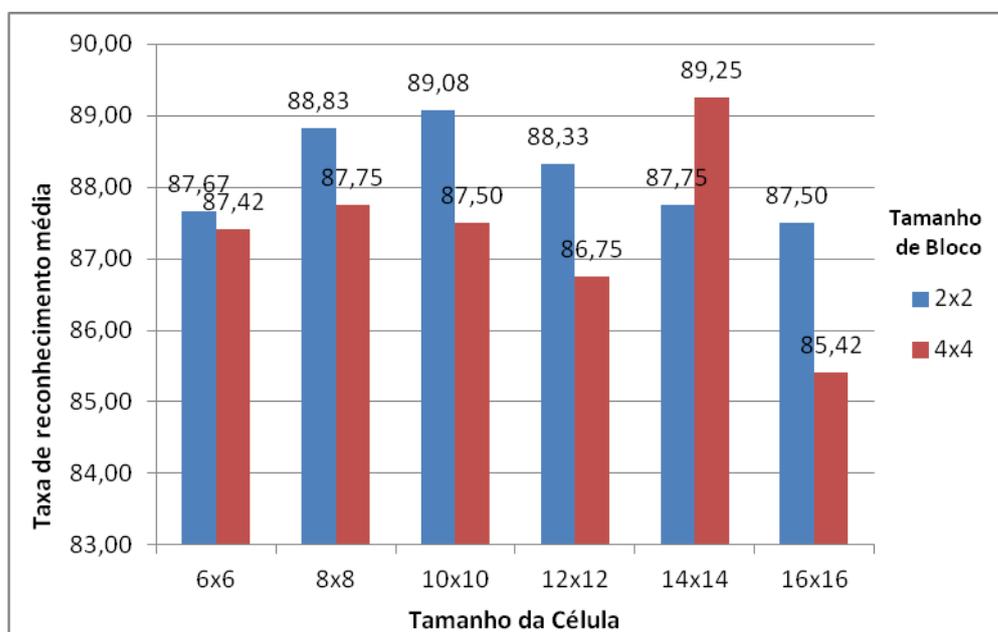
Ao analisar a Figura 28 percebemos que para um tamanho de célula maior, e sobretudo um maior tamanho do bloco, algumas informações das imagens não eram coletadas para a formação do descritor. Esse fato ocorreu porque as dimensões utilizadas para extrair as informações nem sempre coincidiam nos limites da dimensão da imagem. Acima de certa dimensão do bloco e de célula, o vetor gerado era nulo, pois as dimensões de extração de informação extrapolavam as dimensões da imagem.

Na Tabela 9 tem-se a taxa de reconhecimento média das 12 classes do classificador levando em consideração a variação da dimensão de célula e de bloco. Na Figura 29 é possível visualizar a informação da taxa de reconhecimento em gráfico.

Tabela 9. Taxa de reconhecimento em porcentagem do classificador em relação ao tamanho de célula e bloco para o método HOG

Taxa de reconhecimento média do classificador aplicado ao conjunto de teste		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	6x6	87,67	87,42
	8x8	88,83	87,75
	10x10	89,08	87,50
	12x12	88,33	86,75
	14x14	87,75	89,25
	16x16	87,50	85,42

Figura 29. Gráfico da taxa de reconhecimento em porcentagem do classificador em função do tamanho de célula e de bloco

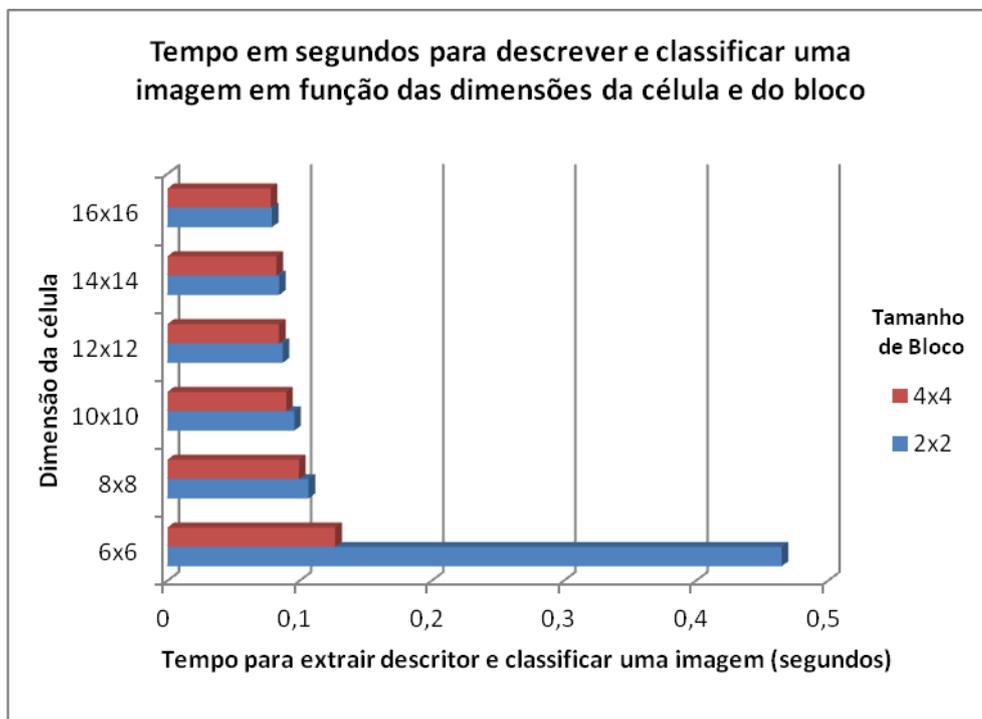


O tempo necessário para extrair o descritor de uma imagem de teste e classificá-la é apresentado na Tabela 10 e seu gráfico na Figura 30.

Tabela 10. Tempo em segundos para extrair descritor e classificar uma imagem de teste em relação ao tamanho de célula e bloco para o método HOG

Tempo médio para extrair descritor e classificar uma imagem (s)		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	6x6	0,46501	0,12711
	8x8	0,10693	0,09963
	10x10	0,09603	0,08994
	12x12	0,08727	0,08451
	14x14	0,08453	0,08246
	16x16	0,07912	0,07827

Figura 30. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste



Apesar do vetor de característica extraído com tamanho de célula de dimensão 14x14 e tamanho de bloco 2x2 conter menos informação visual da palavra em relação a outros de menor dimensão, ele superou os outros vetores utilizados para gerar o classificador. Além disso, ocupa menos espaço de memória e possui um tempo de processamento menor para a avaliação de uma imagem comparado com os outros parâmetros. Sua tabela de confusão é apresentada na Tabela 11, onde as taxas de reconhecimento estão expressas em porcentagem.

Tabela 11. Tabela confusão em porcentagem do classificador gerado com extrator HOG com tamanho de bloco 4x4 e tamanho de célula 14x14

Tamanho de célula: [14 14]; Tamanho de bloco: [4 4]		Classes Reconhecidas											
		jan.	fev.	mar.	abr.	maio	jun.	jul.	ago.	set.	out.	nov.	dez.
Classes Reais	janeiro	89	3	0	0	2	3	1	0	0	0	1	1
	fevereiro	11	86	0	0	1	1	0	0	1	0	0	0
	março	2	2	90	0	4	1	0	1	0	0	0	0
	abril	2	1	0	91	2	0	0	1	2	0	0	1
	maio	0	1	3	0	89	0	0	0	3	3	0	1
	junho	2	0	0	0	4	90	1	0	2	0	1	0
	julho	2	2	0	0	1	6	86	0	1	1	1	0
	agosto	4	1	0	2	1	1	1	89	0	0	0	1
	setembro	1	1	0	0	3	0	0	0	92	0	1	2
	outubro	0	0	1	0	2	0	0	0	2	89	3	3
	novembro	0	1	0	0	1	1	0	0	3	1	88	5
	dezembro	1	0	0	0	0	0	1	3	1	1	1	92

4.1.4 SIFT

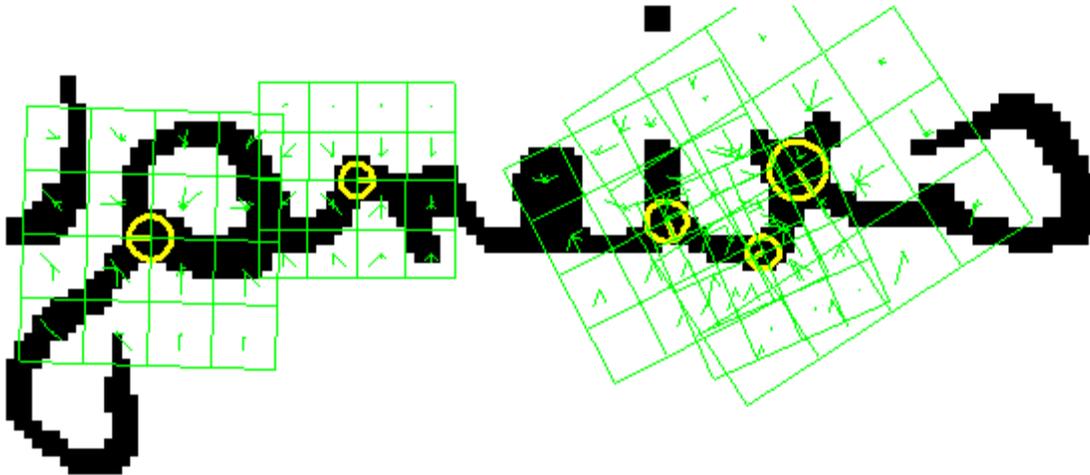
A Figura 31 apresenta os pontos de interesse localizados com o detector SIFT em uma imagem do conjunto de teste. Os traços conectando o centro até a extremidade das regiões representa a orientação.

Figura 31. Pontos de interesse detectados pelo método SIFT



Alguns dos descritores, equivalentes a cinco pontos detectados, são visualizados na Figura 32. Seus diferentes tamanhos correspondem as diferentes escalas em que são detectados.

Figura 32. Descritores de cinco pontos de interesse detectados pelo método SIFT



A Tabela 12 apresenta as taxas de reconhecimento do classificador gerado com o descritor SIFT em função do número de *clusters*, para um conjunto de teste contendo 80% das imagens totais.

Tabela 12. Taxas de reconhecimento do classificador gerado com extração de características pelo método SIFT para diferentes números de *clusters*

Porcentagem de imagens utilizadas para o conjunto de treino	Número de Clusters				
	450	1050	2050	3050	4050
80%	0,4625	0,5242	0,5408	0,545	0,5517

A Tabela 13 apresenta o tempo (segundos) necessário para realizar cada etapa do processo de treinamento e de teste, com ênfase para o tempo de processamento de uma imagem. A divisão das imagens para o conjunto de treino levou em consideração as análises anteriores.

Tabela 13. Tempos de processamento em segundos para cada etapa de geração do classificador e teste utilizando extrator de características SIFT

100 imagens no conjunto de teste		Número de Clusters				
		450	1050	2050	3050	4050
Tempos (s)	Gerar Bag Of Features	157,8717	339,2083	153,414	162,6079	185,8284
	Treinar Classificador	162,1258	187,6274	259,3792	344,0278	432,0209
	Avaliar o Conj. Teste	30,7085	78,8625	30,8455	35,9089	31,2424
	Tempo médio para avaliar uma imagem de teste	0,307085	0,788625	0,308455	0,359089	0,312424

Como a função utilizada para esse método não foi uma das funções de extração padrões do MATLAB®, o tempo de execução de cada atividade não pode ser diretamente comparado aos outros métodos.

