



Reconhecimento Automático de Calorias numa Refeição

Pedro Manuel Pereira Rodrigues

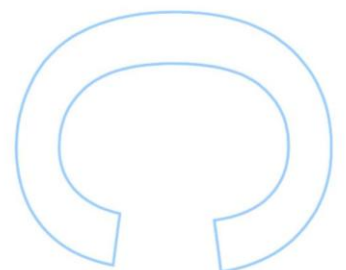
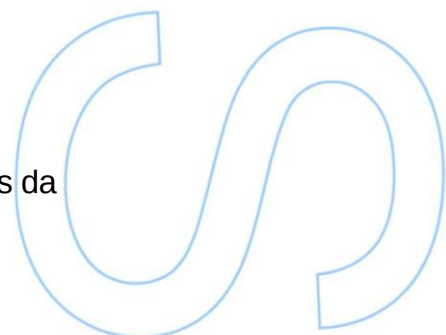
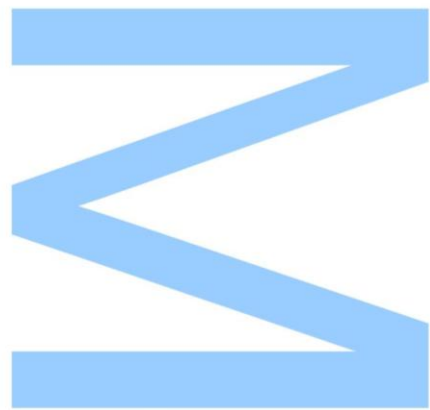
Mestrado integrado em Engenharia de Redes e Sistemas Informáticos
Departamento de Ciência de Computadores
2013

Orientador

Pedro Miguel Alves Brandão, Professor Auxiliar, Faculdade de Ciências da
Universidade do Porto

Coorientador

Miguel Tavares Coimbra, Professor Auxiliar, Faculdade de Ciências da
Universidade do Porto

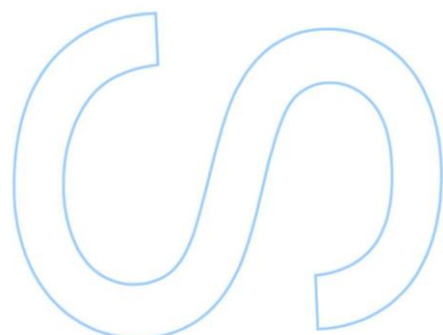
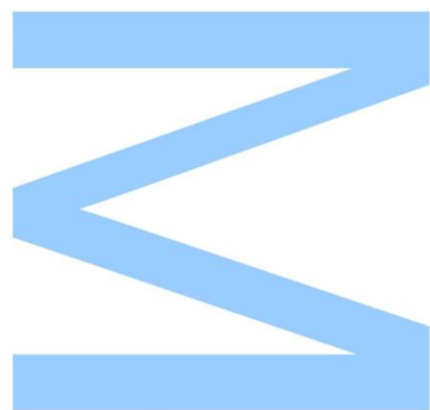




Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____ / ____ / ____



Aos meus Pais,

Maria Pereira e Manuel Rodrigues

À minha Namorada,

Paula Giesta

AGRADECIMENTOS

Aos meus orientadores, Professor Doutor Pedro Miguel Alves Brandão e Professor Doutor Miguel Tavares Coimbra, pela oportunidade concedida de integrar este projeto, pela disponibilidade constante, pelo apoio e incentivos e pelos bons comentários e sugestões que me facultaram no decorrer da execução da dissertação, e que contribuíram de forma decisiva para concretização deste projeto.

Agradeço aos meus pais Maria do Céu Pereira e Manuel Fernando Rodrigues pelo apoio incondicional a todos os níveis, por me terem providenciado a oportunidade de evoluir intelectual e emocionalmente, e por terem possibilitado a minha formação.

À minha namorada e amiga Paula Liliana Giesta um especial agradecimento por todo o carinho, apoio incondicional e incentivo dado ao longo da execução deste projeto e da minha formação como mestre.

Aos meus professores do curso Mestrado Integrado em Engenharia de Redes e Sistemas Informáticos pelos ensinamentos proporcionados e pela dedicação na transmissão e partilha de conhecimentos.

A todos aqueles que contribuíram de forma direta ou indireta para o meu desenvolvimento e crescimento como pessoa e como Mestre os meus sinceros agradecimentos.

QUASE TUDO O QUE FIZERMOS SERÁ INSIGNIFICANTE,
MAS É MUITO IMPORTANTE QUE O FAÇAMOS.

M. Gandhi

RESUMO

A integração das novas tecnologias nas áreas da saúde e bem-estar tem vindo a crescer de forma considerável, promovendo novas formas de levar um estilo de vida mais cuidado, de uma forma mais simples e cómoda. Com este projeto pretendemos propor um modelo para realizar reconhecimento automático de refeições e estimação calórica com base em imagens, capaz de auxiliar qualquer indivíduo, e em especial pessoas que padecem de problemas de saúde, tais como Hipertensão Arterial ou Diabetes, na manutenção de uma alimentação mais saudável através de um controlo das calorias ingeridas.

No decurso deste projeto analisamos as abordagens propostas por outros autores, enquadradas no âmbito da automatização da identificação de refeições e estimação calórica com base em imagens e efetuamos uma descrição e análise do funcionamento dos principais algoritmos do estado de arte de visão computacional para pesquisa de imagens com base no conteúdo (CBIR), nomeadamente SIFT e SURF. Sugerimos duas abordagens distintas para complementar a falta de informação de cor na descrição dos pontos de interesse pelos algoritmos SIFT e SURF, em concreto, a descrição dos pontos-chave para cada camada de cor da imagem e a concatenação de um histograma de cor a cada descritor base. Ainda em relação à introdução de informação de cor, propomos a utilização de histogramas de cor que descrevem a informação de cor globalmente para cada imagem. Analisamos os modelos de cor RGB e HSI por forma a determinar qual o modelo que garante uma melhor descrição da cor. Ao nível da manipulação dos descritores locais, analisamos o desempenho de uma abordagem baseada em *Bag of Words*, e propomos a utilização de *Bag of Words* por Classe com o objetivo de aproveitar a divisão intrínseca das diferentes classes de refeições. Propomos também a normalização do número de descritores por imagem com *Nearest Neighbor Joining* por forma a garantir um maior equilíbrio entre as classes de refeições na construção do *Bag of Words*. Para a estimativa calórica das refeições propomos a utilização do número de descritores locais extraídos como fator de comparação.

Em termos de resultados, as metodologias padrão oferecem uma base muito interessante no reconhecimento automático de refeições. A esta base, as metodologias propostas, nomeadamente de introdução de informação de cor na descrição dos pontos de interesse e aplicação de *Bag of Words* por Classe contribuem

para um aumento significativo do desempenho. Ao nível da estimação calórica, a metodologia apresentada consegue ser satisfatória em cenários relativamente simples. Desenvolveu-se uma interface gráfica com o objetivo de auxiliar na interação com o modelo proposto e proporcionar uma forma simples de testar o desempenho dos algoritmos utilizados.

Palavras Chave: CBIR, SIFT, SUFT, RGB, HSI, Bag of Words, Nearest Neighbor Joining

ABSTRACT

The integration of new technologies in the areas of health and wellness has grown considerably, providing new ways to lead a more careful lifestyle, in a more simple and convenient way. With this project, we propose a model to perform automatic recognition of meals and estimation of calories based on images, able to assist any person, especially those who suffer from health problems such as Hypertension or Diabetes, in maintaining a healthier diet by controlling the ingested calories.

In this project, we analyze approaches proposed by other authors, framed in the context of meals recognition and caloric estimation based on images and we do a description and analysis of the main computer vision algorithms for content-based image retrieval (CBIR), namely SIFT and SURF. We suggest two different approaches to complement the lack of color information in the description of the interest points with SIFT and SURF, specifically the description of the key points for each image color layer and the concatenation of a color histogram in each base descriptor. We also propose the use of color histograms for describing globally the color information for each image. We test this with the color models RGB and HSI in order to determine which model provides a better color description. With regards to manipulation of local descriptors, we analyze the performance of an approach based on Bag of Words, and we propose the use of Bag of Words by Class in order to take advantage of the intrinsic division of the different classes of meals. We also propose the normalization of the number of descriptors per image with Nearest Neighbor Joining to ensure a better balance between the classes of meals in the construction of the Bag of Words. To estimate calorie meals we propose the use of the number of local descriptors extracted as a comparison factor.

In terms of results, the standard methodologies offer a very interesting base for automatic recognition of meals. With the proposed approaches, in particular the introduction of color information in the description of the points of interest and the application of Bag of Words by Class, we obtained a significant increase in performance compared to the standard methodology. At the level of calorie estimation, the proposed methodology can be satisfactory for relatively simple scenarios. We developed a graphical interface in order to assist in the interaction with the proposed model and provide a simple way to test the performance of algorithms used.

Keywords: CBIR, SIFT, SUFT, RGB, HSI, Bag of Words, Nearest Neighbor
Joining

ÍNDICE

AGRADECIMENTOS	4
RESUMO	6
ABSTRACT	8
ÍNDICE.....	10
ÍNDICE DE FIGURAS	12
ÍNDICE DE TABELAS	13
CAPÍTULO I: INTRODUÇÃO	16
1.1. Interesse e Fundamentação	16
1.2. Objetivos	17
1.3. Resultados Esperados	18
1.4. Estrutura da Dissertação	18
CAPÍTULO II: REVISÃO DA LITERATURA.....	21
2.1. Abordagens de Outros Autores.....	21
2.1.1. Reconhecimento de Alimentos e Estimação Calórica	21
2.1.2. Determinação do Tamanho da Porção de Comida.....	24
2.2. Aplicações.....	25
2.2.1. Meal Snap – Calorie Counting Magic.....	25
CAPÍTULO III: METODOLOGIAS	28
3.1. Pesquisa de Imagens com Base no Conteúdo.....	28
3.2. Detecção e Descrição de Características Locais.....	29
3.2.1. SIFT: Scale-Invariant Feature Transform.....	30
3.2.2. SURF: Speeded Up Robust Features	32
3.3. Detecção e Descrição de Características Globais.....	34
3.3.1. Espaços de Cor RGB(A) e HSI	34
3.3.2. Histograma de cor RGB.....	36
3.3.3. Histograma de cor HSI.....	37
3.4. Bag of Words.....	38
3.4.1. Clusterização com K-Means.....	38

CAPÍTULO IV: ALGORITMOS	41
4.1. Pesquisa de Imagens com Base no Conteúdo	41
4.1.1. Detecção e Descrição de Características Locais	42
4.1.1.1. Descrição dos Pontos-chave em cada Camada de Cor da Imagem	43
4.1.1.2. Concatenação de Histograma de Cor ao Descritor Base	44
4.1.2. Detecção e Descrição de Características Globais	46
4.1.3. Bag of Words	46
4.1.3.1. Bag of Words por Classe	47
4.1.3.2. Normalização do número de descritores com K-NNJ	47
4.1.3.3. Classificação	50
4.2. Estimação do Valor Calórico	51
CAPÍTULO V: APLICAÇÃO	54
5.1. Visão Geral do Projeto	54
5.2. Construção de Base de Dados	54
5.3. Pesquisa de Imagens com Base no Conteúdo	56
5.4. Estimação do Valor Calórico	57
5.5. Interface Gráfica de Utilizador	57
CAPÍTULO VI: RESULTADOS	61
6.1. Base de Dados	61
6.2. Reconhecimento de Refeições	62
6.3. Estimação Calórica	76
CAPÍTULO VII: CONCLUSÕES	79
7.1. Comentários Finais	79
7.2. Trabalho Futuro	81
REFERÊNCIAS	83
ANEXOS	86
ANEXO A: Conversão entre espaço de Cor RGB e HSI	86
ANEXO B: Base de Dados	87

ÍNDICE DE FIGURAS

Figura 1: Imagem que respeita os pressupostos impostos pela proposta de Shroff....	22
Figura 2: Etapas seguidas do pré-processamento à segmentação da imagem.....	23
Figura 3: Da esquerda para a direita, início e fim da segmentação da refeição	24
Figura 4: Interface gráfica da aplicação Meal Snap	26
Figura 5: Aplicação do SIFT para deteção dos Pontos-chave	32
Figura 6: Aplicação do SURF para deteção dos Pontos-chave	33
Figura 7: Espaço de Cores RGB.....	34
Figura 8: Espaço de Cores HSI	35
Figura 9: Esquerda para direita e cima para baixo, 53.976 (original), 512, 256 e 64 cores	36
Figura 10: Esquerda para direita e cima para baixo, 49.655 (original), 512, 256 e 64 valores	37
Figura 11: Pipeline de operações na construção da BD e pesquisa de imagens	41
Figura 12: <i>Pipeline</i> da deteção e descrição de características locais.....	42
Figura 13: Descrição dos pontos-chave em cada camada de cor da imagem	44
Figura 14: Concatenação de histograma de cor ao descritor base	45
Figura 15: Clusterização de descritores. Circunferências correspondem aos clusters	48
Figura 16: Deteção e Extração de 980 Descritores usando SURF.....	51
Figura 17: Deteção e Extração de 1985 Descritores usando SURF.....	51
Figura 18: Interface Gráfica.....	58
Figura 19: Menu de Base de Dados	59
Figura 20: Menu de Histórico	59
Figura 21: Exemplos de algumas das imagens utilizadas na análise à metodologia...	61

ÍNDICE DE TABELAS

Tabela 1: SIFT e SURF com <i>Bag of Words</i> de 50, 100 e 200 palavras. Podemos observar para cada imagem de consulta e para cada configuração, a posição em que ficou classificada a classe real da imagem de consulta.	63
Tabela 2: SIFT e BoW de 50 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SIFT e BoW de 50 palavras, sem informação de cor.	64
Tabela 3: SIFT e BoW de 100 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SIFT e BoW de 100 palavras, sem informação de cor.	65
Tabela 4: SURF e BoW de 50 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SURF e BoW de 50 palavras, sem informação de cor.	65
Tabela 5: SURF e BoW de 100 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SURF e BoW de 100 palavras, sem informação de cor.	66
Tabela 6: Resumo dos resultados com SIFT e SURF, Bag of Words de 50 e 100 palavras com Informação Local de Cor com DCC e CHC.	66
Tabela 7: SIFT e BoW de 50 palavras com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT com CHC-32HSI e BoW de 50 palavras.	67
Tabela 8: SURF e BoW de 50 palavras com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF com CHC-32HSI e BoW de 50 palavras.	68
Tabela 9: Bag of Words de 50 e 100 palavras por classe. Nas colunas de referência temos respetivamente os valores obtidos utilizando SIFT com BoW de 50 palavras e 100 palavras, e SURF com BoW de 50 palavras e 100 palavras, em ambos os casos sem introdução de informação de cor.	69
Tabela 10: SIFT e BoW de 50 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT e BoW de 50 palavras por classe, sem informação de cor.	70

Tabela 11: SIFT e BoW de 100 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT e BoW de 100 palavras por classe, sem informação de cor.	70
Tabela 12: SURF e BoW de 50 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF e BoW de 50 palavras por classe, sem informação de cor.	71
Tabela 13: SURF e BoW de 100 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF e BoW de 100 palavras por classe, sem informação de cor.	71
Tabela 14: Resumo dos resultados com SIFT e SURF, Bag of Words de 50 e 100 palavras por Classe com Informação Local de Cor com DCC e CHC.....	72
Tabela 15: SIFT e BoW de 50 palavras por classe com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT com CHC-32HSI e BoW de 50 palavras por classe.	73
Tabela 16: SURF e BoW de 50 palavras por classe com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF com CHC-32HSI e BoW de 50 palavras por classe.	74
Tabela 17: Amostragem de Descritores com NNJ. Nas colunas de referência temos respetivamente os valores obtidos utilizando SIFT com BoW de 50 palavras e 100 palavras, e SURF com BoW de 50 palavras e 100 palavras, em ambos os casos sem introdução de informação de cor.....	75
Tabela 18: Estimação Calórica com SIFT. Nas duas primeiras colunas temos o número de descritores locais extraídos para cada imagem e nas últimas duas colunas a diferença entre o número de descritores.	76
Tabela 19: Estimação Calórica com SURF. Nas duas primeiras colunas temos o número de descritores locais extraídos para cada imagem e nas últimas duas colunas a diferença entre o número de descritores.	77
Tabela 20: Imagens utilizadas durante a análise à metodologia proposta. Os pares de imagens com bordas vermelhas foram utilizados para a estimacão calórica.....	88

CAPÍTULO I: INTRODUÇÃO

CAPÍTULO I: INTRODUÇÃO

1.1. Interesse e Fundamentação

Com a evolução tecnológica assistimos à rápida integração das novas tecnologias no nosso quotidiano. Atualmente lidamos de forma constante com tecnologias, de forma didática ou profissional, seja nas atividades diárias mais simples às operações mais complexas. Podemos ver as novas tecnologias como ferramentas dinamizadoras, que estão presentes em basicamente todas as áreas e proporcionam um melhoramento inequívoco da nossa qualidade de vida ao tornar mais simples, agradável e eficiente a realização de todo o tipo de atividades. É notória também uma crescente preocupação com a saúde e bem-estar por parte da sociedade em geral, conscientes que a resposta a uma vida longa e saudável consegue-se por meio de uma alimentação cuidada e exercícios físicos regulares. Contudo como estamos inseridos numa sociedade “gerida pelo tempo” levar um estilo de vida saudável exige um esforço suplementar por vezes difícil de conciliar com as restantes responsabilidades.

Ao nível da alimentação nem sempre é simples conseguir seguir uma nutrição cuidada e muito menos controlada, e se por um lado pessoas consideradas como saudáveis podem relegar para segundo plano as preocupações com uma alimentação saudável com as respetivas repercussões negativas, isso não é o caso para pessoas que padecem de problemas de saúde, tais como **Hipertensão Arterial** (determinada por elevados níveis de pressão sanguínea nas artérias, o que faz com que o coração tenha que exercer um esforço maior do que o normal para fazer circular o sangue através dos vasos sanguíneos [1]) ou **Diabetes** (doença metabólica caracterizada por um aumento anormal do açúcar ou glicose no sangue [2]), em que a nutrição assume um papel de elevada importância. O tratamento de doenças crónicas como Hipertensão Arterial [3] e Diabetes [4] depende da conjugação de três fatores cruciais: alimentação, atividade física e medicação. Indivíduos com este tipo de doenças necessitam de ter um plano alimentar individualizado, adaptado às suas condições de vida, e equilibrado, quer em quantidade, quer em qualidade, em função do seu peso de referência, sexo e idade .

Assim e aproveitando o fácil acesso as novas tecnologias por parte da população em geral, em conjugação com a vontade ou mesmo necessidade de levar uma vida saudável por meio de uma nutrição correta, pretendemos providenciar um meio de combinar ambos os fatores e recorrendo à Visão Computacional introduzir um modelo simples capaz de auxiliar na manutenção de uma alimentação mais controlada. Podemos ver a Visão Computacional como o conjunto de métodos e técnicas utilizados por sistemas computacionais para interpretar sinais unidimensionais e multidimensionais (tais como imagens ou vídeos). A interpretação de um sinal consiste na transformação de um conjunto de dados digitais numa estrutura de dados descrevendo a semântica deste conjunto de dados num determinado contexto. A Visão Computacional é uma área que se encontra atualmente em grande desenvolvimento e evolução, isto muito influenciado pelo vasto leque de aplicações que possui e que vão desde a **Medicina** com o processamento médico e extração de informação de imagens para realização de diagnósticos, à **Indústria** auxiliando na execução de processos como por exemplo o controle de qualidade, passando por aplicações **Militares** entre muitas outras [5].

1.2. Objetivos

Os objetivos principais deste projeto podem ser divididos nos quatro grupos que se descrevem seguidamente:

- Adaptação e combinação de algoritmos do estado de arte de visão computacional para pesquisa de imagens com base no conteúdo;
- Identificação das limitações dos algoritmos adotados e estruturação de abordagens que permitam colmatar essas limitações;
- Definição de métodos estatísticos para estimação do valor calórico de refeições com base em imagens;
- Desenvolvimento de uma *framework* de suporte ao projeto, com uma interface gráfica, compacta e de simples utilização;

1.3. Resultados Esperados

Esperamos com este projeto desenvolver e apresentar um modelo relativamente flexível para automatizar a identificação de refeições e estimação calórica com base em imagens, seguindo uma abordagem de pesquisa de imagens com base no seu conteúdo. O modelo desenvolvido foi idealizado para ter como suporte um dispositivo móvel e ser o núcleo de uma aplicação orientada ao utilizador comum, onde não se espera qualquer tipo de controlo “laboratorial” sobre a forma como as imagens das refeições são recolhidas, e não sendo estipulada qualquer limitação sobre as refeições a identificar.

De notar que, ao nível da visão computacional aplicada ao processamento de imagens de refeições, estamos a falar de um enorme desafio devido a toda a variabilidade que se verifica nas refeições em termos de forma, disposição, cor, quantidade, etc. Em conjugação com todas estas dificuldades temos ainda que considerar a recolha (e processamento) de imagens a partir de dispositivos móveis o que por si só inviabiliza qualquer tentativa de controlo sobre o ambiente em termos de iluminação do meio, ou mesmo ângulo em que a imagem é recolhida.

1.4. Estrutura da Dissertação

A dissertação está dividida em sete capítulos, sendo que neste primeiro capítulo (**Introdução**) realizamos uma breve reflexão acerca da temática em si, os interesses envolvidos e a sua fundamentação, assim como a definição dos objetivos pretendidos, o estabelecimento dos resultados que se pretende atingir e uma consciencialização em relação à dificuldade inerente à execução do projeto.

No segundo capítulo (**Revisão da Literatura**) efetuamos uma descrição e análise das abordagens propostas por outros autores, enquadradas no âmbito da automatização da identificação de alimentos/refeições e estimação calórica com base em imagens. Fazemos também uma breve referência à aplicação para *smartphone* Meal Snap, a qual possibilita a identificação e estimação calórica de refeições.

No terceiro capítulo (**Metodologias**) damos algumas noções breves sobre processamento de imagem e explicamos de forma sucinta o funcionamento dos principais algoritmos do estado de arte de visão computacional para pesquisa de imagens com base no conteúdo.

No quarto capítulo (**Algoritmos**) identificamos os métodos adotados da biblioteca de visão computacional OpenCV e fazemos uma descrição completa dos algoritmos implementados para compensar as limitações dos algoritmos do estado de arte.

No quinto capítulo (**Aplicação**) temos uma visão geral do *pipeline* da aplicação desenvolvida. Indicamos de forma detalhada a abordagem seguida com vista à concretização dos objetivos propostos, assim como a justificação dos parâmetros escolhidos na configuração dos algoritmos aplicados. Para finalizar o capítulo, apresentamos a interface gráfica desenvolvida para auxiliar na interação com os algoritmos implementados e descrevemos sumariamente o seu funcionamento.

No sexto capítulo (**Resultados**) exibimos os resultados obtidos com a metodologia seguida e realizamos uma análise dos resultados em paralelo com a comparação entre os métodos utilizados, analisando-se o ganho das novas abordagens introduzidas em complemento ao estado de arte.

No sétimo e último capítulo (**Conclusões**) expomos as conclusões alcançadas no decurso deste projeto. Sugerimos também algumas abordagens a serem exploradas em futuros desenvolvimentos no âmbito do reconhecimento automático de calorias em imagens de refeições.

CAPÍTULO II: REVISÃO DA LITERATURA

CAPÍTULO II: REVISÃO DA LITERATURA

2.1. Abordagens de Outros Autores

Ao nível do reconhecimento automático de imagens de refeições e estimação calórica, embora exista ainda um longo caminho para progredir, temos visto algumas propostas interessantes, nomeadamente o trabalho realizado por Geeta Shroff no reconhecimento automático de alimentos e estimação calórica [6], e a proposta de Mingui Sun na determinação do tamanho da porção de comida [7].

2.1.1. Reconhecimento de Alimentos e Estimação Calórica

No trabalho executado por Geeta Shroff, “***Wearable Context-Aware Food Recognition for Nutrition Monitoring***” [6], é proposta a arquitetura DiaWear. O DiaWear é um sistema de reconhecimento de comida e estimação calórica com contexto. Após as imagens dos alimentos serem captadas pelo dispositivo do utilizador, estas são encaminhadas para um servidor que efetua o reconhecimento das imagens e devolve ao utilizador uma estimativa das calorias presentes nos respetivos alimentos.

No trabalho apresentado por Shroff partem-se dos seguintes pressupostos:

- Todos os objetos na imagem estão completos, existe um espaçamento entre cada par de objetos e existe um objeto de referência;
- O *background* da imagem só possui uma única cor, sendo essa cor mais clara que a cor de qualquer um dos objetos presentes na cena;
- As sombras presentes na imagem são mínimas e não é utilizada iluminação proveniente de flash;

São consideradas apenas quatro categorias de alimentos: Hambúrgueres, Batatas fritas, Nuggets de frango e Tarte de maçã.

Na Figura 1 podemos observar uma imagem que obedece aos pressupostos impostos.



Figura 1: Imagem que respeita os pressupostos impostos pela proposta de Shroff

O método apresentado para reconhecimento automático de alimentos e estimação calórica segue o seguinte processo: é colocado um objeto de referência adjacente aos alimentos e usando uma câmara digital é tirada uma foto à cena. Segue-se um pré-processamento de uma versão da imagem em escala de cinzento, onde são ajustados os valores de intensidade, realizada uma equalização adaptativa do histograma para resolver pequenos problemas de iluminação, e por fim aplicado um filtro de média para eliminar algum ruído da imagem. O *background* da imagem é removido com base nos valores de intensidade usando-se *threshold* adaptativo. De notar que este método só funciona para casos em que o fundo é mais claro que os objetos em primeiro plano. Componentes conectados disjuntos são classificados por meio da aplicação de um algoritmo de 4-Means e em conjunto com um filtro morfológico de fecho (dilatação seguida imediatamente de uma erosão). O resultado obtido (máscara) é depois transposto para a imagem original a fim de se obter a imagem segmentada com os objetos detetados separados do *background*. Um algoritmo de extração de características é executado sobre os objetos derivados da segmentação. Usando-se como valores comparativos as características do objeto de referência, são extraídas da imagem características baseadas em cor (intensidades médias ao longo do espectro vermelho, verde e azul), dimensão e textura (entropia local em diferentes áreas do respetivo objeto). Como modelo de classificação é utilizada uma Rede Neuronal 5-10-4 FFBP (Feed Forward Back Propagation). A primeira camada da rede compreende cinco entradas, cada uma representando um elemento do vetor de características extraídas dos objetos. A camada oculta da rede

por motivos de eficiência e fiabilidade é definida por dez neurónios. A camada de saída é constituída por quatro neurónios, cada um representando uma das quatro classes de alimentos. Na Figura 2 podemos seguir os passos da metodologia proposta, desde o pré-processamento à segmentação da imagem.

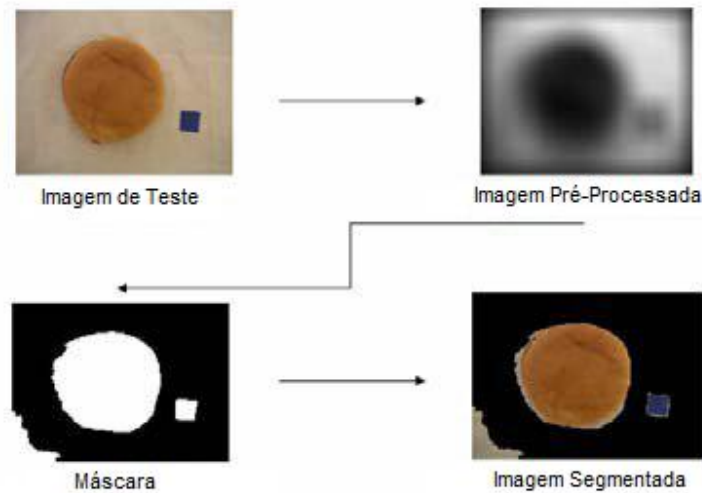


Figura 2: Etapas seguidas do pré-processamento à segmentação da imagem

Em termos de resultados, com a aplicação desta proposta sobre um conjunto de dados com 120 imagens de treino e 20 imagens de teste obtiveram-se valores de exatidão de 75% para Hambúrgueres, 70% para Batatas fritas, 55% para Nuggets de frango e 50% para Tortas de maçã.

O método proposto faz uso de um conjunto de ferramentas simples de visão computacional e *data mining* com o objetivo de realizar reconhecimento automático de alimentos. As imposições do modelo, nomeadamente a necessidade de manter um espaçamento entre os alimentos e utilizar um objeto de referência, assim como o facto de ser necessário garantir que o *background* possui uma cor uniforme mais clara que qualquer alimento presente na cena, retiram flexibilidade ao modelo, pelo que a sua aplicação num contexto real de reconhecimento de refeições é inviável. Refira-se também que mesmo com todas as limitações impostas e inclusive a utilização de apenas quatro categorias de alimentos, os resultados obtidos não são de todo encorajadores à adoção desta base em trabalhos futuros.

2.1.2. Determinação do Tamanho da Porção de Comida

No trabalho apresentado por Mingui Sun et al, “**Determination of Food Portion Size by Image Processing**” [7], é proposta a conjugação de métodos de visão computacional com interação humana para a estimação do valor nutricional e calórico de alimentos. A proposta tem por base a utilização de um cartão de calibração em conjunto com algoritmos para determinar a distância focal da câmara, a distância entre um dado par de pontos e a área de superfície sobre a comida. Estes valores são posteriormente utilizados para determinar o volume dos alimentos.

O método apresentado para determinação do tamanho da porção de comida obedece à seguinte descrição: é colocado um cartão quadriculado adjacente aos alimentos, que funciona como placa de calibração, e usando uma câmara digital é tirada uma foto à cena, captando alimentos e cartão de calibração. Se os alimentos não são completamente consumidos, é tirada uma segunda foto para se proceder a uma subtração entre as quantidades inicial e final. O utilizador identifica os alimentos e mede algumas variáveis geométricas (superfície superior, comprimento e altura). Segue-se uma segmentação automática mais refinada dos alimentos usando um algoritmo de *active contours*. A etapa seguinte consiste na medição do objeto segmentado, o que implica a execução de um processo de calibração da câmara que mapeia um ponto no sistema de coordenadas do mundo (x,y,z) para o índice do pixel (u,v) no plano da imagem. A calibração da câmara exige a determinação da sua distância focal, a qual é estimada a partir de pontos de fuga. O volume de alimentos é calculado tomando-se o produto da área estimada e a altura. Os valores obtidos são passados a uma base de dados da qual se recupera a informação nutricional dos alimentos. Na Figura 3 podemos observar os resultados da segmentação semiautomática de uma imagem de teste.

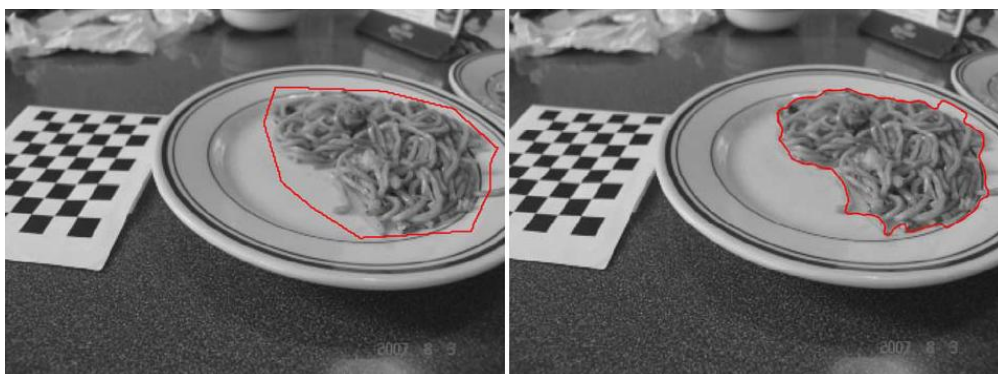


Figura 3: Da esquerda para a direita, início e fim da segmentação da refeição

Os resultados experimentais mostraram que este método consegue uma precisão razoável, com um erro de medição, entre 5% e 8%.