4.2 BASE DE IMAGENS CALTECH 101

Na etapa de divisão de imagens em conjunto de treino e de teste, as 101 classes do banco de imagens Caltech foram classificadas em ordem decrescente de quantidade de imagens a fim de utilizar aquelas com mais elementos. As classes selecionadas foram normalizadas para que contivessem a mesma quantidade de imagens, restando 98 imagens para cada classe. Como dentre os conjuntos selecionados havia dois conjuntos de faces, um deles foi dispensado e adicionado o 13º maior conjunto para completar as doze classes.

É importante considerar que muitas das imagens utilizadas possuem plano de fundo não uniforme, contribuindo com o erro.

4.2.1 DETECTOR E EXTRATOR SURF

Um exemplo de dez pontos detectados pelo extrator SURF na primeira imagem da segunda classe, motocicletas, é apresentado na Figura 33. Novamente, percebe-se a diferença de tamanho das regiões, traduzindo as diferentes escalas em que foram extraídos.

Utilizando o extrator SURF para detectar e descrever as características dos objetos, obteve-se uma taxa de reconhecimento apresentada na Tabela 14, onde é possível visualizar o efeito da variação no número de *clusters* e no número de imagens do conjunto de treino.

Figura 33. Dez pontos de interesse detectados pelo método SURF

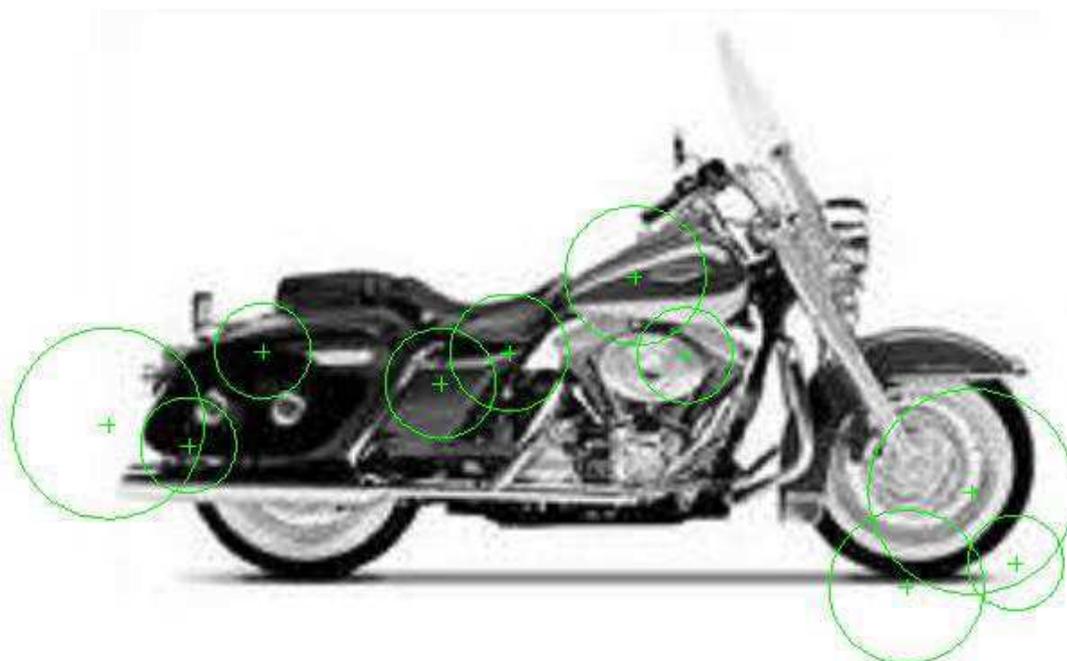


Tabela 14. Taxas de reconhecimento do classificador gerado com extração de características pelo método SURF

Detector e Descritor SURF		Número de Clusters									
		250	450	650	850	1050	1650	2050	2550	3050	4050
Porcentagem de imagens utilizadas para o conjunto de treino	40%	0,7119	0,7542	0,7415	0,7782	0,7684	0,7782	0,7768	0,7811	0,7825	0,7684
	50%	0,7415	0,7653	0,7738	0,7806	0,7585	0,801	0,7789	0,7738	0,7806	0,7908
	60%	0,7671	0,765	0,7885	0,7756	0,7778	0,7821	0,7863	0,7927	0,7671	0,7714
	70%	0,7241	0,7328	0,7644	0,7672	0,7845	0,7989	0,7989	0,773	0,8075	0,7759
	80%	0,7708	0,7708	0,7625	0,775	0,7958	0,7917	0,7708	0,8083	0,8125	0,8
	90%	0,775	0,775	0,7583	0,775	0,7917	0,8333	0,825	0,8083	0,8583	0,825

Como as maiores taxas de reconhecimento foram obtidas com 90% das imagens para o conjunto de treino, estabeleceu-se esse parâmetro para o restante das análises. O gráfico das melhores taxas de reconhecimento em relação ao número de *clusters* é apresentado na Figura 34. Os tempos de processamento em segundos para cada etapa de geração e teste do classificador para a condição citada são apresentados na Tabela 15.

Figura 34. Gráfico da maior taxa de reconhecimento do classificador para cada número de *clusters* (detector e descritor SURF)

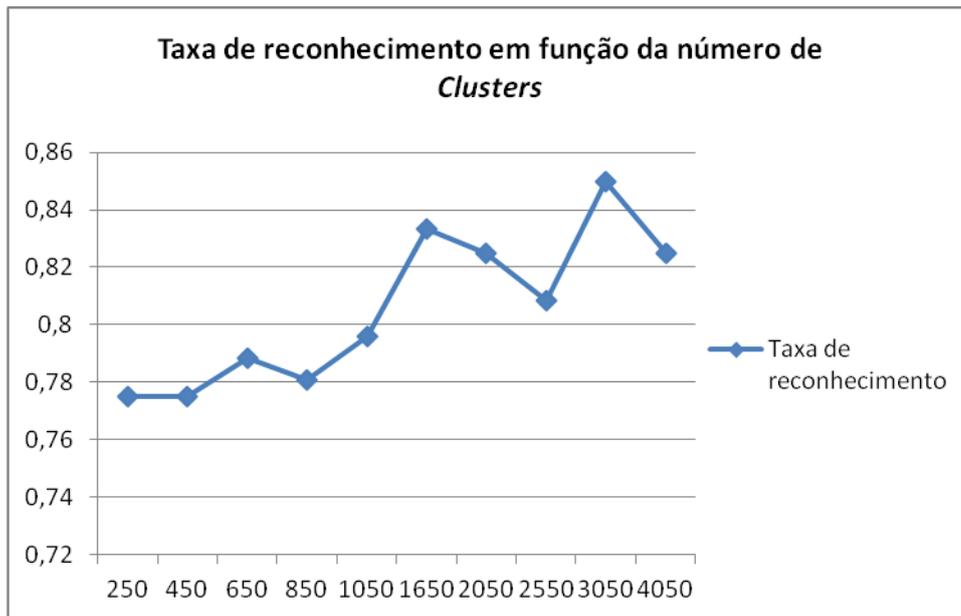
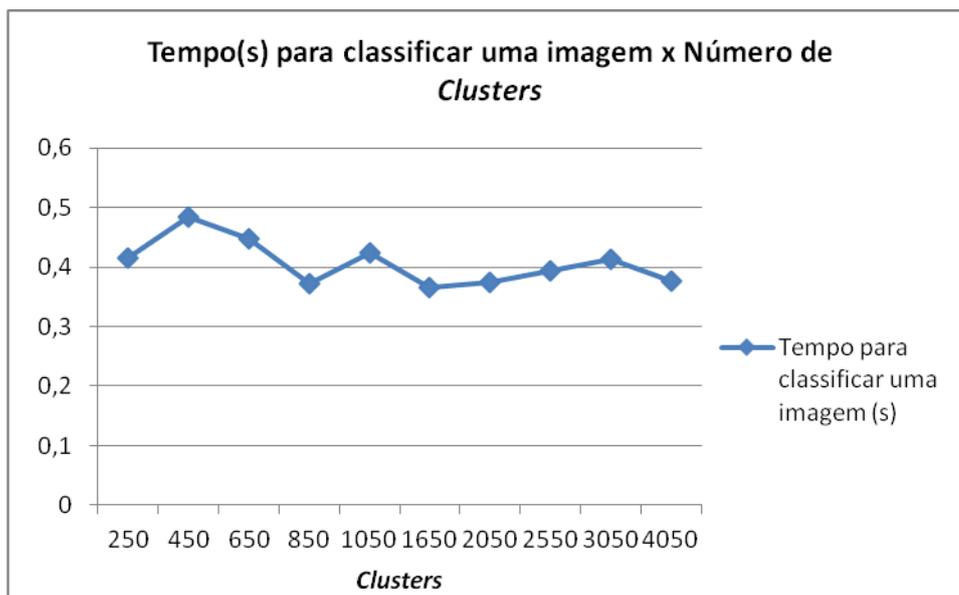


Tabela 15. Tempos de processamento em segundos para cada etapa de geração do classificador e teste utilizando extrator de características SURF

10 imagens no conjunto de treino		Número de Clusters									
		250	450	650	850	1050	1650	2050	2550	3050	4050
Tempos (s)	Gerar Bag Of Features	45,78	51,39	52,19	44,43	50,65	46,95	49,29	65,63	66,85	63,96
	Treinar Classificador	34,81	34,21	37,55	33,06	40,9	35,27	37,31	46,2	46,91	48,54
	Avaliar o Conj. Teste	4,152	4,842	4,466	3,727	4,237	3,667	3,735	3,943	4,127	3,756
	Tempo médio para avaliar uma imagem de teste	0,415	0,484	0,447	0,373	0,424	0,367	0,373	0,394	0,413	0,376

A Figura 35 representa o gráfico do tempo utilizado para classificar uma imagem em função do número de *clusters*.