O modelo proposto por Mingui Sun introduz uma estratégia muito interessante para estimação do tamanho da porção de comida, produzindo resultados com uma precisão considerável. Porém a estimação não é totalmente automática, uma vez que é necessário que o utilizador proceda à identificação e delimitação dos alimentos que compõem a refeição.

2.2. Aplicações

Em termos de aplicações disponíveis, embora exista um número razoável de aplicações ligadas à gestão e controlo de dietas, o número de aplicações que realizam reconhecimento automático é bastante reduzido. Neste setor a aplicação para *smarthphones* desenvolvida pela companhia Daily Burn, “**Meal Snap**” [8], inova ao possibilitar o reconhecimento automático de refeições e estimação calórica.

2.2.1. Meal Snap – Calorie Counting Magic

Meal Snap é uma aplicação para *smarthphone*, desenvolvida pela companhia Daily Burn, que permite o reconhecimento de alimentos e estimação calórica para além do *tracking* das refeições realizadas. O funcionamento ao nível interno desta aplicação não é público, pelo que somente é possível descrever o funcionamento da aplicação ao nível do utilizador. Trata-se de uma aplicação de fácil utilização na qual o utilizador recolhe uma fotografia de uma refeição e insere opcionalmente um título que identifica a refeição (a legenda proporciona resultados mais precisos). Após este procedimento a aplicação identifica a refeição e define um intervalo de calorias para a refeição. Na Figura 4 podemos observar a aplicação Meal Snap em funcionamento, e as sucessivas janelas exibidas no decorrer da identificação e estimação calórica de uma taça de cereais e um copo de sumo de laranja.

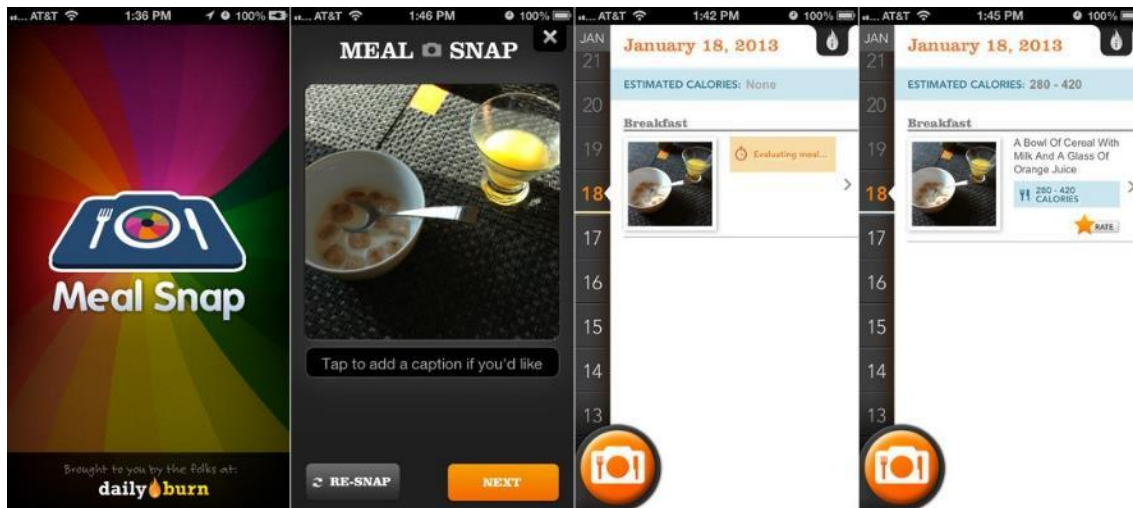


Figura 4: Interface gráfica da aplicação Meal Snap

Ao nível dos resultados, é unânime que a aplicação efetua um reconhecimento de alimentos razoavelmente interessante principalmente com introdução manual de legenda nas imagens recolhidas, porém a estimação calórica não possui uma precisão que permita garantir fiabilidade. Uma referência especial para a análise realizada por Malia Frey, escritora na área da saúde e *personal trainer* certificada, onde menciona que a aplicação Meal Snap consiste numa ferramenta interessante para registar refeições, contudo não sendo recomendada a utilizadores que realmente necessitam obter uma contagem de calorias com precisão [9].

CAPÍTULO III: METODOLOGIAS

CAPÍTULO III: METODOLOGIAS

3.1. Pesquisa de Imagens com Base no Conteúdo

O conceito de Pesquisa de Imagens com Base no Conteúdo (CBIR – Content Based Image Retrieval) foi introduzido por Kato [10] na sua tentativa de procurar imagens automaticamente numa base de dados através de características de cor e forma ao invés de utilizar as tradicionais formas de pesquisa baseadas em texto. Nos sistemas CBIR, as imagens são indexadas pelo seu conteúdo visual (cor, textura, forma) o qual é extraído mediante a utilização de algoritmos de deteção e descrição de características (locais e/ou globais). A utilização de características globais e locais em paralelo permite uma caracterização mais sólida e fiável da imagem [11].

→ Algoritmos de deteção e descrição de características **locais** permitem encontrar correspondências entre imagens, mesmo na existência de grandes mudanças nas condições de visualização, oclusões, e desordem das imagens, além de produzirem uma interessante descrição do conteúdo de cada imagem, possibilitando o reconhecimento da imagem e, portanto, dos objetos representados na imagem.

→ Algoritmos de deteção e descrição de características **globais** funcionam satisfatoriamente quando pretendemos ver a composição da imagem como um todo, em vez dos objetos representados na imagem individualmente (de notar que estes algoritmos misturam a informação que caracteriza os objetos que compõem a cena).

Em CBIR, o processo de pesquisa consiste em dada uma imagem de consulta, calcular a sua similaridade em relação às imagens armazenadas na base de dados de imagens (por comparação entre descritores da imagem de consulta com as imagens da base de dados) e seleccionar as mais similares.

3.2. Detecção e Descrição de Características Locais

Uma característica local é um padrão da imagem que difere da sua vizinhança imediata, geralmente associada com uma alteração de uma ou várias propriedades da imagem simultaneamente, tais como intensidade, cor e textura [12]. Características locais podem ser pontos, bordas ou mesmo pequenos fragmentos da imagem. Para localizar características locais em imagens (pontos de interesse), é necessário analisar vizinhanças locais de pixéis, de modo a dar implicitamente às características locais uma extensão espacial. Normalmente é necessário descrever as características locais de tal forma que as características possam ser identificadas e combinadas. Esta operação requer novamente a utilização de uma vizinhança de pixéis. Ao usar uma vizinhança de pixéis para descrever um ponto de interesse, o processo de extração de características tem que determinar, não apenas a localização do ponto de interesse, mas também o tamanho e, possivelmente, a forma desta vizinhança local. Isto complica consideravelmente o processo, uma vez que o tamanho e forma têm de ser determinados de uma forma invariante. Idealmente, os pontos de interesse devem corresponder (somente) a partes de objetos semanticamente significativas, contudo na prática e atualmente isto não acontece, uma vez que não existe um algoritmo capaz de realizar uma interpretação do conteúdo da imagem de tão alto nível [13].

Um algoritmo de deteção e descrição de características locais deve atender as seguintes propriedades fundamentais [13]:

- **Repetibilidade:** Dadas duas imagens da mesma cena, tiradas em condições de visualização diferentes, uma parte significativa das características visíveis nas imagens devem ser detetadas em ambas as imagens.
 - **Distinção:** Os padrões subjacentes às características detetadas devem mostrar uma grande variabilidade, de modo a possibilitar distinção entre características.
 - **Localidade:** As características devem ser locais, de modo a reduzir a probabilidade de oclusão e permitir aproximações das deformações geométricas e fotométricas através de modelos simples.
-

- **Quantidade:** O número de pontos de interesse deve ser suficientemente grande, de modo a garantir uma descrição mais sólida da imagem.
- **Precisão:** As características detetadas devem ser precisamente localizadas na imagem, incluindo a sua dimensão e forma provável.
- **Eficiência:** Preferencialmente a deteção de características de uma imagem deve ser eficiente ao ponto de possibilitar a utilização do algoritmo em aplicações em que o tempo é um fator crítico.

A repetibilidade é sem dúvida a propriedade mais importante de todas as propriedades mencionadas, e pode ser alcançada de duas formas: por **invariância** ou por **robustez**.

- **Invariância:** Deve ser tolerante a transformações e deformações, ou seja, deve usar métodos para deteção de características que não sejam afetados (dentro do possível) por transformações e deformações.
- **Robustez:** Nos casos de deformações relativamente pequenas, resultantes de ruído, efeitos de discretização, artefactos de compressão, desvios geométricos do modelo matemático utilizado para obter invariância, entre outros, a precisão da deteção pode diminuir, mas não drasticamente.

3.2.1. SIFT: Scale-Invariant Feature Transform

SIFT é um algoritmo de visão computacional usado para detetar e descrever características locais em imagens, sendo formado por duas partes distintas, o **detetor** que é baseado em cálculos de diferenças de Gaussianos e o **descritor** que utiliza histogramas de gradientes orientados para descrever a vizinhança local dos pontos de interesse (pontos-chave).

O funcionamento do SIFT segundo Lowe [14] é definido em quatro etapas principais, sendo que as duas primeiras descrevem a parte do detetor e as duas últimas descrevem a parte do descritor.

- **Detetor:**

- ***Deteção de extremos:*** Consiste na procura de pontos de interesse invariantes a mudanças de escala e orientação da imagem. Isto é conseguido procurando características estáveis em diferentes escalas e orientações, através da utilização de uma função de diferença de Gaussianos.
- ***Localização de Pontos-chave:*** Todos os pontos detetados como extremos são candidatos a pontos-chave. Para cada um dos candidatos é aplicado um modelo detalhado para determinar a localização e escala exatas. O método baseia-se no ajuste de uma função quadrática 3D ao ponto de amostragem local de modo a determinar uma localização interpolada do máximo. Isto é conseguido usando uma expansão de Taylor da função da diferença de Gaussianos aplicada à imagem. Pontos-chave são seleccionados com base nas medidas da sua estabilidade.

- **Descritor:**

- ***Definição da Orientação:*** Para cada ponto-chave são atribuídas uma ou mais orientações, as quais são utilizadas para construção de descritores invariantes. Para cada amostragem da imagem é calculada a magnitude e orientação do gradiente usando as diferenças de pixels, e em seguida é construído um histograma das orientações para os pixels ao redor do ponto-chave. Picos no histograma correspondem a direcções dominantes dos gradientes locais, sendo usados para definir a orientação.
- ***Descrição dos Pontos-chave:*** O descritor de cada ponto-chave é criado computando-se a magnitude e orientação do gradiente na região, definida por 4x4 sub-regiões, em torno da localização do ponto-chave. Uma função Gaussiana é utilizada para dar peso à magnitude e orientação do gradiente. Para cada sub-região é construído um histograma com oito valores de orientação. O descritor resultante tem um tamanho fixo de 128 dimensões.

O algoritmo SIFT tem como aplicações o reconhecimento de objetos, *tracking* de vídeo, reconhecimento de gestos, reconstrução 3D, entre muitas outras [15]. Na Figura 5 podemos ver o resultado da aplicação do SIFT na deteção de pontos de interesse de uma imagem de caldo verde. A dimensão de cada circunferência corresponde à escala do respetivo ponto-chave, e os raios definem a sua orientação. Dependendo da simetria do ponto-chave a determinação da orientação pode ser ambígua, pelo que neste caso podemos ter mais que uma possível orientação.



Figura 5: Aplicação do SIFT para deteção dos Pontos-chave

3.2.2. SURF: Speeded Up Robust Features

SURF é um algoritmo de visão computacional usado para detetar e descrever características locais em imagens. O funcionamento do SURF segundo Bay et al. [16] pode ser dividido em três passos principais:

- São selecionados pontos de interesse em locais distintos da imagem como cantos, manchas, e junções T.
- A vizinhança de cada ponto de interesse é representada num vetor de características (descriptor) distintivo e robusto ao ruído, às deformações geométricas e fotométricas da imagem e aos erros.
- Os descritores são comparados entre diferentes imagens. A comparação é baseada na distância entre vetores.

No SURF a localização dos pontos-chave é conseguida mediante a utilização de um detetor Hessiano. Para cada ponto-chave é determinada a orientação utilizando-se a transformada de Haar (Haar wavelets). Os descritores são formados a partir dos pulsos dos pontos ao redor dos pontos-chave. Uma região quadrada, centrada em torno do ponto de interesse é dividida em 4x4 sub-regiões, com cada sub-região possuindo 5x5 pontos de amostragem regularmente espaçados. Em cada sub-região são somados os pulsos horizontais (d_x), verticais (d_y) e os seus módulos $|d_x|$ e $|d_y|$, formando-se um descritor de quatro dimensões para cada sub-região ($v = (\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|)$), resultando num descritor global de 64 dimensões para o ponto de interesse. Para se obter um descritor de 128 dimensões, o somatório dos pulsos d_x e $|d_x|$ é calculado separadamente para $d_y < 0$ e $d_y \geq 0$, e da mesma forma o somatório dos pulsos d_y e $|d_y|$ é dividido de acordo com o sinal de d_x . De notar que a dimensão do descritor interfere diretamente na eficiência da aplicação, sendo que descritores de menores dimensões proporcionam tempos de execução menores, contudo estes são menos distintivos que descritores de maiores dimensões.

As aplicações do SURF são similares ao SIFT. Na Figura 6 podemos ver o resultado da aplicação do SURF na deteção de pontos de interesse de uma imagem de caldo verde. A dimensão das circunferências correspondem à escala de cada ponto-chave, e os raios definem a sua orientação.



Figura 6: Aplicação do SURF para deteção dos Pontos-chave

3.3. Deteção e Descrição de Características Globais

Um descritor global consiste num elemento que caracteriza a imagem na sua totalidade. Ao nível das características visuais podemos ter descritores que refletem diretamente características específicas tais como cor, textura ou forma. As técnicas mais genéricas usadas para a extração de um descritor que caracterize globalmente a imagem ao nível da cor envolvem normalmente um processo de quantização. A quantização consiste na realização de um mapeamento do espaço de cores para um espaço mais reduzido, e portanto uma redução de informação e detalhe.

3.3.1. Espaços de Cor RGB(A) e HSI

Uma imagem pode ser representada como uma grelha de pixéis, sendo cada pixel definido por um escalar que define a sua Cor. No modelo de Cor RGBA, cada pixel resulta da combinação de três camadas monocromáticas (**Vermelho**, **Verde**, **Azul**) e uma quarta camada de transparência (**Alfa**). Cada um dos componentes é representado por um inteiro de oito bits e assume valores no intervalo [0, 255]. Na Figura 7 podemos ver o cubo que define o espaço de cores RGB.