Figura 35. Tempo em segundos gasto para classificar uma imagem em função do número de *clusters* (detecção e descritor SURF)



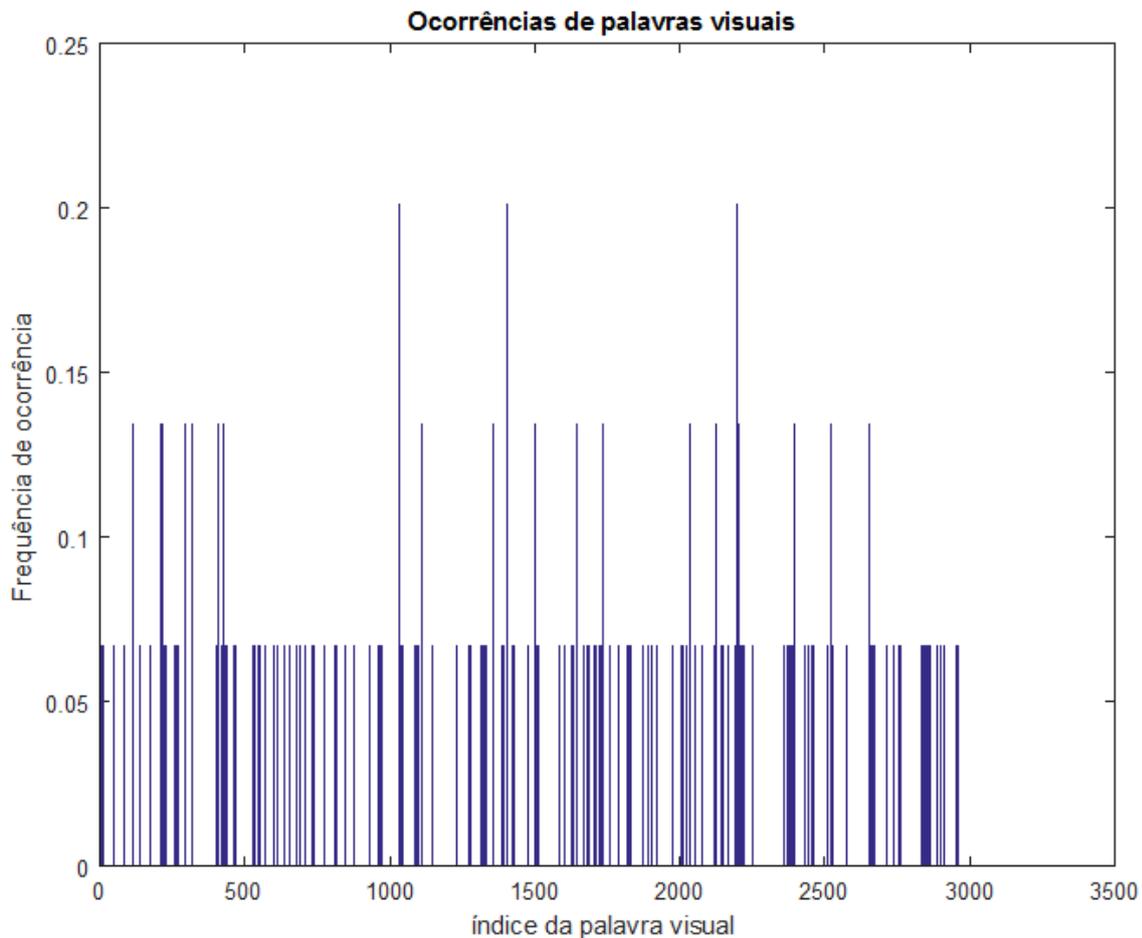
Ao analisar os resultados percebe-se que houve uma taxa maior de reconhecimento quando se utilizou 3050 *clusters* para um conjunto de treino com 90% das imagens. Sua tabela de confusão é apresentada na Tabela 16.

Tabela 16. Tabela confusão da taxa de reconhecimento para 3050 *clusters* e 90% das imagens utilizadas para treino para método com extrator SURF

Classes Verdadeiras ↓	Avião	Moto.	Face	Rel.	Leop.	Bonsai	Lat. car.	Barco	Cand.	Tart.	Piano	Cére.
Avião	0,90	0,00	0,00	0,00	0,00	0,00	0,10	0,00	0,00	0,00	0,00	0,00
Motocicleta	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Face	0,00	0,00	0,90	0,00	0,00	0,00	0,00	0,00	0,10	0,00	0,00	0,00
Relógio de pulso	0,00	0,00	0,00	0,90	0,00	0,00	0,00	0,00	0,00	0,00	0,10	0,00
Leopardo	0,00	0,00	0,00	0,00	0,60	0,00	0,00	0,00	0,00	0,40	0,00	0,00
Bonsai	0,00	0,00	0,00	0,00	0,00	0,80	0,00	0,00	0,00	0,20	0,00	0,00
Lateral de carro	0,10	0,00	0,00	0,00	0,00	0,00	0,90	0,00	0,00	0,00	0,00	0,00
Barco (brigue)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00
Candelabro	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,10	0,80	0,00	0,10	0,00
Tartaruga	0,00	0,00	0,00	0,00	0,30	0,00	0,00	0,00	0,10	0,60	0,00	0,00
Piano	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,10	0,00	0,00	0,90	0,00
Cérebro	0,00	0,00	0,10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,90

Após a criação do vocabulário de palavras visuais, cada imagem tem seu vetor de característica gerado pelo histograma da frequência relativa de ocorrências de cada palavra na imagem. Na Figura 36 é apresentado o histograma da frequência relativa de ocorrências de cada palavra do vocabulário na imagem da Figura 25.

Figura 36. Exemplo de histograma de frequência relativa de ocorrências de palavras do vocabulário



4.2.2 EXTRAÇÃO POR GRADE E DESCRITOR SURF

Foram utilizados diferentes espaçamentos de grade para a extração dos pontos para geração de descritores das imagens. As dimensões analisadas foram de 16, 24 e 32 *pixels*, e um exemplo de cada dimensão utilizada para a localização dos pontos é ilustrado na Figura 37, 38 e 39 respectivamente. Mais testes foram realizados com diferentes dimensões, entretanto apenas estas apresentaram resultados interessantes para comparação.

Figura 37. Exemplo de imagem com os pontos de interesse selecionados pelo método de grade 16x16 e descritor SURF

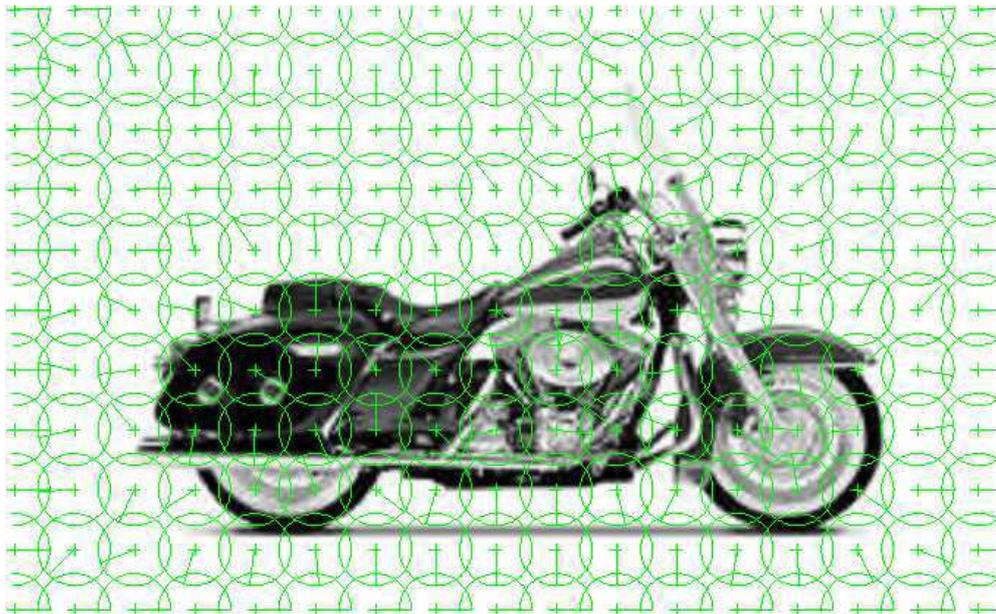


Figura 38. Exemplo de imagem com os pontos de interesse selecionados pelo método de grade 24x24 e descritor SURF

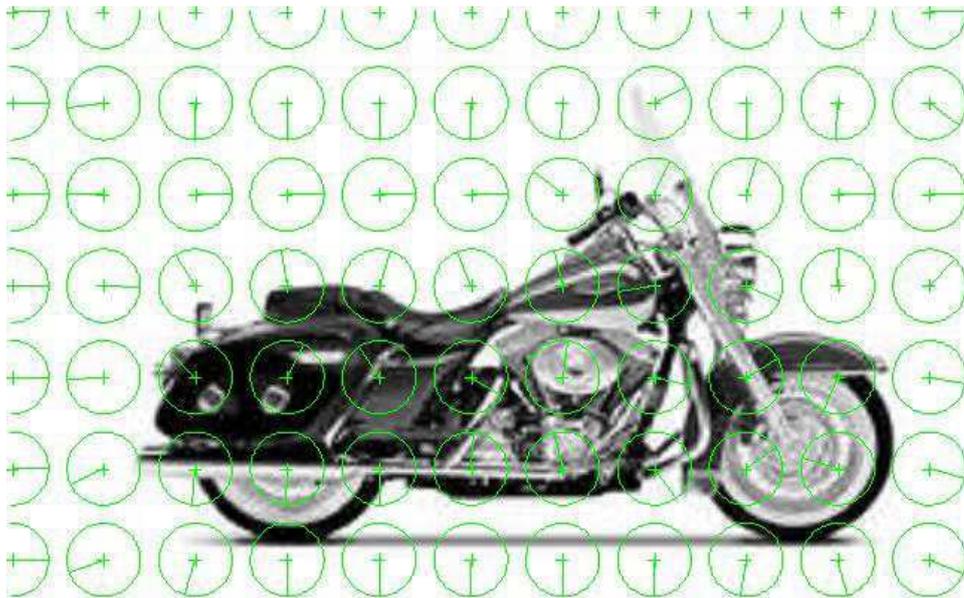
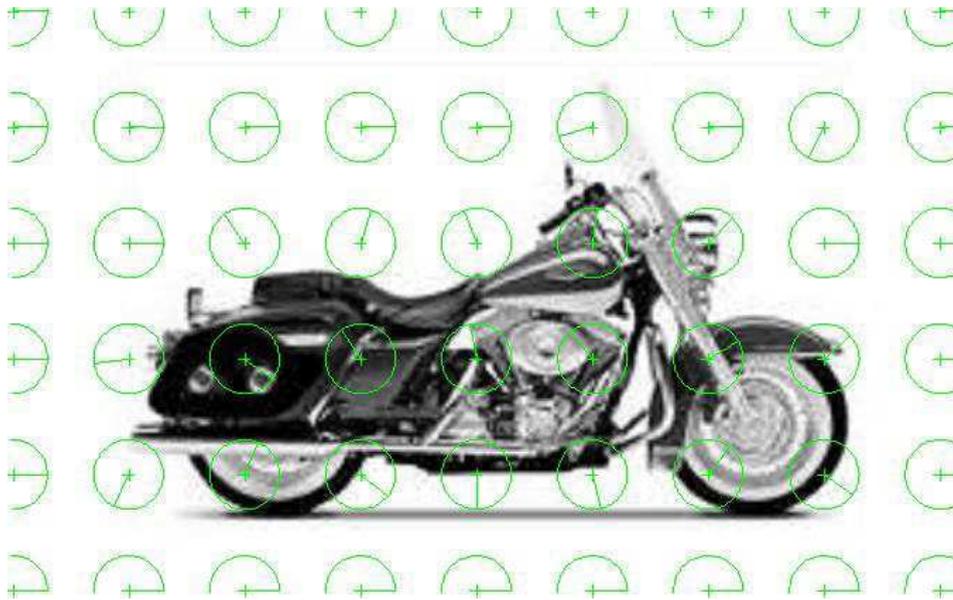


Figura 39. Exemplo de imagem com os pontos de interesse seleccionados pelo método de grade 32x32 e descritor SURF



Os melhores resultados foram obtidos com grade de dimensão 16x16. A taxa de reconhecimento em função do número de *clusters* e do número de imagens no conjunto de treino é apresentada para esse espaçamento de detecção na Tabela 17.

Tabela 17. Taxas de reconhecimento do classificador gerado com detecção de pontos de interesse por grade 16x16 e descritor SURF

Grade [16 16] e Descritor SURF		Número de Clusters				
		450	1050	2050	3050	4050
Porcentagem de imagens utilizadas para o conjunto de treino	40%	0,7119	0,7345	0,7472	0,7542	0,7415
	50%	0,7364	0,7517	0,767	0,7806	0,7857
	60%	0,7201	0,7735	0,7778	0,7991	0,7991
	70%	0,7155	0,7328	0,7557	0,7615	0,7787
	80%	0,7375	0,7542	0,7917	0,7917	0,7875
	90%	0,7583	0,7833	0,7917	0,8583	0,8417

Os tempos de cada etapa de geração de classificador e teste são apresentados na Tabela 18, com o parâmetro de porcentagem de imagens de treino fixado em 90%.

Tabela 18. Tempos de processamento em segundos para cada etapa de geração do classificador e teste utilizando detecção por grade 16x16 e descritor SURF

10 imagens no conjunto de treino		Número de Clusters				
		450	1050	2050	3050	4050
Tempos (s)	Gerar Bag Of Features	129,01	150,3	188,81	178,65	244,29
	Treinar Classificador	57,345	57,709	74,109	76,89	80,748
	Avaliar o Conj. Teste	7,0433	6,7149	9,3577	7,8162	8,2805
	Tempo médio para avaliar uma imagem de teste	0,7043	0,6715	0,9358	0,7816	0,8281

Os gráficos da taxa de reconhecimento e o tempo para classificação de uma imagem são apresentados na Figura 40 e 41 respectivamente. A taxa de reconhecimento máxima para esse método ocorreu nas mesmas situações que a obtida com o método com detecção e descrição SURF, com um valor de 0,8583 ou 85,83% de reconhecimento.

Figura 40. Gráfico da maior taxa de reconhecimento do classificador para cada número de clusters (detecção por grade 16x16 e descritor SURF)

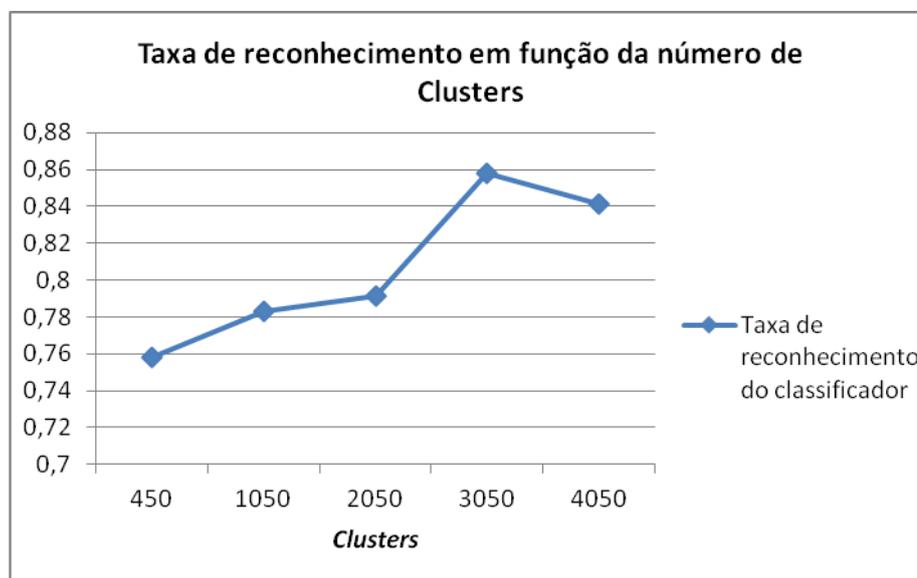
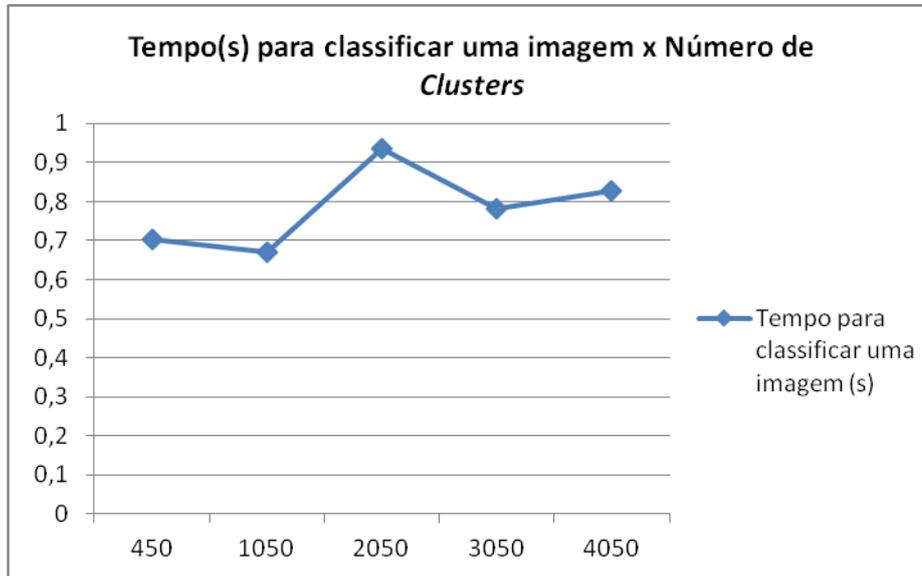


Figura 41. Tempo em segundos gasto para classificar uma imagem em função do número de *clusters* (detecção por grade 16x16 e descritor SURF)



4.2.3 HOG

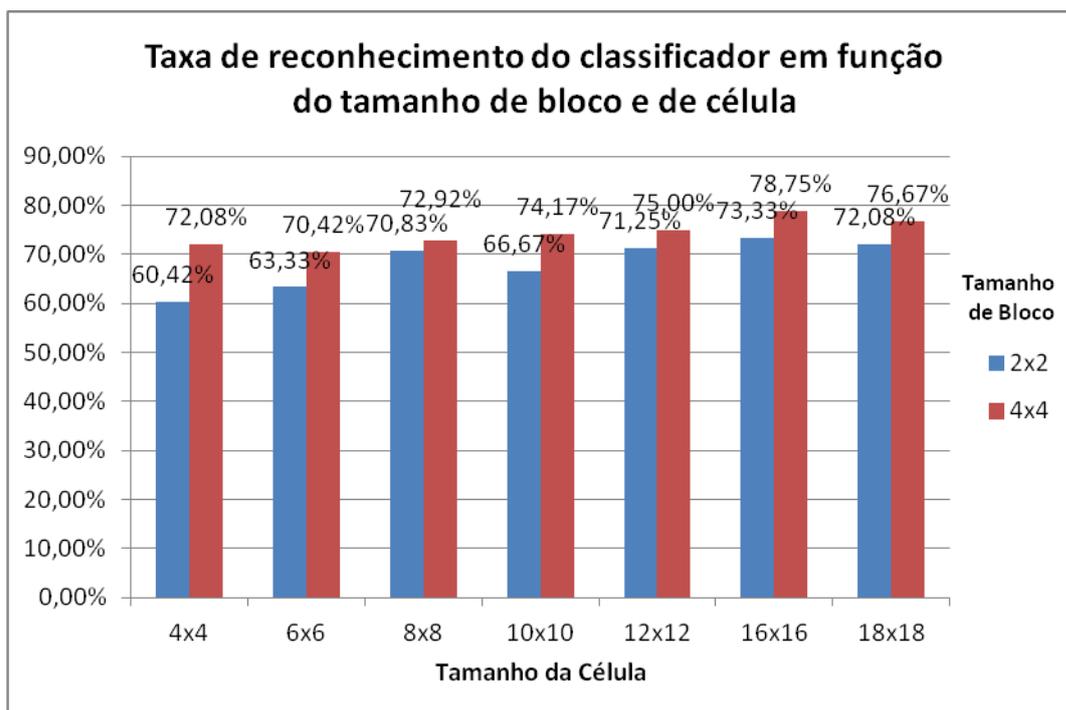
Com a utilização do descritor HOG foi possível obter bons resultados de maneira local e também global. Para detectar os pontos de interesse foram utilizados o detector SURF e o método por grade.

Pela detecção de pontos com o SURF, as taxas de reconhecimento foram calculadas em função das diferentes dimensões de célula e bloco utilizadas para descrever as características. A número de *clusters* foi fixada em 1050 e o conjunto de treino contou com 80% das imagens totais. A Tabela 19 apresenta o resumo dos testes realizados para descobrir a melhor taxa de reconhecimento para esse método. Na Figura 42 tem-se o gráfico da taxa de reconhecimento em função do tamanho de célula e de bloco.

Tabela 19. Taxa de reconhecimento do classificador em relação ao tamanho de célula e bloco para o método com detecção SURF e descrição HOG, para 1050 clusters e 80% das imagens para treino

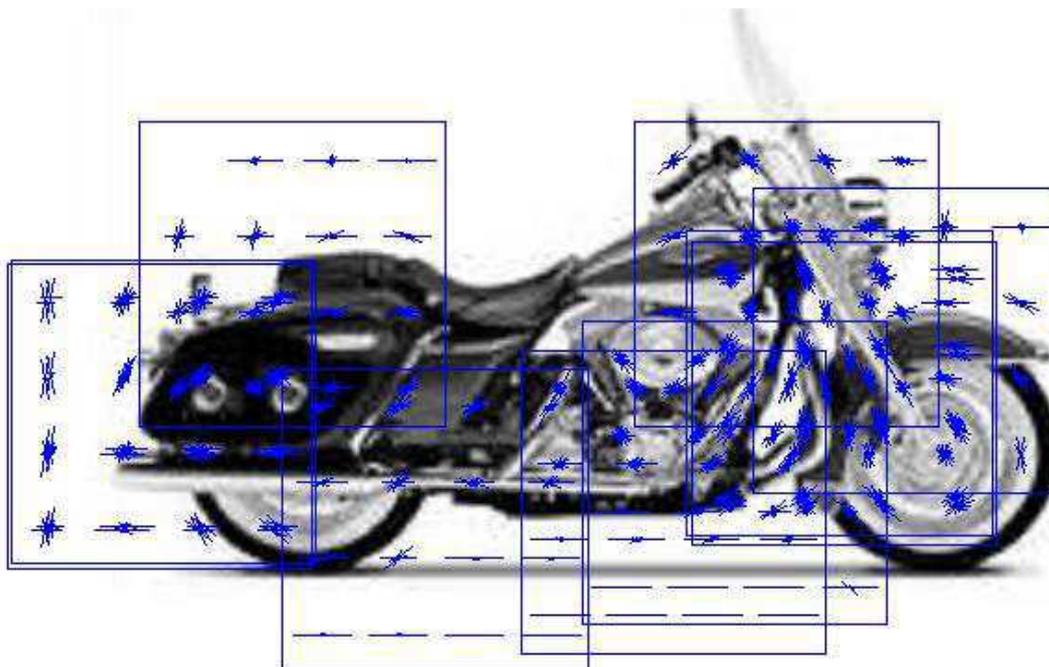
Taxa de reconhecimento média do classificador aplicado ao conjunto de teste		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	4x4	60,42%	72,08%
	6x6	63,33%	70,42%
	8x8	70,83%	72,92%
	10x10	66,67%	74,17%
	12x12	71,25%	75,00%
	16x16	73,33%	78,75%
	18x18	72,08%	76,67%

Figura 42. Gráfico da taxa de reconhecimento em porcentagem do classificador em função do tamanho de célula e de bloco



A Figura 43 apresenta visualização dos descritores HOG sobre os dez pontos mais significantes localizados pelo SURF.