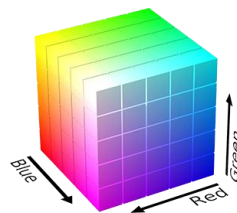


Figura 7: Espaço de Cores RGB

O valor de cada Cor é obtido pela combinação das camadas da seguinte forma:

$$Cor_c = Alfa \ll 24 \mid Vermelho \ll 16 \mid Verde \ll 8 \mid Azul$$

Analogamente, pode-se separar uma Cor nas camadas que a compõem:

$$Alfa = (Cor_c \gg 24) \& 0xFF$$

$$Vermelho = (Cor_c \gg 16) \& 0xFF$$

$$Verde = (Cor_c \gg 8) \& 0xFF$$

$$Azul = Cor_c \& 0xFF$$

Por norma a componente Alfa não é tida em consideração na descrição de cor. O modelo de cor RGB(A) ajusta-se com as curvas de sensibilidade do olho humano, contudo não é o modelo mais eficiente para descrever as cores em termos práticos para o nosso sistema percetivo visual (**Exemplo:** A cor de um feijão é: 60% de Vermelho, 10% de Verde e 30% Azul). Uma alternativa ao espaço de cores RGB consiste no modelo HSI¹, baseado em coordenadas cilíndricas. O espaço HSI permite a separação da cor nas componentes de Matiz, Saturação e Intensidade. Na Figura 8 podemos ver uma representação do espaço de cores HSI.

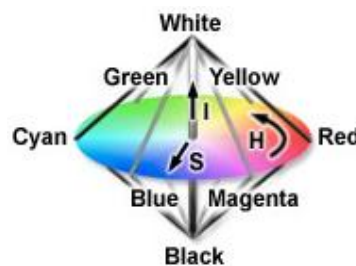


Figura 8: Espaço de Cores HSI

Matiz, Saturação e Intensidade:

- **Matiz:** Medida do comprimento de onda médio da luz refletida ou emitida por um objeto, definindo a sua cor. Este atributo é definido no intervalo $[0, 2\pi]$.
- **Saturação:** Intervalo de comprimentos de onda ao redor do comprimento de onda médio no qual a energia é refletida ou transmitida. Um valor alto de saturação corresponde a uma cor pura, ao passo que um valor baixo indica uma maior diluição da cor. Este atributo é definido no intervalo $[0, 1]$;
- **Intensidade:** Medida da energia total envolvida em todos os comprimentos de onda, sendo responsável pela sensação de brilho. Definida no intervalo $[0, 1]$;

Por ser um modelo mais natural e intuitivo para a percepção humana, o modelo HSI é ideal para o desenvolvimento de algoritmos de processamento de imagem baseados na descrição de cores. A conversão entre o modelo RGB e o modelo HSI pode ser consultado no **ANEXO A: Conversão entre espaço de Cor RGB e HSI**.

¹ Referência para dois modelos muito comuns, baseados em coordenadas cilíndricas, nomeadamente HSL (Matiz, Saturação e Luminosidade) e HSB (Matiz, Saturação e Brilho). Luminosidade é a denominação dada à capacidade que possui qualquer cor de refletir a luz. Brilho é a percepção provocada pela luminosidade de um alvo visual.

3.3.2. Histograma de cor RGB

Na construção convencional de um histograma de cor [17], é efetuado um mapeamento do espaço de cores para um espaço com dimensões menores, porém com um tamanho suficiente grande para se conseguir uma descrição consistente da cor da imagem. A construção do histograma de cor é conseguida pela combinação para cada pixel dos valores quantizados das camadas de Vermelho, Verde e Azul num inteiro. O valor obtido corresponde ao índice da coluna do histograma a incrementar. Supondo que se pretende usar três bits para representar as camadas de Vermelho e Verde, e dois bits para representar a camada de Azul, temos: (*descritor com* $2^{(3+3+2)} = 256$ *dimensões*)

$$\text{Índice} = (Cor_c \gg 16) \& 0xE0 \mid (Cor_c \gg 11) \& 0x1C \mid (Cor_c \gg 6) \& 0x3$$

Por forma a verificar a degradação na imagem, resultante da quantização do espaço de cores, podemos aplicar o seguinte mapeamento a cada pixel:

$$Cor_{Final} = Cor_{Inicial} \& 0xFF\alpha\beta\gamma$$

Onde os fatores de oito bits α , β e γ estão dependentes da dimensão desejada para o espaço de cores final. Podemos observar na Figura 9 a quantização para espaços de cores de tamanhos **512** (0xFFE0E0E0), **256** (0xFFE0E0C0) e **64** (0xFFC0C0C0).



Figura 9: Esquerda para direita e cima para baixo, 53.976 (original), 512, 256 e 64 cores

3.3.3. Histograma de cor HSI

Nesta abordagem o primeiro passo consiste na conversão do espaço de cor RGB para o espaço HSI. Considerem-se os valores de Matiz, Saturação e Intensidade quantizados e normalizados no intervalo [0, 255]. A construção do descritor de cor é conseguida pela combinação para cada pixel destes valores num inteiro, correspondente ao índice da coluna do histograma a incrementar. Supondo a quantização dos valores de Matiz para 16 níveis (4 bits), Saturação para 4 níveis (2 bits) e Intensidade para 4 níveis (2 bits), temos:

$$\text{Índice} = \text{Matiz} \& 0xF0 \mid (\text{Saturação} \gg 4) \& 0xC \mid (\text{Intensidade} \gg 6) \& 0x3$$

Podemos analisar a degradação na imagem, resultante da quantização do espaço de cores, quantizando individualmente os valores de Matiz, Saturação e Intensidade para cada pixel e em seguida reconvertendo para o espaço de cor RGB. Na Figura 10 vemos o efeito da quantização para **512** (16 níveis de Matiz, 8 níveis de Saturação e 4 níveis de Intensidade), **256** (16 níveis de Matiz, 4 níveis de Saturação e 4 níveis de Intensidade) e **64** (4 níveis de Matiz, Saturação e Intensidade) valores respetivamente.



Figura 10: Esquerda para direita e cima para baixo, 49.655 (original), 512, 256 e 64 valores

3.4. Bag of Words

O processo de extração de descritores resulta na obtenção de um elevado número de descritores por imagem, pelo que realizar uma comparação descritor a descritor não é computacionalmente viável principalmente para uma base de dados de tamanho considerável. Uma abordagem para atacar este problema consiste na utilização de um *Bag of Words* [18]. Neste método é escolhido um conjunto de N descritores de referência que são usados como base para a construção de um histograma de N *bins* (palavras) para cada imagem da base de dados. Desta forma a pesquisa de imagens da base de dados é feita mediante a comparação entre pares de histogramas (ao invés de se comparar pares de descritores).

→ **Construção do Bag of Words:** Para o correto funcionamento desta abordagem é necessário garantir que o núcleo de descritores escolhidos para servir de base aos processos subsequentes é representativo do contexto que se pretende identificar. Esta amostragem pode ser realizada usando um algoritmo simples de clusterização como o K-Means que agrupa os descritores em K clusters, sendo o centróide de cada cluster uma palavra do *Bag of Words*.

→ **Construção dos Histogramas:** Cada um dos descritores do *Bag of Words* corresponde a um *bin* no histograma. Para cada imagem, é comparado cada um dos seus descritores com cada descritor do *Bag of Words*, sendo incrementado o *bin* associado ao descritor para o qual a comparação resulta na menor distância.

Após a construção do *Bag of Words* e cálculo dos respetivos histogramas para cada imagem da base de dados é possível proceder à pesquisa de imagens com base no conteúdo, usando uma imagem de consulta. A pesquisa de imagens implica a construção de um histograma para a imagem de consulta, e a sua comparação com cada um dos histogramas das imagens da base de dados.

3.4.1. Clusterização com K-Means

O algoritmo de clusterização K-Means [19] é utilizado em *data mining* para particionar um conjunto de n observações em k clusters, sendo cada observação atribuída ao cluster mais próximo.

O funcionamento do K-Means segue as seguintes fases:

Inicialização: São escolhidos aleatoriamente K elementos (descritores). Cada elemento é o centróide de uma classe;

Iteração:

- (1) É calculada a distância entre cada elemento e todos os centróides;
- (2) Cada elemento é adicionado à classe do centróide com a menor distância;
- (3) Para cada classe é recalculado o respetivo centróide;
- (4) Processo é repetido até se atingir convergência;

O K-Means é um algoritmo bastante simples e apresenta normalmente resultados satisfatórios, porém é necessário ter em consideração que a elevada densidade de elementos por cluster e a aleatoriedade inicial na escolha dos centróides podem ser em determinadas situações desvantajosas.

CAPÍTULO IV: ALGORITMOS

CAPÍTULO IV: ALGORITMOS

4.1. Pesquisa de Imagens com Base no Conteúdo

O processo padrão de pesquisa de imagens com base no conteúdo segue uma série de etapas sequenciais, nas quais se englobam deteção e descrição de características locais e/ou globais de cada imagem, construção de um *Bag of Words* e por fim pesquisa das imagens em base de dados usando uma imagem de consulta e seleção das imagens que apresentaram uma maior semelhança à imagem de consulta. Na Figura 11 podemos observar o *pipeline* de operações realizadas, envolvendo as fases de construção da base de dados, identificação de refeições e estimação calórica.

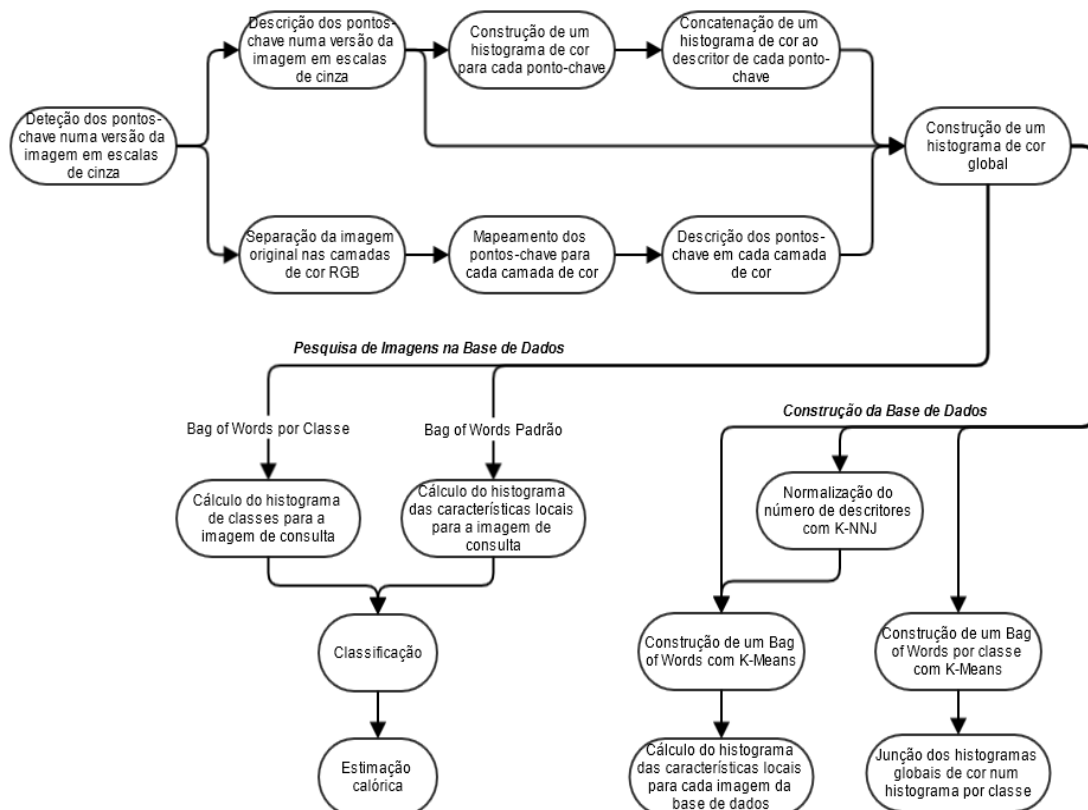


Figura 11: Pipeline de operações na construção da BD e pesquisa de imagens

4.1.1. Detecção e Descrição de Características Locais

Na deteção e descrição de características locais das imagens são utilizados os algoritmos SIFT (Scale-invariant feature transform) e SURF (Speeded Up Robust Features), os quais são invariantes a redimensionamento e a mudanças de orientação e parcialmente invariantes a adição de ruído e mudanças de iluminação. Ambos os algoritmos aplicados são disponibilizados pela biblioteca de visão computacional OpenCV. OpenCV (Open Source Computer Vision) [20] é uma biblioteca de programação multiplataforma, de código aberto e totalmente livre tanto para uso académico como comercial. O OpenCV possui módulos de estruturas de dados, álgebra linear, interfaces gráficas e implementa uma variedade de ferramentas de processamento e interpretação de imagens que possibilitam operações tais como aplicação de filtros, segmentação, extração de descritores, reconhecimento de objetos, entre muitas outras. A biblioteca OpenCV foi idealizada com a finalidade de tornar a visão computacional acessível a utilizadores e programadores em áreas tais como a interação pessoa-máquina, reconhecimento facial, reconhecimento gestual, controlo de movimento, robótica, realidade aumentada, entre muitas outras áreas.

Os algoritmos aplicados, SIFT e SURF, não têm em consideração a informação da cor nem ao nível da deteção nem ao nível da descrição dos pontos de interesse. A deteção dos pontos-chave é realizada sobre uma versão em escala de cinzento da imagem original, e a descrição dos pontos-chave é igualmente realizada usando como base a imagem em escala de cinzento (e normalmente usando a mesma vizinhança de píxeis utilizada na deteção). Na Figura 12 podemos ver o *pipeline* da deteção e descrição de características locais em imagens.

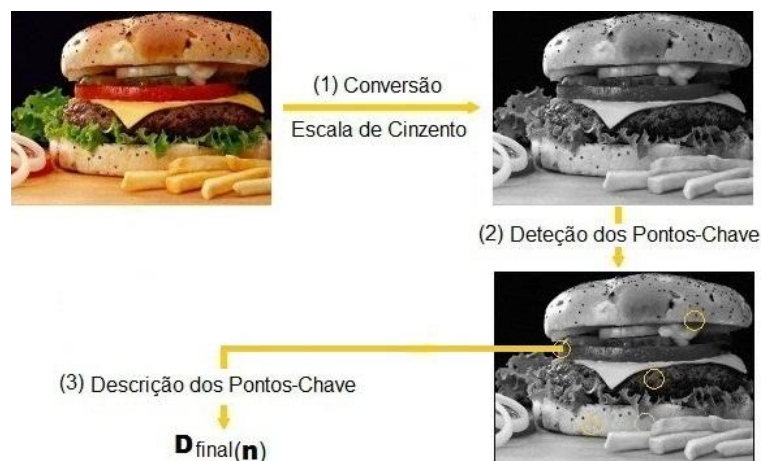


Figura 12: Pipeline da deteção e descrição de características locais

Porém, a cor é indubitavelmente uma propriedade importante nas imagens, e ganha ainda maior relevo perante imagens de refeições. Podemos facilmente verificar que existem diversos tipos de alimentos que possuem a mesma textura e inclusive a mesma forma, como por exemplo, no caso das várias qualidades de feijão, o que torna a informação de cor imprescindível para a sua correta classificação. Portanto, tendo em consideração as características físicas dos alimentos, consideramos que a adição de informação de cor nos descritores locais pode contribuir para um melhoramento na discriminação e/ou categorização dos alimentos. Assim com o objetivo de introduzir informação de cor na descrição de características locais, propomos a análise do desempenho de duas abordagens distintas: *descrição dos pontos-chave para cada camada de cor da imagem* que implica a separação da imagem original nas camadas de cor RGB e mapeamento dos pontos-chave para as localizações homólogas em cada uma das camadas e *concatenação de um histograma de cor a cada descritor base*, que implica a utilização de informação sobre a escala de cada ponto-chave para delimitação da área de interesse e subsequente construção de um histograma de cor e concatenação do histograma ao descritor.