Figura 43. Exemplo de imagem com os dez pontos de interesse mais significativos detectados pelo SURF descritos por HOG com célula 14x14 e bloco 4x4

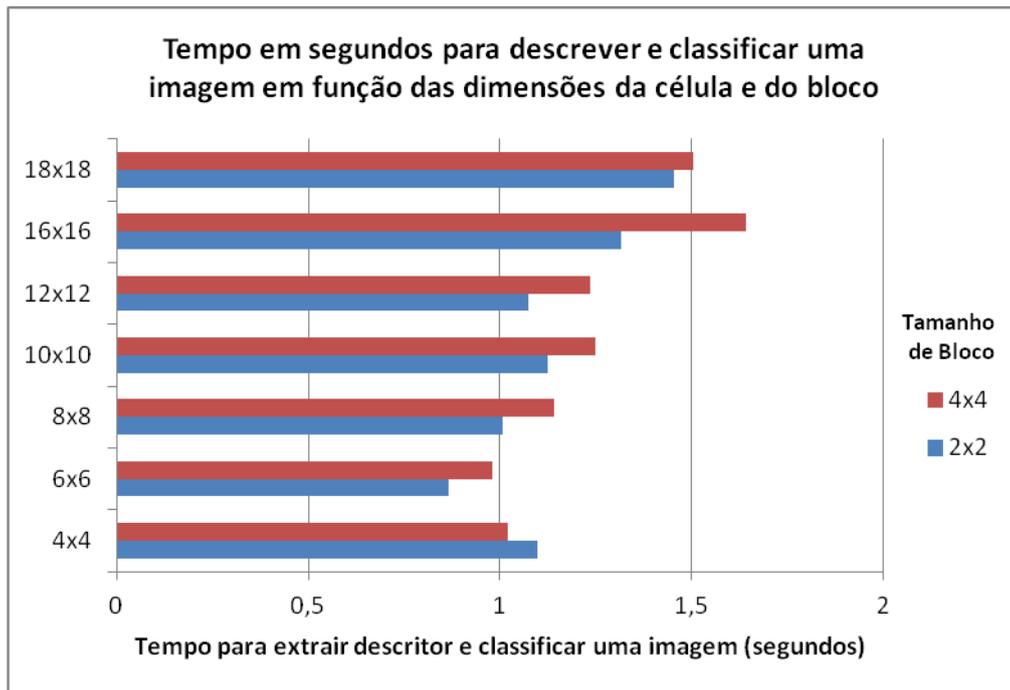


O tempo gasto para classificar uma imagem em função das dimensões do extrator HOG é apresentado na Tabela 20. Na Figura 44 tem-se o gráfico do tempo de descrição e classificação de uma imagem para os diferentes tamanhos de célula e bloco testados.

Tabela 20. Tempo para classificar uma imagem em segundos para o método de detecção por SURF e descrição por HOG, para 1050 clusters e 80% das imagens para treino

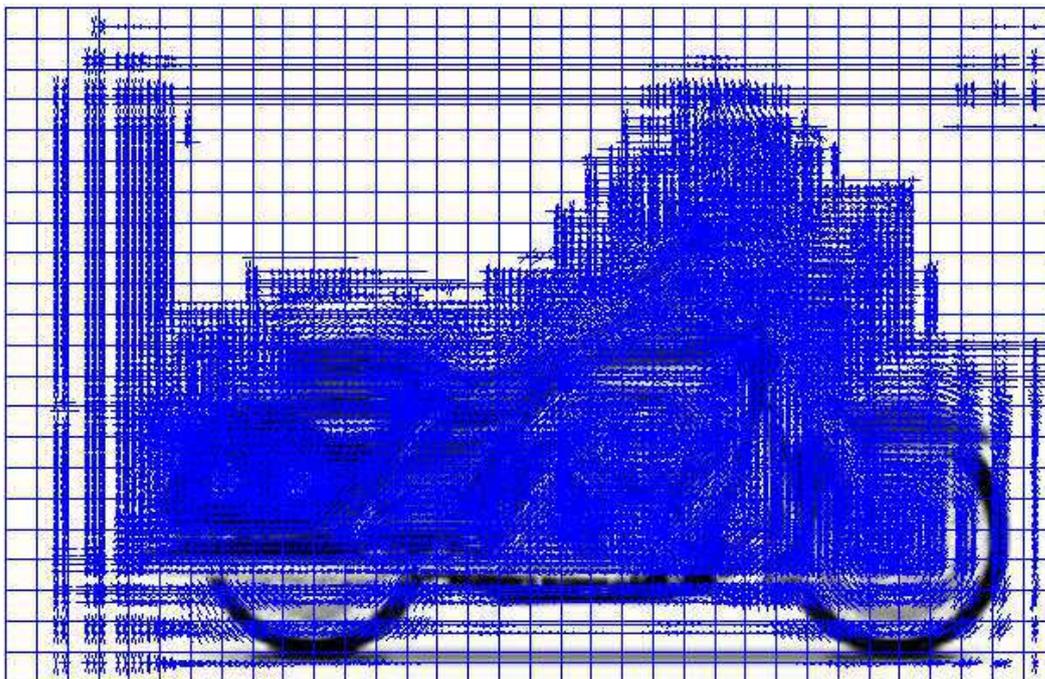
Tempo médio para extrair descritor e classificar uma imagem (s)		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	4x4	1,096305	1,02075
	6x6	0,86659	0,980095
	8x8	1,00708	1,14289
	10x10	1,12366	1,24766
	12x12	1,07425	1,23588
	16x16	1,316505	1,6407
	18x18	1,452025	1,503495

Figura 44. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste



Utilizando a forma de detecção por grade 8x8 obtêm-se uma descrição de cada região espaçada de 8 *pixels* na horizontal e na vertical da imagem como pode ser visualizado na Figura 45.

Figura 45. Exemplo de imagem com pontos de interesse selecionados por grade 8x8 e descritos por HOG



A taxa de reconhecimento em função das dimensões de bloco e célula é apresentada na Tabela 21 e seu gráfico na Figura 46. Da mesma forma, foi fixada a número de *clusters* em 1050 e o conjunto de treino contou com 80% das imagens totais. Os tempos de classificação de uma imagem são apresentados na Tabela 22 e seu gráfico na Figura 47.

Tabela 21. Taxa de reconhecimento do classificador em relação ao tamanho de célula e bloco para o método de detecção por grade 8x8 e HOG, para 1050 clusters e 80% das imagens para treino

Taxa de reconhecimento média do classificador aplicado ao conjunto de teste		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	6x6	71,67%	74,58%
	10x10	72,50%	80,83%
	14x14	73,33%	77,92%
	18x18	75,00%	77,92%

Figura 46. Gráfico da taxa de reconhecimento em porcentagem do classificador em função do tamanho de célula e de bloco

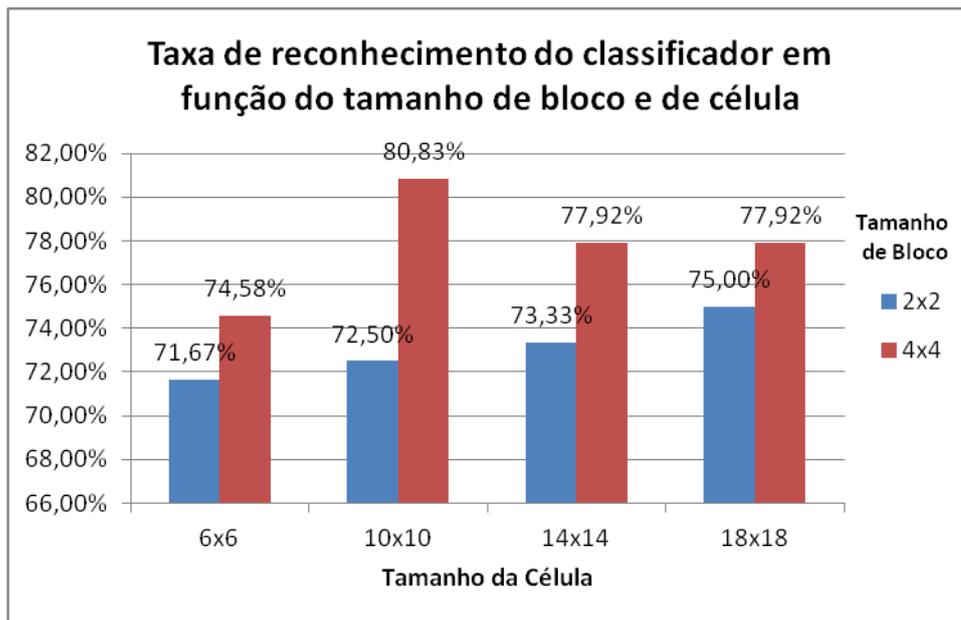
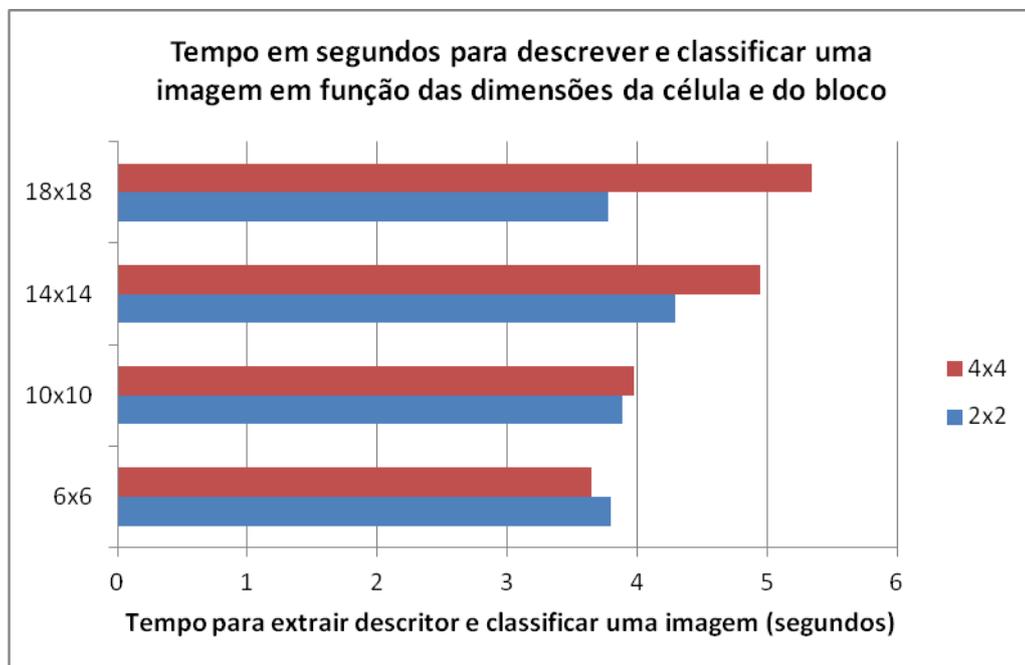


Tabela 22. Tempo para classificar uma imagem em segundos para o método de detecção por grade 8x8 e descrição por HOG, para 1050 clusters e 80% das imagens para treino

Tempo médio para extrair descritor e classificar uma imagem (s)		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	6x6	3,798105	3,65041
	10x10	3,889495	3,9766
	14x14	4,292005	4,950745
	18x18	3,77953	5,34092

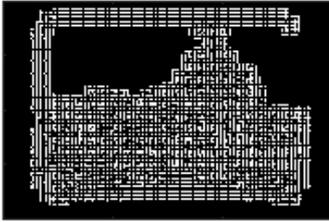
Figura 47. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste



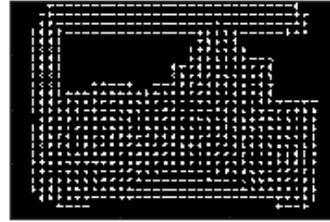
O HOG Global foi analisado variando-se as dimensões de célula e bloco conforme foi realizado com as palavras manuscritas anteriormente, a visualização de seus efeitos é apresentada na Figura 48 e 49, para tamanho de bloco de 2x2 e 4x4 respectivamente.

Figura 48. Visualização do HOG utilizando bloco de dimensão 2x2 com diversos tamanhos de célula

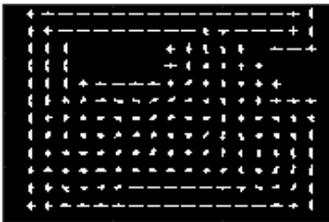
Tamanho de Célula = [4 4]
Tamanho de vetor de característica = 130536



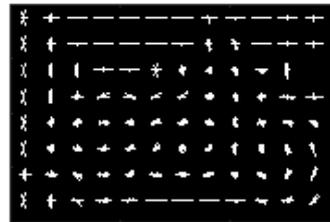
Tamanho de Célula = [8 8]
Tamanho de vetor de característica = 31104



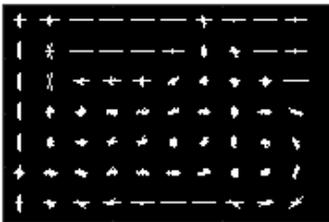
Tamanho de Célula = [16 16]
Tamanho de vetor de característica = 6732



Tamanho de Célula = [24 24]
Tamanho de vetor de característica = 2772



Tamanho de Célula = [28 28]
Tamanho de vetor de característica = 1944



Tamanho de Célula = [32 32]
Tamanho de vetor de característica = 1440

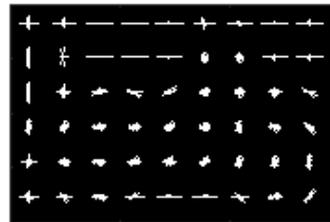
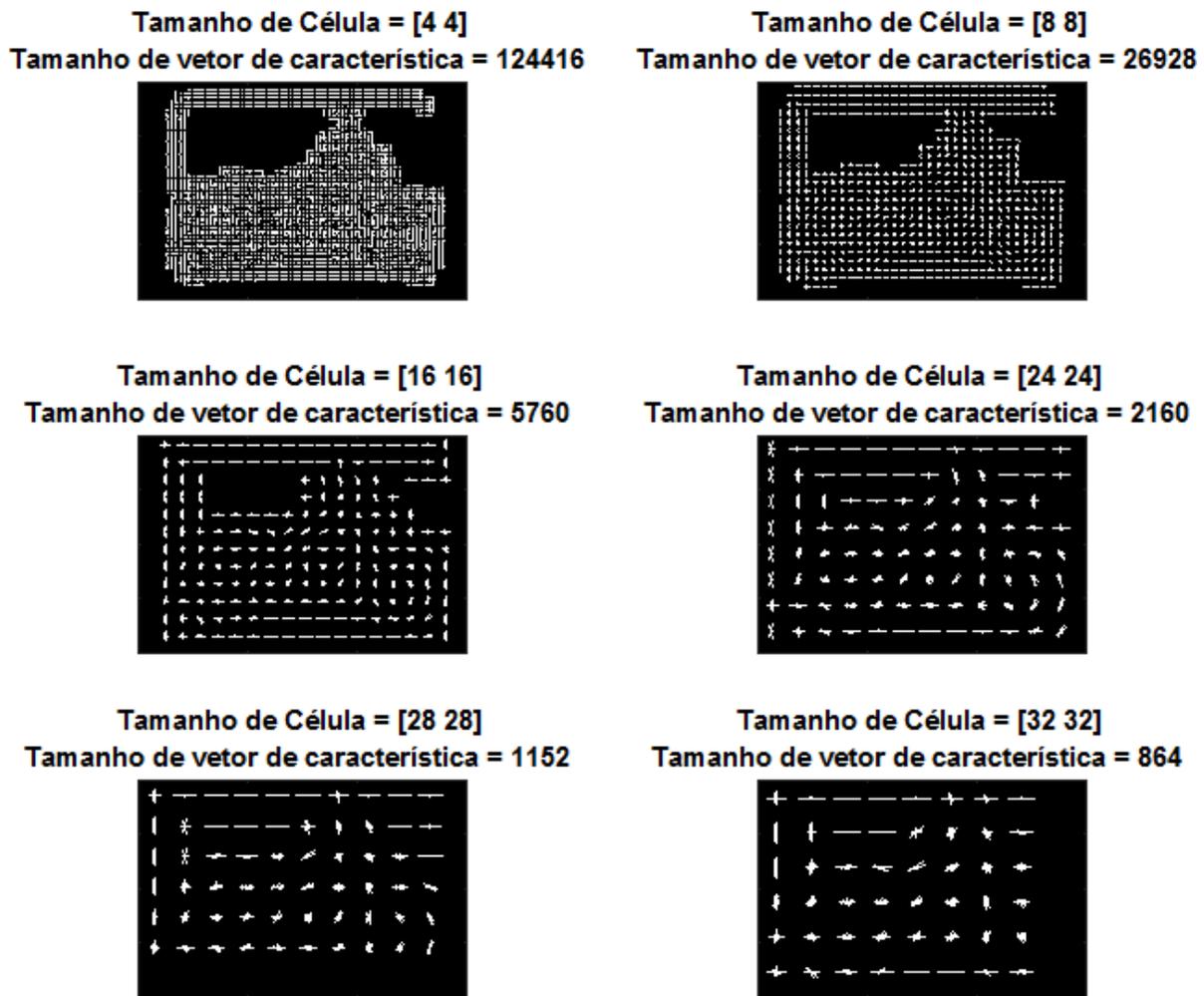


Figura 49. Visualização do HOG utilizando bloco de dimensão 4x4 com diversos tamanhos de célula



A taxa de reconhecimento para o classificador é apresentada na Tabela 23 sob as mesmas condições de 1050 *clusters* e 80% de imagens utilizadas para treino, e seu gráfico é apresentado na Figura 50. Os tempos médios de descrição e classificação de uma imagem de teste são apresentados na Tabela 24 e seu gráfico na Figura 51.

Tabela 23. Taxa de reconhecimento do classificador pelo método de extração global por HOG, para 1050 clusters e 80% das imagens para treino

Taxa de reconhecimento média do classificador aplicado ao conjunto de teste		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	4x4	76,67%	79,58%
	8x8	79,58%	81,67%
	16x16	83,33%	81,67%
	24x24	85,83%	81,25%
	28x28	86,67%	81,67%
	32x32	84,17%	84,17%
	40x40	85,00%	75,00%

Figura 50. Gráfico da taxa de reconhecimento em porcentagem do classificador em função do tamanho de célula e de bloco

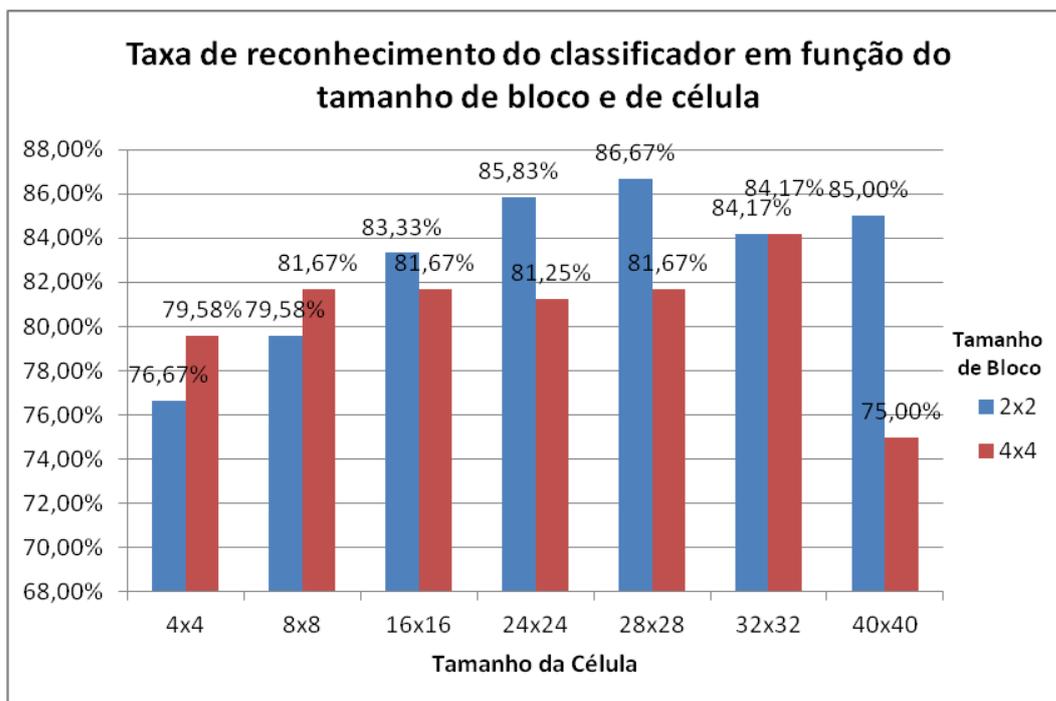
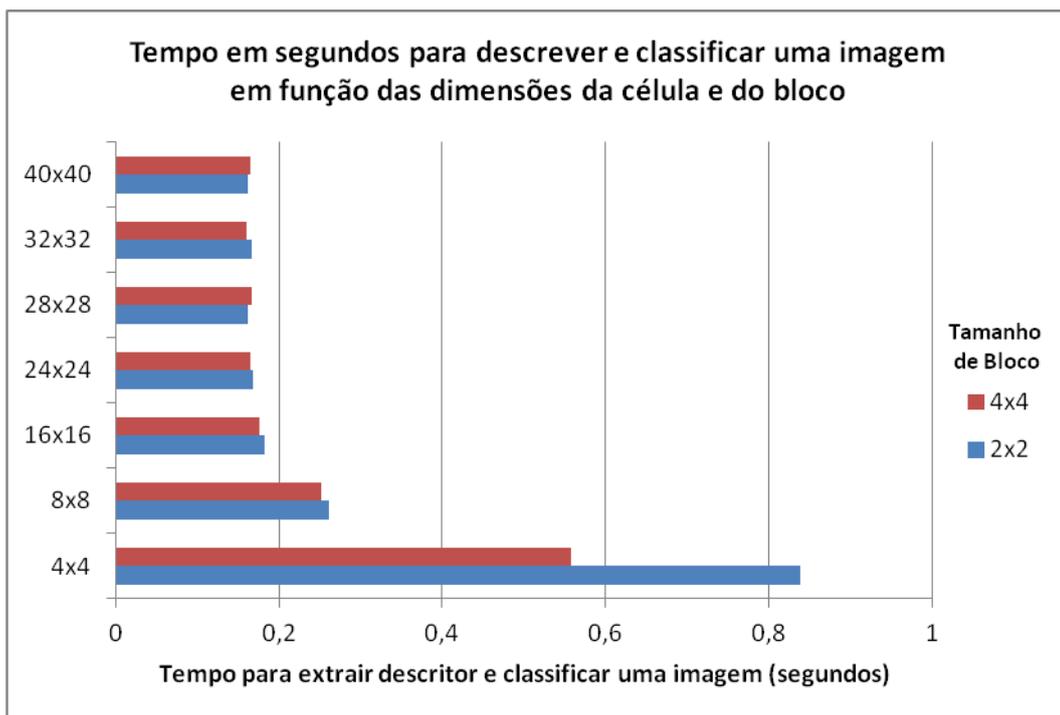