4.1.1.1. Descrição dos Pontos-chave em cada Camada de Cor da Imagem

A deteção dos pontos-chave da imagem é executada da forma padrão sobre a imagem em escala de cinzento. Após termos a localização dos pontos de interesse, separamos a imagem original nas três camadas de cor que a constituem (Vermelho, Verde e Azul) e mapeamos a posição dos pontos de interesse para a localização homóloga em cada uma das camadas de cor. Em seguida, para cada ponto de interesse é obtida a sua descrição em cada uma das camadas de cor, resultando desta operação três descritores, os quais são concatenados num descritor final de 384 dimensões. Com esta metodologia temos à partida alguns fatores negativos a considerar. Embora se consiga de certa forma incluir informação de cor nos descritores, a deteção dos pontos de interesse é realizada sobre a imagem em escala de cinzento, ou seja, na deteção dos pontos de interesse é descartada toda a informação de cor. Outro elemento a ter em consideração consiste no agravamento do custo computacional resultante da triplicação do tamanho de cada descritor. Este aumento de dimensões traduz-se diretamente tanto na triplicação do custo ao nível do processamento como de memória.

Podemos observar na Figura 13 o *pipeline* de operações na descrição dos pontos-chave em cada camada de cor.

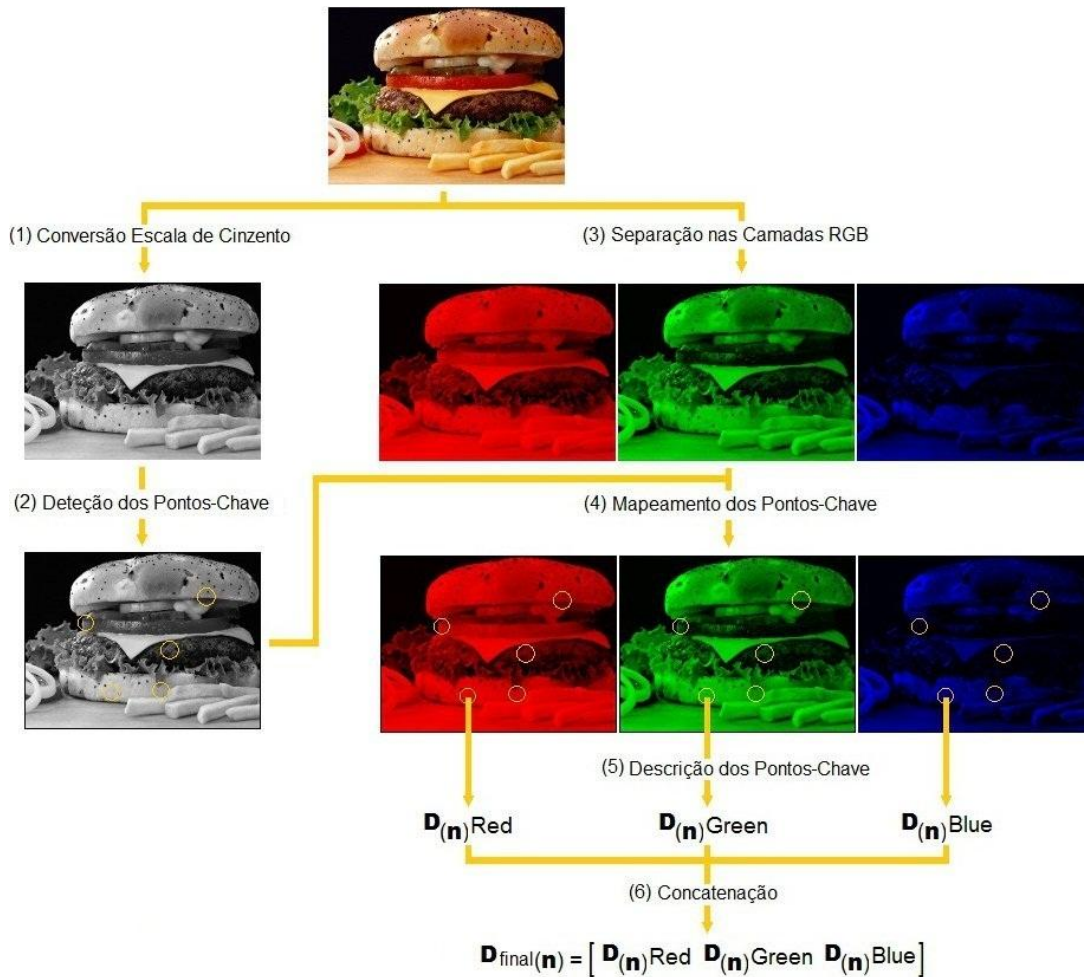


Figura 13: Descrição dos pontos-chave em cada camada de cor da imagem

4.1.1.2. Concatenação de Histograma de Cor ao Descritor Base

Uma vez determinados os pontos de interesse, e sabendo as coordenadas de cada ponto e o diâmetro da circunferência que o envolve, i.e. a área considerada relevante, podemos tirar partido destes elementos para extrair informação da área em torno do ponto de interesse que o caracterize ao nível da cor. Como na generalidade, para cada imagem é detetado um elevado número de pontos-chave, determinar a equação da circunferência para cada ponto-chave e verificar quais pixéis se encontram abrangidos pela circunferência é computacionalmente custoso. Assim, optamos por considerar para cada ponto-chave a área abrangida por um quadrado com lado igual ao diâmetro da circunferência para construção do histograma de cor. O histograma resultante é concatenado ao descritor base proporcionando assim descrição da cor.

Com este método conseguimos incluir informação de cor nos descritores, porém, da mesma forma que com o procedimento anterior, a deteção dos pontos de interesse é realizada sobre a imagem em escala de cinzento, ou seja, na deteção dos pontos de interesse é descartada toda a informação de cor. Note-se também a importância na escolha do fator de quantização na construção do histograma de cor. Uma quantização demasiado severa pode induzir uma perda de diferenciação entre os descritores ao nível da cor, por outro lado uma quantização demasiado flexível para além de provocar um aumento considerável do tamanho de cada descritor (com o respetivo agravamento ao nível do processamento e memória), pode provocar um aumento considerável da semelhança entre os descritores ao nível da cor devido à área abrangida por cada descritor ser relativamente reduzida. Na Figura 14 podemos visualizar o *pipeline* de operações realizadas na adição de um histograma de cor a cada descritor base.

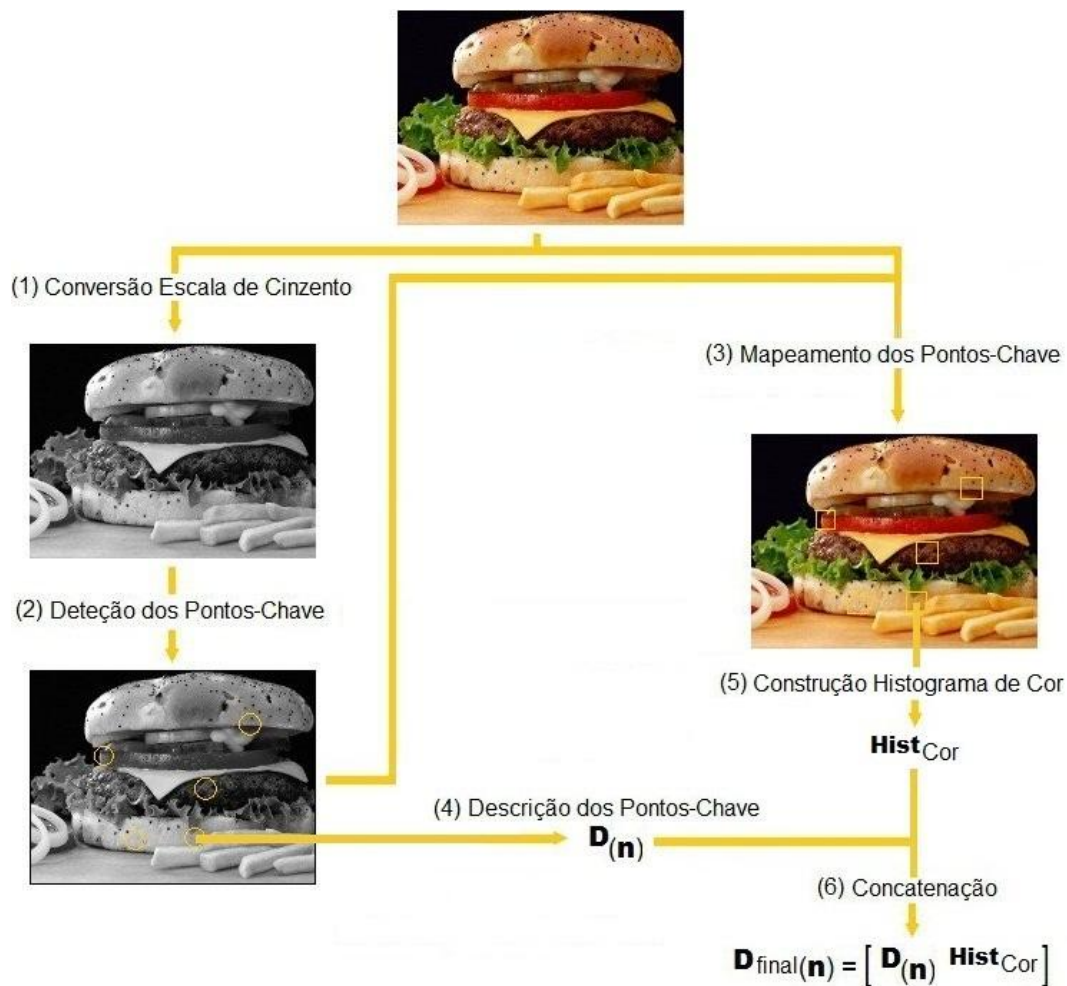


Figura 14: Concatenação de histograma de cor ao descritor base

4.1.2. Detecção e Descrição de Características Globais

Ao nível das características globais consideramos somente a descrição da informação da cor sob a forma de um histograma de cor. A motivação na construção de um descritor global da cor da imagem resulta do fato de por um lado na descrição de características locais a descrição de informação de cor não ser uma característica intrínseca dos algoritmos aplicados, SIFT e SURF, e por outro lado os modelos apresentados para inclusão de informação de cor ao nível local terem as mencionadas limitações. Na construção do histograma de cor são considerados os espaços de cor RGB e HSI, seguindo um procedimento similar ao descrito nas secções 3.3.2 e 3.3.3 respetivamente.

4.1.3. Bag of Words

Numa abordagem *Bag of Words* padrão é realizada uma amostragem de N elementos do universo total, onde neste caso cada elemento consiste num descritor. A amostragem é executada de forma que os elementos selecionados permitam caracterizar o universo total, isto é, a base de dados. Seguindo um procedimento conforme o descrito na secção 3.4, estamos a misturar todos os descritores de todas as classes de refeições num único universo de descritores ao qual se segue uma amostragem. Desta forma perde-se a informação sobre as classes, para além das pesquisas de imagens implicarem a comparação da semelhança entre a imagem de consulta e cada uma das imagens da base de dados. Uma vez que temos disponível à partida uma categorização das imagens das refeições (a categorização é realizada manualmente através de anotações sobre cada uma das imagens da base de dados), é interessante aproveitar esta situação no reconhecimento automático de imagens, através da utilização de um *Bag of Words* por classe. Outro fator a ter em consideração reside na forma como é feita a amostragem dos descritores. Uma prática genérica consiste na utilização do algoritmo de clusterização K-Means, contudo a elevada densidade de elementos por cluster em conjunto com a diferença no número de descritores extraídos entre imagens de diferentes refeições pode influenciar negativamente os resultados. Portanto, consideramos que no caso do *Bag of Words* padrão uma normalização do número de descritores por imagem pode trazer melhores resultados. Assim propomos a análise do ganho de utilizar um *Bag of Words* que caracterize cada classe de refeições, e a normalização do número de descritores para cada imagem com K-Nearest Neighbor Joining.

4.1.3.1. Bag of Words por Classe

A ideia geral deste procedimento consiste em gerar um *Bag of Words* para cada classe de refeições, ou dito de outra forma, para cada tipo específico de refeição fazer uma amostragem do universo de descritores definido pela aglomeração dos descritores de todas as imagens que representam esse mesmo tipo de refeição, conseguindo-se assim caracterizar a classe da refeição. Por exemplo, construir um *Bag of Words* que caracteriza a classe de refeições “Bolonhesa”, outro *Bag of Words* que caracterize a classe “Lasanha”, e assim adiante para cada classe de refeições, por forma a obter uma caracterização independente para cada classe. A pesquisa de uma imagem na base de dados passa assim a implicar a comparação da imagem de consulta com cada uma das classes de refeições (ao invés de se comparar diretamente com cada uma das imagens de refeições da base de dados). Em detalhe este procedimento exige a comparação de cada descritor da imagem de consulta com cada um dos descritores presentes nos *Bag of Words* que definem cada uma das classes. O resultado final consiste num **histograma de classes** onde cada *bin* corresponde a uma classe. Como exemplo, supondo $\mathbf{d}_c$ um descritor da imagem de consulta, e $\mathbf{d}_\tau$ o descritor mais semelhante ao primeiro e incluído no *Bag of Words* da classe τ , o *bin* do histograma correspondente à classe τ é portanto incrementado em uma unidade. Note-se também na possibilidade de efetuar uma amostragem dos descritores da imagem de consulta por forma a reduzir o número de comparações subsequentes a realizar.

4.1.3.2. Normalização do número de descritores com K-NNJ

Como vimos anteriormente, o algoritmo K-Means é uma prática padrão para o particionamento de um conjunto de observações por K clusters, e, portanto aplicado diretamente neste tipo de abordagens baseadas em *Bag of Words*. Porém seguindo uma abordagem de *Bag of Words* padrão, a aplicação do algoritmo K-Means resulta na obtenção de *clusters* com uma elevada densidade de elementos, o que combinado com o facto de existir uma diferença considerável entre o número de descritores extraídos em imagens de diferentes classes de refeições pode contribuir negativamente na construção do *Bag of Words*, resultando em palavras que não caracterizam corretamente o contexto que se pretende descrever. Como exemplo, consideremos que temos três imagens de refeições, nomeadamente hambúrguer, pizza e sopa, cada imagem com um determinado número de pontos-chave associados. Supondo que o número de pontos-chave detetados na imagem de sopa

em comparação com as restantes imagens é claramente inferior. Então, espera-se que usando 3-Means, os descritores dos pontos-chave da imagem de sopa não tenham um peso relevante na escolha das palavras do Bag of Words, resultando em dificuldades posteriores na classificação de imagens de sopa. Na Figura 14 podemos observar uma ilustração simplificada daquilo que se pretende demonstrar, onde cada símbolo corresponde a um descritor de uma das três refeições.

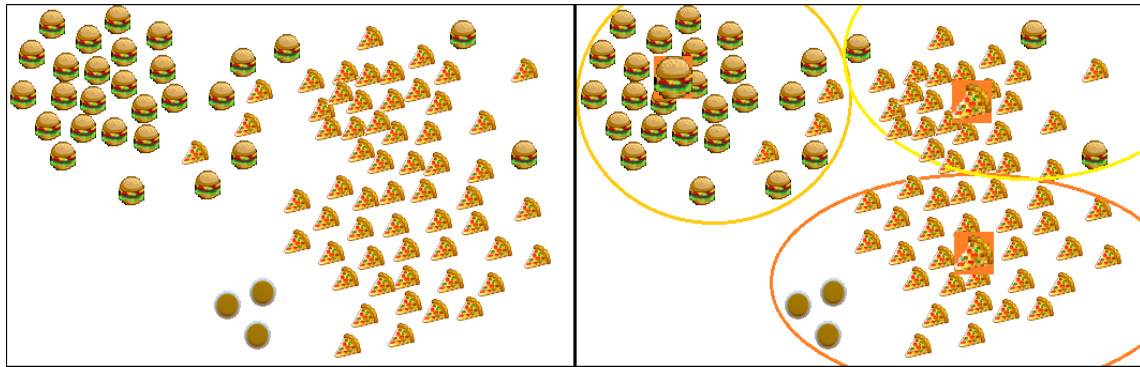


Figura 15: Clusterização de descritores. Circunferências correspondem aos clusters

Uma normalização do número de descritores por imagem com K-Means não é uma solução positiva devido à forma de operação do algoritmo. Na clusterização de grandes conjuntos de dados é espetável que os centróides se situem nos grandes aglomerados, o que faz com que os (possíveis) *outsiders* percam importância. Isto poderá ser um fator negativo para imagens de refeições compostas por múltiplos alimentos, se não existir um equilíbrio entre o número de pontos-chave detetados em cada alimento. A título de exemplo, consideremos que temos uma imagem de Esparguete com Atum e outra de Esparguete com Carne. Se o número de pontos-chave detetados no esparguete for muito superior aos detetados no atum e na carne, existe uma probabilidade significativa de com a normalização se perder capacidade de discriminar corretamente estas duas imagens.

Assim, propomos a utilização do algoritmo K-Nearest Neighbor Joining [21], com o objetivo de normalizar o número de descritores por imagem de forma a obter uma maior igualdade entre as várias classes de refeições no momento da seleção das palavras que irão representar a base de dados. Referência também para a redução do gasto computacional nos processos subsequentes conseguido pela redução do número de descritores por imagem.

O funcionamento do algoritmo segue os seguintes passos:

Inicialização: Cada descritor define uma classe na qual é o centróide. São calculadas e armazenadas as distâncias entre cada descritor e os restantes descritores;

Iteração:

- (1) O par de centróides com menor distância entre si é agrupado numa única classe, e é (re)calculado o centróide da classe;
- (2) São calculadas as distâncias entre o novo centróide e os restantes;
- (3) Processo é repetido até sobraem somente K classes;

Uma aplicação direta deste algoritmo a um conjunto de descritores pertencente a uma classe de refeições poderá ser inviável se o número de imagens da classe for grande, devido ao número total de descritores ser provavelmente bastante elevado, pelo que a sua aplicação deve ser realizada individualmente para cada imagem ou em pequenos conjuntos de imagens. Após se normalizar o número de descritores por imagem ou classe de imagens com K-Nearest Neighbor Joining, podemos proceder à aplicação do algoritmo de clusterização K-Means para se encontrar as palavras do *Bag of Words*. O valor K escolhido para K-NNJ é independente do valor K' definido para K-Means. De notar que a aplicação do K-Nearest Neighbor Joining para determinação das palavras do *Bag of Words* resultaria muito provavelmente na seleção de um número considerável de *outliers* para palavras do *Bag of Words*, traduzindo-se numa errada representação da base de dados.

4.1.3.3. Classificação

Com aplicação de uma metodologia de *Bag of Words* padrão, o resultado é um histograma para cada imagem, o qual pode ser usado como meio de comparação por forma a encontrar imagens similares entre si. Assim, podemos comparar o histograma da imagem de consulta com o histograma de cada imagem da base de dados e selecionar o conjunto de imagens mais semelhantes à imagem de consulta. Por fim, podemos determinar as classes que possuem maior probabilidade de corresponderem à classe da imagem de consulta recorrendo a *Majority Vote*, selecionando e exibindo de forma ordenada as classes com mais representantes no conjunto de imagens com melhor classificação. A utilização de *Majority Vote*, para selecionar as classes mais prováveis, não tem em consideração a informação de classificação das imagens mais prováveis na ordenação das classes. Assim, supondo que temos n imagens que foram consideradas as mais semelhantes à imagem de consulta, das quais p imagens correspondem à classe τ e q imagens correspondem à classe β , sendo p ligeiramente menor que q . Mesmo que as imagens pertencentes à classe τ sejam em termos de semelhança à imagem de consulta claramente superior em comparação com as imagens da classe β , a classe mais provável será a classe β . Para contornar este problema, propomos um modelo alternativo de ordenar as classes mais prováveis baseado num sistema de pontos. Com base no valor da distância entre a imagem mais semelhante e a menos semelhante à imagem de consulta, os valores de semelhança são normalizados segundo uma escala. Para cada classe representada no conjunto de imagens mais prováveis, a sua pontuação irá corresponder à soma dos valores de semelhança normalizados das imagens pertencentes à classe. Finalmente, as classes são ordenadas segundo a pontuação obtida.

Seguindo uma abordagem baseada em *Bag of Words* por Classe, o resultado é um histograma que indica a probabilidade da imagem de consulta corresponder a cada uma das classes de refeições. Os picos no histograma correspondem às classes mais favoráveis. Assim, com base na normalização da diferença entre a classe mais provável (classificada com maior probabilidade) e a classe menos provável (com pior classificação) são exibidas de forma ordenada as classes com maior probabilidade de corresponderem à classe da imagem de consulta.

4.2. Estimação do Valor Calórico

Propomos a estimação do valor calórico de uma refeição pela comparação entre o número de descritores locais da imagem de consulta e o número médio de descritores do conjunto de imagens da base de dados pertencentes à mesma classe da imagem de consulta. Como exemplo, consideremos que temos na base de dados uma imagem de um croquete com 86 calorias (valor calórico de referência), pré-processada pelo algoritmo de detecção e extração de características locais SURF², que resultou na detecção e extração de 980 descritores. Na Figura 16 podemos observar o resultado da aplicação do algoritmo SURF na detecção de pontos de interesse.



Figura 16: Detecção e Extração de 980 Descritores usando SURF

Agora supondo uma segunda imagem em que temos o dobro da quantidade de alimentos da imagem anterior. Podemos verificar na Figura 17 o resultado de aplicar o algoritmo SURF sobre uma imagem com aproximadamente o dobro da proporção dos alimentos da Figura 16.



Figura 17: Detecção e Extração de 1985 Descritores usando SURF

² A escolha do algoritmo SURF em detrimento do SIFT não teve como base nenhum fator relevante, note-se que são esperados resultados semelhantes com a utilização de qualquer dos dois algoritmos.

A aplicação do SURF resultou na deteção e extração de 1985 descritores. Assim, e sabendo que 980 descritores correspondem a 86 calorias (para o caso específico dos croquetes quando analisados com o algoritmo SURF), podemos fazer uma estimativa do valor calórico da imagem em análise:

$$Calorias \cong \frac{86 \times 1985}{980} = 174$$

Temos que a estimativa obtida são 174 calorias, que corresponde a aproximadamente duas vezes o valor calórico de referência de um croquete.

Esta abordagem é extremamente simples e permite a obtenção de uma estimativa do valor calórico para cenários igualmente simples, porém em cenários mais complexos em que, por exemplo, existe sobreposição/oclusão de alimentos, as estimativas não são de todo precisas. Note-se também que seguindo esta metodologia, quando se efetua uma estimativa calórica de imagens de refeições completas, sem segmentar a imagem nos vários alimentos que compõem a refeição, parte-se do princípio que embora as dimensões da refeição possam variar, a proporção entre os vários alimentos são mantidos de forma aproximadamente constante.

CAPÍTULO V: APLICAÇÃO

CAPÍTULO V: APLICAÇÃO

5.1. Visão Geral do Projeto

Podemos estruturar o projeto em três fases distintas interligadas entre si: construção de uma base de dados de imagens de refeições, a pesquisa de imagens na base de dados com base no conteúdo, e a estimação do valor calórico da refeição. Para auxiliar nestas tarefas desenvolveu-se uma interface gráfica de suporte, desenhada para ser um primeiro esboço para interface de uma aplicação de reconhecimento e estimação calórica de refeições para dispositivos móveis.

5.2. Construção de Base de Dados

A estratégia de pesquisa de imagens com base no conteúdo é iniciada pela construção da base de dados (DB), portanto começamos pela deteção e descrição de características locais de cada imagem a incluir na DB. Para o efeito, testamos o desempenho dos algoritmos SIFT e SURF ambos disponibilizados pela biblioteca de visão computacional OpenCV, os quais proporcionam invariância a redimensionamento e a mudanças de orientação e invariância parcial a adição de ruído e mudanças de iluminação. Ao nível do algoritmo SURF, valorizando a precisão em detrimento da eficiência usamos a opção intrínseca de duplicar o tamanho do descritor de 64 para 128 dimensões.

Por forma a introduzir informação de cor na descrição de características locais, analisamos o ganho das abordagens descritas em 4.1.1, nomeadamente descrição dos pontos-chave para cada camada de cor da imagem e concatenação de um histograma de cor a cada descritor base, em comparação com a abordagem padrão em que não é usada informação de cor. Para a construção dos histogramas de cor consideramos o modelo RGB com histogramas de tamanhos 8 ($R_1G_1B_1$)³ e 32 ($R_2G_2B_1$) e o modelo HSI com histogramas de tamanhos 16 ($H_2S_1I_1$)⁴ e 32 ($H_3S_1I_1$). Os tamanhos para os histogramas de cor foram escolhidos de forma que a componente

³ $R_\alpha G_\beta B_\gamma$: Histograma RGB com α , β e γ bits para representar as camadas R, G e B respetivamente.

⁴ $H_\alpha S_\beta I_\gamma$: Histograma HSI quantizado a α níveis de matiz, β níveis de saturação e γ níveis intensidade.

de cor não tivesse um peso demasiado elevado no descritor, ao ponto de eclipsar a informação dada pelos algoritmos de deteção e extração de características locais. Em complemento à adição de informação de cor localmente, procedemos à extração de um descritor global ao nível da cor conforme definido em 3.3. Para a construção dos histogramas de cor consideramos o modelo RGB com histogramas de tamanhos 64 ($R_2G_2B_2$), 256 ($R_3G_3B_2$) e 512 ($R_3G_3B_3$) e o modelo HSI com histogramas de tamanhos 64 ($H_3S_2I_1$), 256 ($H_4S_2I_2$) e 512 ($H_4S_3I_2$).

Posteriormente, seguimos uma abordagem *Bag of Words* com o objetivo de efetuar uma seleção de um conjunto de descritores representativo da base de dados. Nesta fase analisamos o desempenho das abordagens:

- ***Bag of Words Padrão:*** Utilizando o algoritmo de clusterização K-Means selecionamos K descritores do universo global de descritores locais da base de dados. Avaliamos o desempenho de 50-Means, 100-Means e 200-Means.
- ***Bag of Words por Classe:*** Dividimos a base de dados por classes, e utilizando o algoritmo de clusterização K-Means selecionamos K descritores para cada classe de imagens de refeições. Avaliamos o desempenho de 50-Means e 100-Means.

Ao nível do *Bag of Words* padrão, e por forma a normalizar o número de descritores locais por classe de refeições, analisamos as vantagens de aplicar um algoritmo K-Nearest Neighbor Joining, para K igual a 150, individualmente ao conjunto de descritores de cada imagem. Ao resultado obtido segue-se a aplicação do algoritmo de clusterização K-Means para selecionar as palavras do *Bag of Words*. A última fase da construção da base de dados consiste na construção de um histograma para cada imagem da base de dados, onde cada *bin* do histograma correspondente a uma das palavras do *Bag of Words*. Para cada imagem, é comparado cada um dos seus descritores locais com cada palavra do *Bag of Words*, sendo incrementado o *bin* associado à palavra para a qual a comparação resulta na menor distância. De todo este processamento, resulta um histograma das características locais e um histograma de cor global para cada imagem da base de dados.

Com *Bag of Words* por Classe a última fase da construção da base de dados consiste em para cada classe juntar os descritores globais de cor (histograma de cor) de cada uma das imagens num único descritor. A junção é realizada através de uma média ponderada entre os histogramas. De todo este processamento resulta um *Bag of Words* e um histograma global de cor por cada classe.

5.3. Pesquisa de Imagens com Base no Conteúdo

Após a construção da base de dados podemos prosseguir com a pesquisa de imagens na base de dados usando uma imagem de consulta. A primeira fase deste processo consiste em, da mesma forma que com as imagens da base de dados, detetar e descrever os pontos de interesse e construir um descritor global de cor. Conforme na construção da base de dados, é seguida uma abordagem *Bag of Words* padrão ou por classe, temos a seguinte metodologia:

→ ***Bag of Words* Padrão:** É construído um histograma das características locais para a imagem de consulta, e em seguida os histogramas das características locais e da cor global são comparados com os respetivos histogramas de cada uma das imagens da base de dados. Ao histograma das características locais é dado um peso de 80% e ao histograma de cor global é dado um peso de 20%. Os pesos foram definidos usando como critério a relevância de cada histograma. São selecionadas as 15 imagens mais semelhantes à imagem de consulta, e usando os métodos definidos em 4.1.3.3 são escolhidas as classes com maior probabilidade de serem a classe da imagem de consulta.

→ ***Bag of Words* por Classe:** É construído um histograma de classes para a imagem de consulta. O histograma de cor global é comparado com o respetivo histograma de cor de cada uma das classes de refeições na base de dados. Ao histograma de classes é dado um peso de 80% e ao histograma de cor global é dado um peso de 20%. Os pesos foram definidos usando como critério a relevância de cada histograma. Por fim são selecionadas as classes com maior probabilidade de serem a classe da imagem de consulta.

5.4. Estimação do Valor Calórico

A estimação do valor calórico da refeição é, conforme descrito em 4.2, realizada por meio de comparação do número de descritores locais extraídos da imagem de consulta e o número médio de descritores extraídos das imagens da base de dados pertencentes à mesma classe da imagem de consulta. De referir novamente que esta técnica apenas apresenta resultados aceitáveis em situações simples, onde se mantém a proporção entre os alimentos que compõem a refeição e não existem muitas oclusões.

5.5. Interface Gráfica de Utilizador

Utilizando o Java GUI *widget toolkit* Swing desenvolveu-se uma interface gráfica para auxiliar na interação com os algoritmos desenvolvidos para CBIR e estimação calórica. Para além do objetivo principal de suporte, a interface gráfica foi idealizada para servir como um primeiro esboço para uma interface de uma aplicação de reconhecimento e estimação calórica de refeições para dispositivos móveis. Assim no seu desenvolvimento foram tidos em consideração os seguintes conceitos de *design*:

- **Percetibilidade:** Numa aplicação que se pretende que seja de simples utilização, é imperativo que o utilizador consiga deduzir facilmente a partir da estrutura visual como a aplicação deve ser usada. Para isto contribui uma interface intuitiva com aspeto apelativo.
 - **Mapeamento:** Simplificação dos controles e aumento da expressividade das ações garante um melhor estabelecimento de relações entre as ações e os seus efeitos na aplicação.
 - **Feedback:** Mudanças visuais na aplicação permitem ao utilizador acompanhar de forma contínua o resultado das suas ações assim como o estado da aplicação.
 - **Visibilidade:** Uma redução dos controles e menus disponíveis para os estritamente necessários, associado a uma exibição simplificada do sistema resulta numa boa visibilidade das funções presentes.
-

A aplicação desenvolvida corre em qualquer PC que disponha de um sistema operativo que suporte Java. A interface conta com três menus principais:

- **Foto:** Permite recolher e processar uma imagem através de um clique no respetivo ícone (máquina fotográfica) ou na área central da aplicação. Após recolhida a imagem, o processamento é inicializado automaticamente, sendo exibido no menu lateral direito as classes de refeições com maior afinidade à imagem de consulta. É exibido também em cada classe de refeições o valor calórico estimado para a imagem de consulta, assim como a pontuação que essa classe obteve na classificação. No final do processamento a aplicação aguarda um clique por parte do utilizador na classe correta, ao qual se segue a adição de uma nova entrada no histórico de refeições. Na Figura 18 temos a janela principal da interface gráfica desenvolvida.