Tabela 24. Tempo de extração do descritor e classificação de imagem de teste pelo método de extração global por HOG, para 1050 clusters e 80% das imagens para treino

Tempo médio para extrair descritor e classificar uma imagem (s)		Tamanho do Bloco	
		2x2	4x4
Tamanho da Célula	4x4	0,837505	0,558305
	8x8	0,261065	0,25186
	16x16	0,18239	0,17682
	24x24	0,16819	0,165695
	28x28	0,16235	0,16605
	32x32	0,167425	0,160935
	40x40	0,16266	0,165705

Figura 51. Gráfico do tempo necessário em segundos para extrair descritor e classificar uma imagem de teste



Percebe-se que os métodos com o descritor HOG local não superaram os anteriores utilizando SURF. Contudo, o método utilizando HOG global obteve uma alta taxa de sucesso para um tamanho de célula de 28x28 e dimensão de bloco 2x2.

4.2.4 SIFT

Com a ferramenta disponibilizada pelo autor do SIFT e pela disponível no VLFeat foi possível analisar a extração de características dos objetos com esse método. Entretanto, os resultados não foram conforme os esperados e apresentados por outros trabalhos e pesquisas.

Os pontos detectados utilizando a função desenvolvida no VLFeat apresentaram maior informação visual quando apresentados sobre a imagem. Um exemplo é apresentado na Figura 52. Quanto aos pontos detectados pela função demonstrativa do autor, obteve-se sua origem e orientação com magnitudes diferentes representando as escalas de extração, conforme pode ser visualizado na Figura 53.

Figura 52. Pontos de interesse detectados pela função SIFT do VLFeat

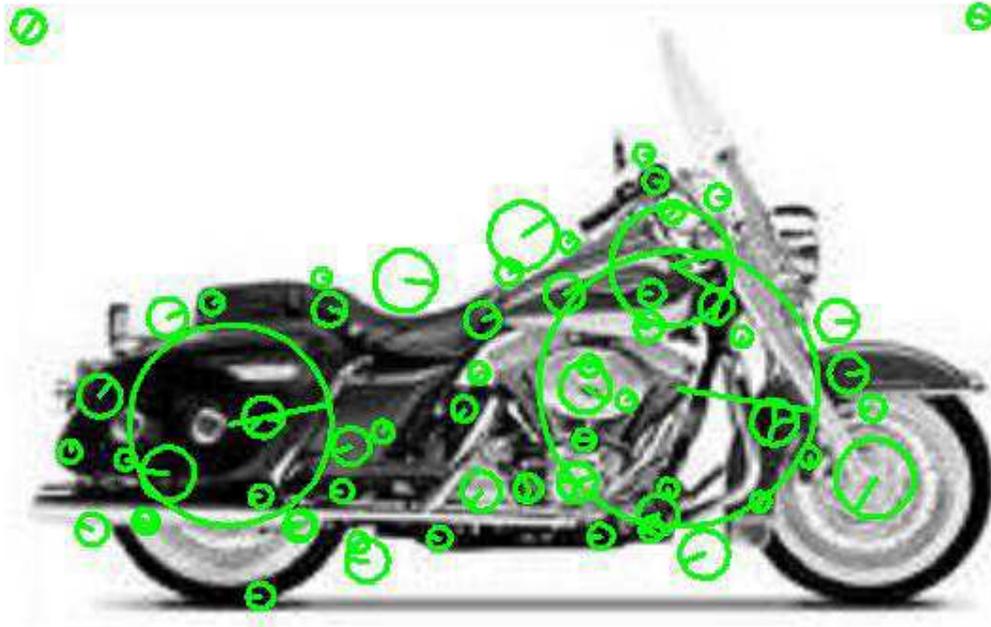
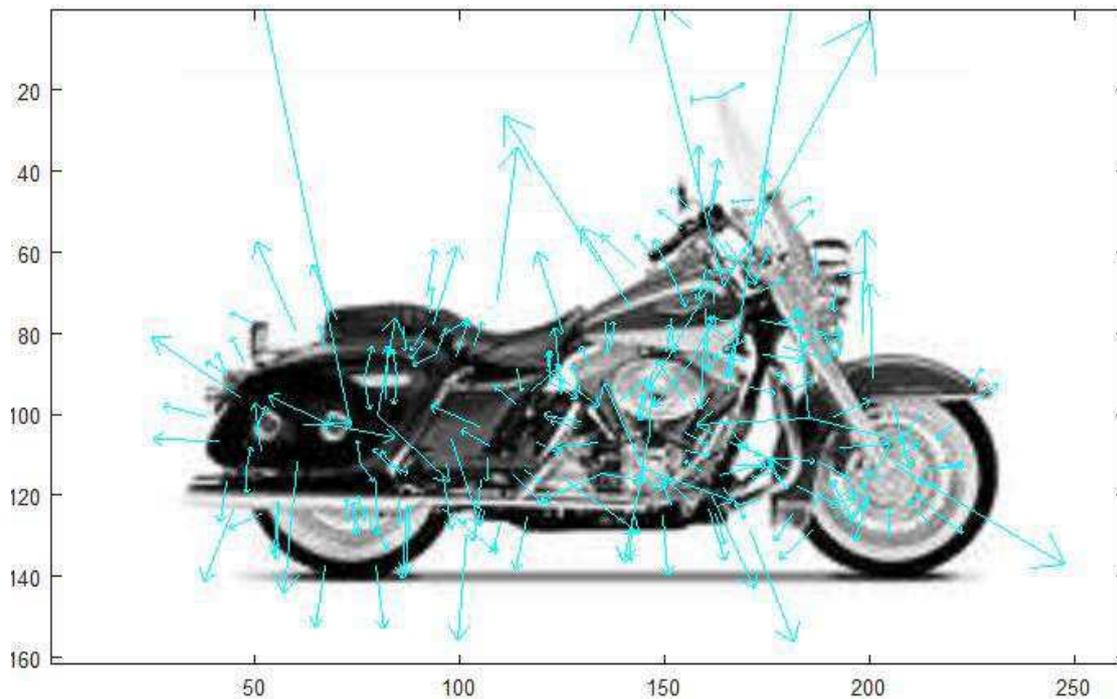


Figura 53. Pontos de interesse detectados pela função demonstrativa disponibilizada por David Lowe



A taxa de reconhecimento do classificador gerado com as informações extraídas pelo código demonstrativo de David Lowe apresentou melhores resultados que podem ser visualizados na Tabela 25.

Tabela 25. Taxa de reconhecimento do classificador pelo método de extração com SIFT

Porcentagem de imagens utilizadas para o conjunto de treino	Número de Clusters				
	450	1050	2050	3050	4050
80%	0,6917	0,7208	0,7417	0,7333	0,7208

4.3 ANÁLISE COMPARATIVA

As maiores taxas de reconhecimento dos métodos aplicados ao conjunto de imagens de manuscritos são apresentadas na Figura 54, e para acompanhar sua análise de desempenho também são apresentados na Figura 55 os tempos gastos correspondentes para analisar uma imagem. Ressaltando o fato que o método realizado com SIFT foi obtido utilizando-se de uma função de biblioteca externa, não tendo condições de comparar seu

desempenho diretamente aos demais. Nessas condições seu tempo de classificação de imagem foi retirado da análise gráfica. É relevante lembrar que ambos os classificadores trabalharam com 12 classes distintas e que com o maior número de classes maior a probabilidade de erro.

Figura 54. Maior taxa de reconhecimento do classificador de cada método aplicado a base de imagens dos manuscritos

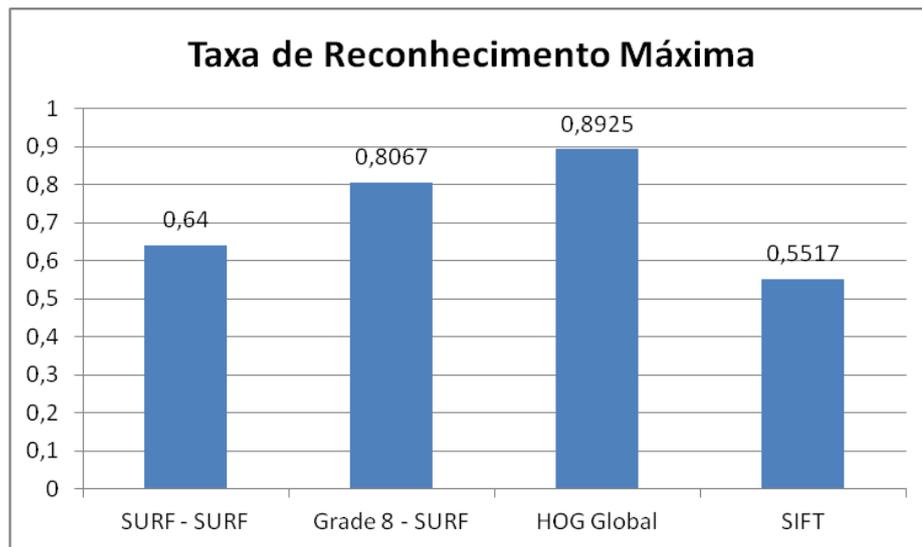
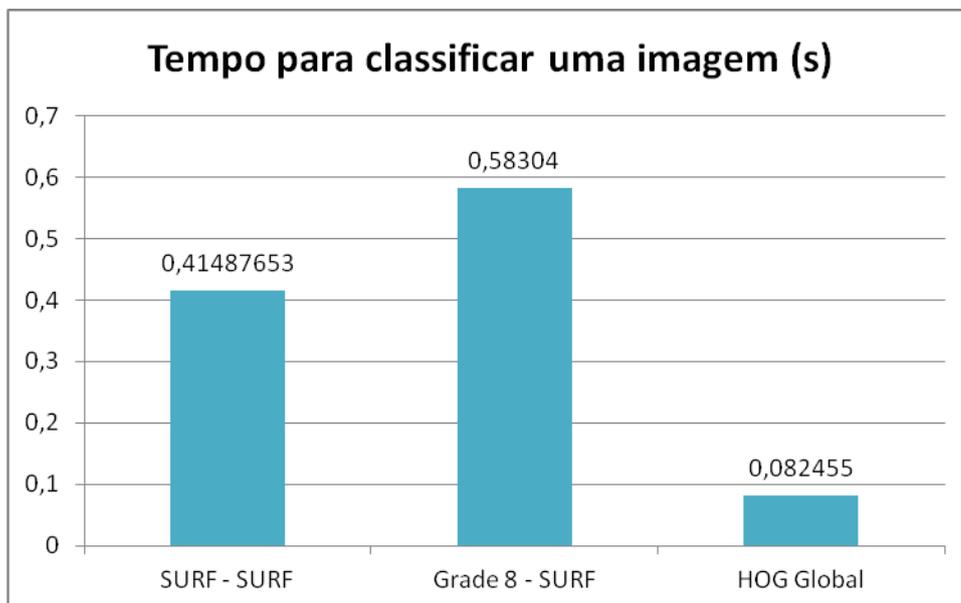


Figura 55. Tempo em segundos para classificar uma imagem correspondente às condições de maior reconhecimento de cada método com 1050 clusters aplicado à base de imagens dos manuscritos



Os mesmos dados também são visualizados para os métodos aplicados à base de imagens do Caltech na Figura 56 e 57.