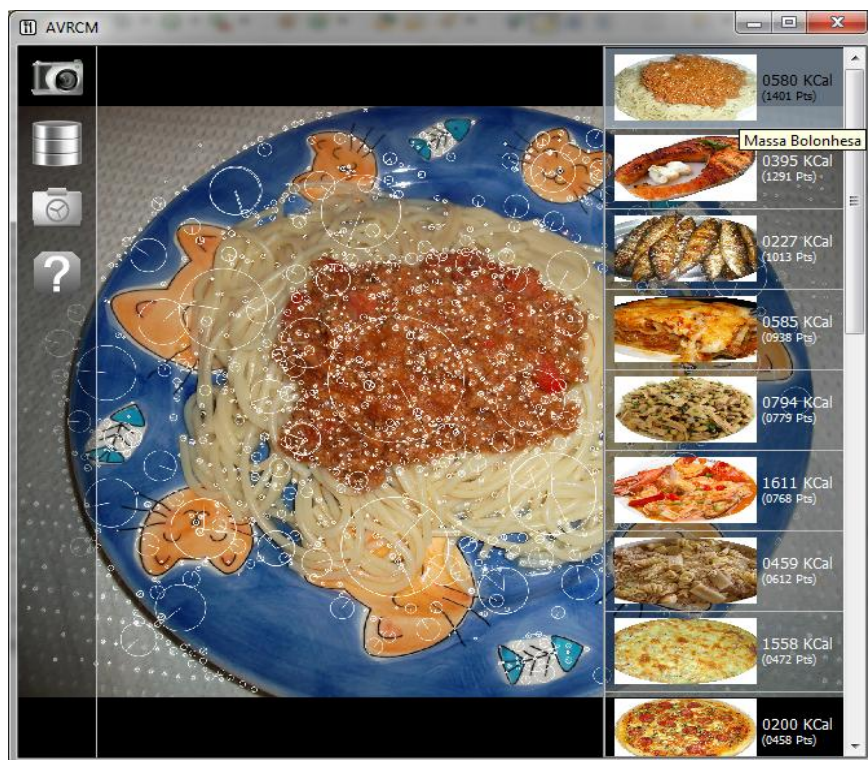


Figura 18: Interface Gráfica

- **Base de Dados:** Permite construir uma base de dados a partir de um conjunto de imagens de refeições identificadas (nome das imagens definidos da forma Nome[TAG_CALORIAS].ext), definir o algoritmo a usar para extração de características, pesquisar, editar e exportar a base de dados corrente e importar bases de dados armazenadas em disco. Na Figura 19 vemos a janela de auxílio à criação e gestão da base de dados.

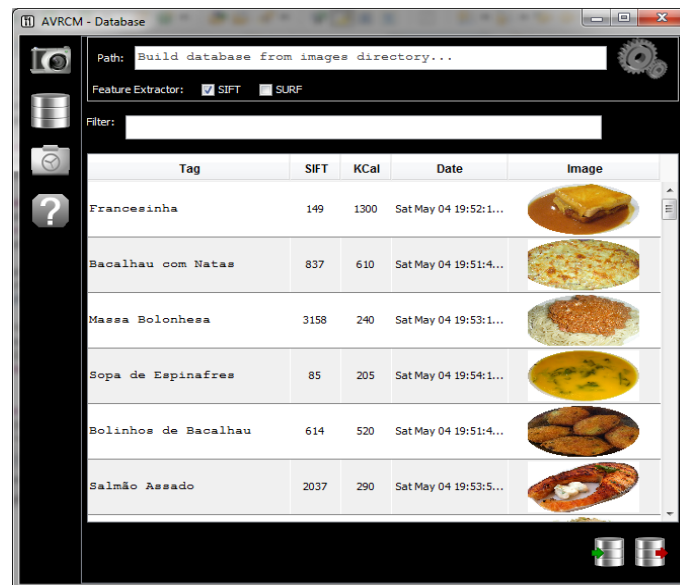


Figura 19: Menu de Base de Dados

- **Histórico:** Permite pesquisar, editar, exportar e importar o histórico de imagens de refeições processadas. Na Figura 20 temos a janela de auxílio à gestão do histórico de imagens processadas.

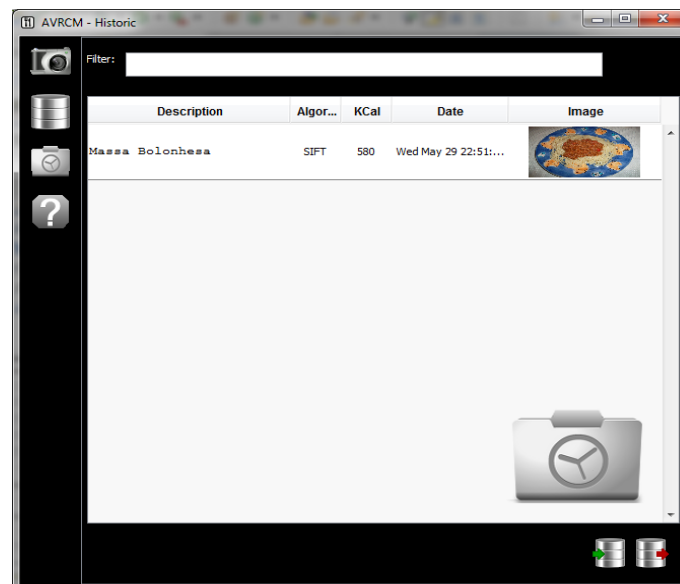


Figura 20: Menu de Histórico

CAPÍTULO VI: RESULTADOS

CAPÍTULO VI: RESULTADOS

6.1. Base de Dados

Para a realização da análise à metodologia proposta utilizamos 17 conjuntos de imagens, com cada conjunto a conter cinco imagens de cada classe de refeições (85 imagens no total). As imagens foram recolhidas usando uma câmara fotográfica de 2 mega pixels de um telemóvel LG modelo GS290. Na recolha das imagens procurou-se que existisse uma variação considerável em termos de disposição e quantidade dos alimentos dentro de cada classe de refeições. Em cada um dos conjuntos de imagens escolhemos aleatoriamente uma imagem para servir de imagem de consulta, sendo as restantes usadas na construção da base de dados. Na Figura 21 temos alguns exemplos de imagens de refeições utilizadas para análise à metodologia proposta.



Figura 21: Exemplos de algumas das imagens utilizadas na análise à metodologia

No **ANEXO B**: Base de Dados podem ser consultados os *thumbnails* de todas as imagens utilizadas na análise à metodologia proposta.

6.2. Reconhecimento de Refeições

No fim do reconhecimento automático de refeições a aplicação devolve, de forma ordenada, as classes mais prováveis de corresponderem à classe da imagem de consulta. Por forma a simplificar a visualização dos resultados, vamos considerar somente os resultados até à nona classe mais favorável, isto é, se para uma determinada imagem de consulta a classe correspondente ficar classificada fora dos nove primeiros lugares, vamos considerar que o reconhecimento da imagem falhou, e entrará para as estatísticas como imagem não identificada. Para as restantes imagens vamos exibir o lugar em que ficou classificada a classe à qual a imagem de consulta pertence. Por exemplo, supondo que temos imagens de Bolonhesa, Massa de Carne e Prego no Prato como imagens de consulta, e na classificação obtivemos as seguintes listagens de classes ordenadas da mais provável à menos provável para cada uma das imagens de consulta: (considerando os resultados somente até à terceira classe mais favorável)

- **Bolonhesa:** (1) Massa de Carne, (2) Bolonhesa, (3) Massa de Atum
- **Massa de Carne:** (1) Massa de Carne, (2) Massa de Atum, (3) Massa de Frango
- **Prego no Prato:** (1) Rojões à Minhota, (2) Panados com Arroz, (3) Feijoada

Então, temos que apenas a Massa de Carne foi classificada corretamente, sendo que no caso da imagem de Bolonhesa a sua classe real ficou classificada como sendo a segunda mais provável e no caso do Prego no Prato a classe real ficou mesmo fora da lista de classes mais favoráveis. Assim, para 1/3 das imagens (33,3%) a classe real correspondeu à classe mais favorável, para 2/3 das imagens (66,7%) a classe real ficou classificada entre as duas mais favoráveis e 1/3 das imagens não foram identificadas.

Nas tabelas de resultados que se seguem podemos verificar para cada imagem de consulta e para cada uma das diferentes configurações adotadas, a posição em que ficou classificada a classe real da imagem de consulta. Em complemento a estes valores juntamos no fim de cada tabela uma secção de dados estatísticos a fim de se conseguir uma visão geral dos resultados obtidos. Começamos por analisar o resultado de variar as dimensões do *Bag of Words* na prestação da metodologia mais tradicional de pesquisa de imagens com base em características locais. Na Tabela 1 podemos observar os resultados obtidos com a utilização de *Bag of Words* padrão de 50, 100 e 200 palavras, e os algoritmos SIFT e SURF respetivamente.

Imagem de Consulta	SIFT			SURF		
	50	100	200	50	100	200
Batatas com Bacalhau	2	2	3	8	7	---
Caldo Verde	---	---	---	1	2	8
Carne Assada com Batata e Arroz	1	3	2	2	4	2
Entrecosto Grelhado com Arroz	2	2	2	1	2	2
Feijoada	3	1	4	1	1	1
Frango Assado com Batata e Arroz	7	7	---	7	7	---
Massa com Atum	1	1	1	1	1	1
Massa Bolonhesa	1	1	1	1	1	1
Massa de Carne	1	1	1	1	1	1
Massa de Frango	1	1	2	2	1	2
Panados com Arroz	2	2	2	1	1	1
Pataniscas de Bacalhau	1	1	1	1	1	1
Prego no Prato	5	3	4	3	6	4
Rojões à Minhota	3	4	3	4	3	3
Salada de Alface e Tomate	9	---	---	8	9	---
Salada de Feijão Frade	1	1	1	1	1	1
Sardinhas Assadas	1	1	1	1	1	1
Corretamente Classificado (%)	47,1	47,1	35,3	58,8	52,9	47,1
Entre as 2 classes mais favoráveis (%)	64,7	64,7	58,8	70,6	64,7	64,7
Entre as 3 classes mais favoráveis (%)	76,5	76,5	70,6	76,5	70,6	70,6
Entre as 5 classes mais favoráveis (%)	82,4	82,4	82,4	82,4	76,5	76,5
Não Identificado (%)	5,9	11,8	17,6	0,0	0,0	17,6

Tabela 1: SIFT e SURF com *Bag of Words* de 50, 100 e 200 palavras. Podemos observar para cada imagem de consulta e para cada configuração, a posição em que ficou classificada a classe real da imagem de consulta.

Como se pode constatar por observação dos resultados da Tabela 1, obtemos melhores resultados com a utilização de *Bag of Words* de tamanhos menores tanto com a utilização do SIFT como do SURF (pelo menos para uma base de dados de pequena dimensão como a que é usada), o que é claramente positivo uma vez que implicitamente um *Bag of Words* de tamanho menor permite reduzir o gasto computacional no reconhecimento da refeição. Em concreto os *Bags of Words* de 50 e 100 palavras oferecem resultados bastante similares entre si, e consideravelmente superiores aos obtidos com a utilização de 200 palavras. De notar os maus resultados na identificação de Batatas com Bacalhau, Caldo Verde, Frango Assado com Batata e Arroz e Salada de Alface e Tomate, situação que pode ser explicada pelo baixo número de descritores extraídos nas imagens destas refeições⁵.

⁵ Na Tabela 18 e Tabela 19 temos o número de descritores extraídos com SIFT e SURF respetivamente, para um par de imagens de cada classe de refeições, com cada imagem a corresponder a aproximadamente uma dose média.

Em seguida analisamos os melhoramentos obtidos com a introdução de informação de cor ao nível dos descritores locais com a descrição dos pontos de interesse nas três camadas de cor e com concatenação de histogramas de cor. Como base usamos os algoritmos SIFT e SURF, e usamos *Bags of Words* de 50 e 100 palavras. Na Tabela 2 e Tabela 3 podemos observar os resultados obtidos com SIFT e *Bags of Words* de 50 e 100 palavras respetivamente. Na Tabela 4 e Tabela 5 temos os resultados obtidos com SURF e *Bags of Words* de 50 e 100 palavras respetivamente. Segue uma lista na nomenclatura usada, comum às próximas tabelas:

- **DCC:** Descrição dos pontos de interesse nas **Camadas de Cor**;
- **CHC:** Concatenação de **Histograma de Cor** ao Descritor Base;
- **HCG:** Histograma de **Cor Global**;
- Modelos de Histogramas:
 - **nR_αG_βB_γ:** Histograma RGB com n dimensões, e **α**, **β** e **γ** bits para representar as camadas R, G e B respetivamente;
 - **nH_αS_βI_γ:** Histograma HSI com n dimensões, quantizado a **α** níveis de matiz, **β** níveis de saturação e **γ** níveis intensidade;

Imagem de Consulta	SIFT e Bag of Words de 50 Palavras					Refv
	DCC	CHC				
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁	
Batatas com Bacalhau	2	3	6	3	---	
Caldo Verde	---	---	---	---	---	
Carne Assada com Batata e Arroz	1	2	2	1	2	
Entrecosto Grelhado com Arroz	2	2	3	4	2	
Feijoadada	2	3	2	3	2	
Frango Assado com Batata e Arroz	6	3	6	4	8	
Massa com Atum	1	1	1	1	1	
Massa Bolonhesa	1	1	1	1	1	
Massa de Carne	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	
Panados com Arroz	2	2	2	2	2	
Pataniscas de Bacalhau	1	1	1	1	1	
Prego no Prato	5	4	5	5	1	
Rojões à Minhota	2	3	2	4	5	
Salada de Alface e Tomate	4	---	---	7	5	
Salada de Feijão Frade	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	
Corretamente Classificado (%)	47,1	41,2	41,2	47,1	47,1	47,1
Nas 2 classes mais favoráveis (%)	76,5	58,8	64,7	52,9	70,6	64,7
Nas 3 classes mais favoráveis (%)	76,5	82,4	70,6	64,7	70,6	76,5
Nas 5 classes mais favoráveis (%)	88,2	88,2	76,5	88,2	82,4	82,4
Não Identificado (%)	5,9	11,8	11,8	5,9	11,8	5,9

Tabela 2: SIFT e BoW de 50 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SIFT e BoW de 50 palavras, sem informação de cor.

Imagem de Consulta	SIFT e Bag of Words de 100 Palavras					Refv
	DCC	CHC				
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁	
Batatas com Bacalhau	8	2	3	2	3	
Caldo Verde	---	---	---	---	---	
Carne Assada com Batata e Arroz	2	2	2	2	1	
Entrecosto Grelhado com Arroz	2	2	2	2	2	
Feijoada	3	4	4	4	3	
Frango Assado com Batata e Arroz	7	8	9	7	8	
Massa com Atum	1	1	1	1	1	
Massa Bolonhesa	1	2	1	1	1	
Massa de Carne	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	
Panados com Arroz	2	2	2	2	1	
Pataniscas de Bacalhau	1	1	1	1	1	
Prego no Prato	4	5	4	4	5	
Rojões à Minhota	4	2	4	5	2	
Salada de Alface e Tomate	5	5	4	---	7	
Salada de Feijão Frade	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	
Corretamente Classificado (%)	41,2	35,3	41,2	41,2	52,9	47,1
Nas 2 classes mais favoráveis (%)	58,8	70,6	58,8	64,7	64,7	64,7
Nas 3 classes mais favoráveis (%)	64,7	70,6	64,7	64,7	76,5	76,5
Nas 5 classes mais favoráveis (%)	82,4	88,2	88,2	82,4	82,4	82,4
Não Identificado (%)	5,9	5,9	5,9	11,8	5,9	11,8

Tabela 3: SIFT e BoW de 100 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SIFT e BoW de 100 palavras, sem informação de cor.

Imagem de Consulta	SURF e Bag of Words de 50 Palavras					Refv
	DCC	CHC				
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁	
Batatas com Bacalhau	1	2	2	1	1	
Caldo Verde	4	1	1	1	2	
Carne Assada com Batata e Arroz	1	1	1	1	1	
Entrecosto Grelhado com Arroz	1	1	1	1	1	
Feijoada	1	1	1	1	1	
Frango Assado com Batata e Arroz	7	8	3	1	1	
Massa com Atum	1	1	1	1	1	
Massa Bolonhesa	1	1	2	1	1	
Massa de Carne	1	1	1	2	1	
Massa de Frango	1	1	1	1	1	
Panados com Arroz	1	3	1	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	
Prego no Prato	4	1	1	3	2	
Rojões à Minhota	3	1	1	2	1	
Salada de Alface e Tomate	3	2	1	1	1	
Salada de Feijão Frade	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	
Corretamente Classificado (%)	70,6	76,5	82,4	82,4	88,2	58,8
Nas 2 classes mais favoráveis (%)	70,6	88,2	94,1	94,1	100,0	70,6
Nas 3 classes mais favoráveis (%)	82,4	94,1	100,0	100,0	100,0	76,5
Nas 5 classes mais favoráveis (%)	94,1	94,1	100,0	100,0	100,0	82,4
Não Identificado (%)	0,0	0,0	0,0	0,0	0,0	0,0

Tabela 4: SURF e BoW de 50 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SURF e BoW de 50 palavras, sem informação de cor.