Figura 56. Maior taxa de reconhecimento do classificador de cada método aplicado a base de imagens do Caltech

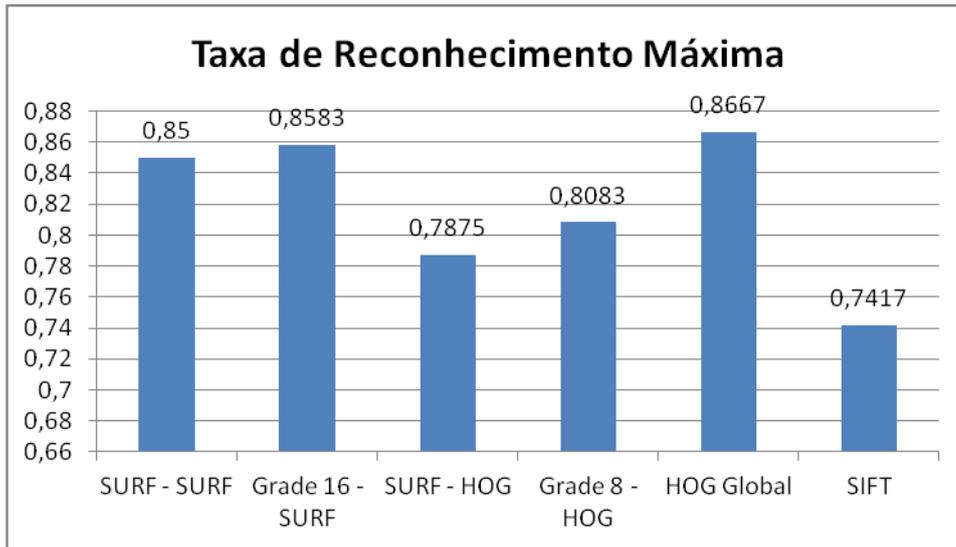
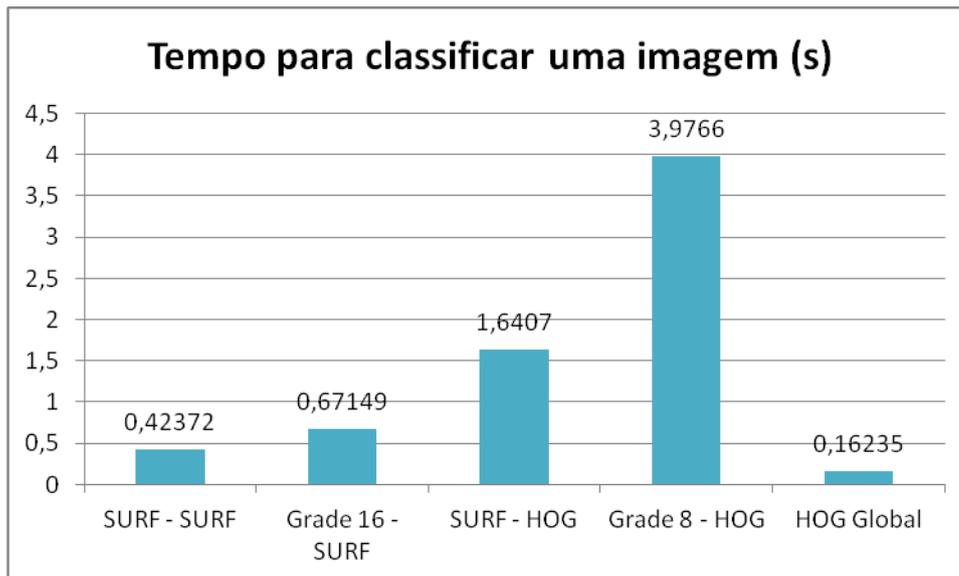


Figura 57. Tempo em segundos para classificar uma imagem correspondente às condições de maior reconhecimento de cada método com 1050 clusters aplicado à base de imagens do Caltech



5 CONCLUSÃO

Neste trabalho foi possível reproduzir o método de pacote de palavras visuais com o auxílio do *software* MATLAB para testes de extratores de característica em uma aplicação de reconhecimento de palavras e outra para reconhecimento de objetos. Dentre os extratores, a maioria dos testes foi realizada com extratores locais, como o SURF e o SIFT, e o extrator HOG foi utilizado tanto como extrator local quanto global.

O extrator SIFT não reproduziu o êxito esperado em relação a pesquisas anteriores, provavelmente pela maneira como as funções foram implementadas. Tendo em vista o perfil comparativo da análise, o fato das demais funções serem desenvolvidas pelo MATLAB favoreceu seu estudo e aplicação em detrimento da função do extrator SIFT.

O descritor HOG superou a taxa de reconhecimento de outros extratores e quase completamente no quesito tempo, superando expectativas para a função de classificação. Embora tenha tido um bom desempenho, o HOG utilizado para descrever a imagem globalmente não possui as vantagens de outros métodos para rastreamento e correspondência de objetos entre imagens.

Observando os resultados experimentais foi possível perceber que cada método aplicado teve um desempenho diferente em relação ao tipo de informação que deveria tratar. Tal comportamento se explica pelas diferentes maneiras com que cada algoritmo extrai as características da imagem, sua maneira de descrevê-lo e se é invariante ou não a certas modificações. Dessa maneira, é importante estudar e analisar que tipo de extrator melhor se encaixa para o tipo de informação a ser tratada.

A pesquisa limitou-se às imagens da base de dados dos manuscritos e do Caltech. Uma análise com outras imagens com diferentes escalas, iluminação e perspectivas enriqueceria o experimento ao aproximar-se mais de uma aplicação mais ligada a prática. Para melhores comparações e resultados mais refinados será necessário investigar outras alternativas para cada passo do algoritmo, como ferramentas para agrupamento dos *clusters*, classificadores, outros detectores de característica, etc.

6 REFERÊNCIAS

BAY, Herbert et al. Speeded-up robust features (SURF). **Computer vision and image understanding**, v. 110, n. 3, p. 346-359, 2008.

BAY, Herbert; TUYTELAARS, Tinne; VAN GOOL, Luc. Surf: Speeded up robust features. In: **Computer vision–ECCV 2006**. Springer Berlin Heidelberg, 2006. p. 404-417.

CHUM, Ondřej; ZISSERMAN, Andrew. An exemplar model for learning object classes. In: **Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on**. IEEE, 2007. p. 1-8.

CSURKA, Gabriella et al. Visual categorization with bags of keypoints. In: **Workshop on statistical learning in computer vision, ECCV**. 2004. p. 1-2.

DALAL, Navneet; TRIGGS, Bill. Histograms of oriented gradients for human detection. In: **Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on**. IEEE, 2005. p. 886-893.

FEI-FEI, Li; FERGUS, Rob; PERONA, Pietro. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. **Computer Vision and Image Understanding**, v. 106, n. 1, p. 59-70, 2007.

JÉGOU, Hervé; DOUZE, Matthijs; SCHMID, Cordelia. Packing bag-of-features. In: **Computer Vision, 2009 IEEE 12th International Conference on**. IEEE, 2009. p. 2357-2364.

JOACHIMS, Thorsten. Text categorization with support vector machines: Learning with many relevant features. Springer Berlin Heidelberg, 1998.

KIM, H. Y.; LÓPEZ, G. A. P. *Reconhecimento de Padrões Invariante Afim Baseado nos Coeficientes de Fourier de Projeções Radiais*. In: VI Workshop de Visão Computacional, 2010. FCT/UNESP - P. Prudente -São Paulo, 2010.

LOWE, David G. Distinctive image features from scale-invariant keypoints. **International journal of computer vision**, v. 60, n. 2, p. 91-110, 2004.

LOWE, David G. Object recognition from local scale-invariant features. In: **Computer vision, 1999. The proceedings of the seventh IEEE international conference on**. Ieee, 1999. p. 1150-1157.

MATHWORKS, Computer Vision System Toolbox Documentation. Disponível em: < <http://www.mathworks.com/help/vision/index.html>>. Acesso em: fevereiro de 2016.

MIKOLAJCZYK, Krystian; SCHMID, Cordelia. A performance evaluation of local descriptors. **Pattern Analysis and Machine Intelligence, IEEE Transactions on**, v. 27, n. 10, p. 1615-1630, 2005.

MORDVINTSEV, A; K. A. *Introduction to SIFT (Scale-Invariant Feature Transform)*. OpenCV-Python Tutorials. Disponível em: < http://opencv-python-tutroals.readthedocs.io/en/latest/py_tutorials/py_tutorials.html >, Acesso em: mai. 2016.

OLIVEIRA JUNIOR, J.J. *Reconhecimento de Palavras Manuscritas Usando Análise Multi-Vistas*. Outubro, 2006. 91 f. Tese (Doutorado em Engenharia Elétrica) - Universidade Federal de Campina Grande, Campina Grande, Paraíba, Brasil. 2006.

Pattern Recognition Systems – Lab 5: Histograms of Oriented Gradients. Disponível em: < <http://spotidoc.com/doc/322697/pattern-recognition-systems-%E2%80%93-lab-5-histograms-of-oriente...>>. Acesso em: 06 de maio de 2016.

PHILBIN, James et al. Object retrieval with large vocabularies and fast spatial matching. In: **Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on**. IEEE, 2007. p. 1-8.

SALIS, Turíbio Tanus; PEREIRA, Guilherme Augusto Silva. *Contagem Automática de Tarugos de Aço por Meio de Visão Computacional (01)*. 2007.

SAVELONAS, Michalis A.; PRATIKAKIS, Ioannis; SFIKAS, Konstantinos. An overview of partial 3D object retrieval methodologies. **Multimedia Tools and Applications**, v. 74, n. 24, p. 11783-11808, 2015.

SIVIC, Josef et al. *Discovering object categories in image collections*. 2005.

SZELISKI, Richard. **Computer vision: algorithms and applications**. Springer Science & Business Media, 2010.

VASCONCELOS, F. N. de O.; GIBSON, R. P. *Reconhecimento de Objetos em Tempo Real Utilizando Visão Computacional para o Futebol de Robôs*. Trabalho de Conclusão de Curso (Tecnologia em Análise e Desenvolvimento de Sistemas) - Instituto Federal de Educação, Ciência e Tecnologia de Pernambuco, Recife, Brasil, Agosto de 2014

VLFeat. Disponível em: < <http://www.vlfeat.org/index.html>>. Acesso em: 14 de maio de 2016.