Imagem de Consulta	SURF e Bag of Words de 100 Palavras					Refv
	DCC	CHC				
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁	
Batatas com Bacalhau	5	2	2	3	1	
Caldo Verde	8	1	1	1	2	
Carne Assada com Batata e Arroz	1	1	1	1	1	
Entrecosto Grelhado com Arroz	1	1	1	1	1	
Feijoada	1	1	1	2	2	
Frango Assado com Batata e Arroz	8	7	3	1	3	
Massa com Atum	1	1	1	1	1	
Massa Bolonhesa	1	1	2	1	1	
Massa de Carne	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	
Panados com Arroz	1	4	1	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	
Prego no Prato	3	1	2	2	2	
Rojões à Minhota	2	2	1	2	1	
Salada de Alface e Tomate	---	6	1	1	1	
Salada de Feijão Frade	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	
Corretamente Classificado (%)	64,7	70,6	76,5	76,5	76,5	52,9
Nas 2 classes mais favoráveis (%)	70,6	82,4	94,1	94,1	94,1	64,7
Nas 3 classes mais favoráveis (%)	76,5	82,4	100,0	100,0	100,0	70,6
Nas 5 classes mais favoráveis (%)	82,4	88,2	100,0	100,0	100,0	76,5
Não Identificado (%)	5,9	0,0	0,0	0,0	0,0	0,0

Tabela 5: SURF e BoW de 100 palavras com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de apenas SURF e BoW de 100 palavras, sem informação de cor.

Na Tabela 6 temos um sumário dos resultados obtidos com a introdução de informação de cor nos descritores.

			= 1	≤ 2	≤ 3	≤ 5	NI	
SIFT	BoW 50 Palavras	DCC		47,1	76,5	76,5	88,2	5,9
		CHC	8R ₁ G ₁ B ₁	41,2	58,8	82,4	88,2	11,8
			32R ₂ G ₂ B ₁	41,2	64,7	70,6	76,5	11,8
			16H ₂ S ₁ I ₁	47,1	52,9	64,7	88,2	5,9
			32H ₃ S ₁ I ₁	47,1	70,6	70,6	82,4	11,8
	BoW 100 Palavras	DCC		41,2	58,8	64,7	82,4	5,9
		CHC	8R ₁ G ₁ B ₁	35,3	70,6	70,6	88,2	5,9
			32R ₂ G ₂ B ₁	41,2	58,8	64,7	88,2	5,9
			16H ₂ S ₁ I ₁	41,2	64,7	64,7	82,4	11,8
			32H ₃ S ₁ I ₁	52,9	64,7	76,5	82,4	5,9
SURF	BoW 50 Palavras	DCC		70,6	70,6	82,4	94,1	0,0
		CHC	8R ₁ G ₁ B ₁	76,5	88,2	94,1	94,1	0,0
			32R ₂ G ₂ B ₁	82,4	94,1	100,0	100,0	0,0
			16H ₂ S ₁ I ₁	82,4	94,1	100,0	100,0	0,0
			32H ₃ S ₁ I ₁	88,2	100,0	100,0	100,0	0,0
	BoW 100 Palavras	DCC		64,7	70,6	76,5	82,4	5,9
		CHC	8R ₁ G ₁ B ₁	70,6	82,4	82,4	88,2	0,0
			32R ₂ G ₂ B ₁	76,5	94,1	100,0	100,0	0,0
			16H ₂ S ₁ I ₁	76,5	94,1	100,0	100,0	0,0
			32H ₃ S ₁ I ₁	76,5	94,1	100,0	100,0	0,0

Tabela 6: Resumo dos resultados com SIFT e SURF, Bag of Words de 50 e 100 palavras com Informação Local de Cor com DCC e CHC.

De uma forma geral, a introdução de informação de cor nos descritores locais produziu melhores resultados, com maior relevo para o algoritmo SURF. O tamanho do *Bag of Words* (50 ou 100 palavras) não teve um papel significativo nos resultados obtidos. Como era previsível, o modelo de cor HSI garante uma melhor descrição da informação de cor, e por sua vez melhores resultados. Na generalidade a concatenação de histogramas de cor foi mais vantajosa que a descrição da cor nas várias camadas, o que é positivo principalmente em termos de custo computacional.

Para finalizar o estudo da abordagem com *Bag of Words* padrão, segue-se os resultados obtidos com a introdução de informação global de cor, usando os modelos de cor RGB e HSI e histogramas de 64, 256 e 512 cores. Como base usamos a melhor das configurações obtidas anteriormente (descrição local de cor com concatenação de histogramas 32HSI e *Bag of Words* de 50 palavras), e para comparação adicionamos às tabelas os resultados anteriormente obtidos com a configuração adotada. Na Tabela 7 e Tabela 8 podemos observar a contribuição da introdução de descritores globais de cor para os algoritmos SIFT e SURF respetivamente.

Imagem de Consulta	SIFT e BoW de 50 Palavras						
	64 R ₂ G ₂ B ₂	256 R ₃ G ₃ B ₂	512 R ₃ G ₃ B ₃	64 H ₃ S ₂ I ₁	256 H ₄ S ₂ I ₂	512 H ₄ S ₃ I ₂	
Batatas com Bacalhau	2	2	2	2	2	2	
Caldo Verde	4	4	4	4	4	5	
Carne Assada com Batata e Arroz	1	2	2	1	1	1	
Entrecosto Grelhado com Arroz	2	2	2	2	2	2	
Feijoada	1	1	1	1	1	1	
Frango Assado com Batata e Arroz	---	9	9	8	7	9	
Massa com Atum	1	1	1	1	1	1	
Massa Bolonhesa	1	1	1	1	1	1	
Massa de Carne	1	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	1	
Panados com Arroz	1	1	1	1	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	1	
Prego no Prato	1	1	1	1	1	1	
Rojões à Minhota	2	2	2	2	2	2	
Salada de Alface e Tomate	7	5	4	3	3	3	
Salada de Feijão Frade	1	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	1	
Corretamente Classificado (%)	64,7	58,8	58,8	64,7	64,7	64,7	Refv 47,1
Nas 2 classes mais favoráveis (%)	82,4	82,4	82,4	82,4	82,4	82,4	70,6
Nas 3 classes mais favoráveis (%)	82,4	82,4	82,4	88,2	88,2	88,2	70,6
Nas 5 classes mais favoráveis (%)	88,2	94,1	94,1	94,1	94,1	94,1	82,4
Não Identificado (%)	5,9	0,0	0,0	0,0	0,0	0,0	11,8

Tabela 7: SIFT e BoW de 50 palavras com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT com CHC-32HSI e BoW de 50 palavras.

Imagem de Consulta	SURF e BoW de 50 Palavras						
	64 $R_2G_2B_2$	256 $R_3G_3B_2$	512 $R_3G_3B_3$	64 $H_3S_2I_1$	256 $H_4S_2I_2$	512 $H_4S_3I_2$	
Batatas com Bacalhau	1	1	1	1	1	1	
Caldo Verde	2	2	2	2	2	2	
Carne Assada com Batata e Arroz	1	1	1	1	1	1	
Entrecosto Grelhado com Arroz	1	1	1	1	1	1	
Feijoada	1	1	1	2	1	1	
Frango Assado com Batata e Arroz	2	2	2	3	1	2	
Massa com Atum	1	1	1	1	1	1	
Massa Bolonhesa	1	1	1	1	1	1	
Massa de Carne	1	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	1	
Panados com Arroz	1	1	1	1	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	1	
Prego no Prato	1	1	1	1	1	1	
Rojões à Minhota	1	1	1	1	1	1	
Salada de Alface e Tomate	1	1	1	1	1	1	
Salada de Feijão Frade	1	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	1	Refv
Corretamente Classificado (%)	88,2	88,2	88,2	82,4	94,1	88,2	88,2
Nas 2 classes mais favoráveis (%)	100,0	100,0	100,0	94,1	100,0	100,0	100,0
Nas 3 classes mais favoráveis (%)	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Nas 5 classes mais favoráveis (%)	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Não Identificado (%)	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Tabela 8: SURF e BoW de 50 palavras com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF com CHC-32HSI e BoW de 50 palavras.

Por observação dos resultados da Tabela 7 e Tabela 8 podemos constatar que no caso do algoritmo SIFT a introdução de informação global de cor resultou numa contribuição muito positiva, já no caso do SURF os resultados obtidos mantiveram-se de forma geral idênticos aos valores de referência. Novamente podemos verificar que o modelo de cor HSI proporciona uma melhor descrição da informação de cor, e por sua vez melhores resultados em comparação com o modelo RGB. Ao nível do tamanho do histograma, obtivemos um melhor desempenho com 256 e 512 cores.

Segue-se a análise ao método proposto de construção de um *Bag of Words* para cada classe de refeições. Na Tabela 9 vemos os resultados obtidos com a utilização dos algoritmos SIFT e SURF, e *Bags of Words* de 50 e 100 palavras por Classe. Por forma a facilitar a comparação entre metodologias, adicionamos à tabela os resultados obtidos com a utilização da metodologia padrão para cada uma das configurações.

Imagem de Consulta	SIFT				SURF			
	50		100		50		100	
Batatas com Bacalhau	1		1		1		9	
Caldo Verde	9		8		3		8	
Carne Assada com Batata e Arroz	1		1		1		1	
Entrecosto Grelhado com Arroz	1		2		2		2	
Feijoada	2		2		1		2	
Frango Assado com Batata e Arroz	7		9		---		8	
Massa com Atum	1		1		1		2	
Massa Bolonhesa	2		4		2		1	
Massa de Carne	1		1		1		1	
Massa de Frango	1		3		1		1	
Panados com Arroz	1		1		1		1	
Pataniscas de Bacalhau	1		1		1		1	
Prego no Prato	1		1		1		1	
Rojões à Minhota	1		1		1		1	
Salada de Alface e Tomate	2		2		---		8	
Salada de Feijão Frade	1		1		1		1	
Sardinhas Assadas	1	<i>Refv</i>	1	<i>Refv</i>	1	<i>Refv</i>	1	<i>Refv</i>
Corretamente Classificado (%)	70,6	47,1	58,8	47,1	70,6	58,8	58,8	52,9
Nas 2 classes mais favoráveis (%)	88,2	64,7	76,5	64,7	82,4	70,6	76,5	64,7
Nas 3 classes mais favoráveis (%)	88,2	76,5	82,4	76,5	88,2	76,5	76,5	70,6
Nas 5 classes mais favoráveis (%)	88,2	82,4	88,2	82,4	88,2	82,4	76,5	76,5
Não Identificado (%)	0,0	5,9	0,0	11,8	11,8	0,0	0,0	0,0

Tabela 9: Bag of Words de 50 e 100 palavras por classe. Nas colunas de referência temos respetivamente os valores obtidos utilizando SIFT com BoW de 50 palavras e 100 palavras, e SURF com BoW de 50 palavras e 100 palavras, em ambos os casos sem introdução de informação de cor.

Por análise dos resultados da Tabela 9 podemos verificar que existe um claro ganho ao utilizar um *Bag of Words* por Classe, em todas as configurações adotadas. De uma forma geral a probabilidade de a classe da imagem de consulta estar entre as três classes selecionadas como mais prováveis aumentou consideravelmente, confirmando-se a suposição de que não descartando a informação das classes conseguimos obter uma maior precisão.

Tendo-se obtido resultados promissores com a utilização de *Bag of Words* por Classe, é interessante verificar o ganho de combinar esta abordagem com introdução de cor ao nível local. Assim, seguem-se na Tabela 10 e Tabela 11 os resultados obtidos com SIFT e *Bags of Words* de 50 e 100 palavras por classe respetivamente, e na Tabela 12 e Tabela 13 os resultados obtidos com SURF e *Bags of Words* de 50 e 100 palavras por classe respetivamente. Adicionamos às tabelas como referência os valores obtidos para essas mesmas configurações (algoritmos SIFT e SURF, e *Bags of Words* de 50 e 100 palavras por classe), mas sem inclusão de informação de cor.

Imagem de Consulta	SIFT e BoW de 50 Palavras por Classe					Refv
	DCC	CHC				
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁	
Batatas com Bacalhau	1	1	1	1	1	
Caldo Verde	1	1	2	1	2	
Carne Assada com Batata e Arroz	1	1	1	2	1	
Entrecosto Grelhado com Arroz	1	1	1	1	1	
Feijoada	5	5	2	2	3	
Frango Assado com Batata e Arroz	6	9	3	---	---	
Massa com Atum	1	1	1	1	1	
Massa Bolonhesa	1	3	1	4	2	
Massa de Carne	1	1	1	1	1	
Massa de Frango	2	1	1	3	2	
Panados com Arroz	1	1	1	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	
Prego no Prato	1	1	1	1	1	
Rojões à Minhota	2	3	2	3	4	
Salada de Alface e Tomate	1	2	2	2	2	
Salada de Feijão Frade	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	
Corretamente Classificado (%)	76,5	70,6	70,6	58,8	58,8	70,6
Nas 2 classes mais favoráveis (%)	88,2	76,5	94,1	76,5	82,4	88,2
Nas 3 classes mais favoráveis (%)	88,2	88,2	100,0	88,2	88,2	88,2
Nas 5 classes mais favoráveis (%)	94,1	94,1	100,0	94,1	94,1	88,2
Não Identificado (%)	0,0	0,0	0,0	5,9	5,9	0,0

Tabela 10: SIFT e BoW de 50 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT e BoW de 50 palavras por classe, sem informação de cor.

Imagem de Consulta	SIFT e BoW de 100 Palavras por Classe				
	DCC	CHC			
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁
Batatas com Bacalhau	1	1	1	1	1
Caldo Verde	2	9	4	---	7
Carne Assada com Batata e Arroz	2	1	1	1	1
Entrecosto Grelhado com Arroz	1	1	1	2	1
Feijoada	5	4	3	3	4
Frango Assado com Batata e Arroz	---	5	8	7	---
Massa com Atum	1	1	1	1	1
Massa Bolonhesa	2	2	2	3	2
Massa de Carne	1	1	1	1	1
Massa de Frango	2	1	1	1	1
Panados com Arroz	1	1	1	1	1
Pataniscas de Bacalhau	1	1	1	1	1
Prego no Prato	1	1	1	1	1
Rojões à Minhota	4	3	3	2	3
Salada de Alface e Tomate	3	2	5	6	1
Salada de Feijão Frade	1	1	1	1	1
Sardinhas Assadas	1	1	1	1	1
Corretamente Classificado (%)	52,9	64,7	64,7	58,8	70,6
Nas 2 classes mais favoráveis (%)	76,5	76,5	70,6	70,6	76,5
Nas 3 classes mais favoráveis (%)	82,4	82,4	82,4	82,4	82,4
Nas 5 classes mais favoráveis (%)	94,1	94,1	94,1	82,4	88,2
Não Identificado (%)	5,9	0,0	0,0	5,9	5,9

Refv

58,8

76,5

82,4

88,2

0,00

Tabela 11: SIFT e BoW de 100 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT e BoW de 100 palavras por classe, sem informação de cor.

Imagem de Consulta	SURF e BoW de 50 Palavras por Classe					Refv
	DCC	CHC				
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁	
Batatas com Bacalhau	2	2	1	1	1	
Caldo Verde	2	1	1	1	1	
Carne Assada com Batata e Arroz	1	1	1	1	1	
Entrecosto Grelhado com Arroz	1	1	1	1	1	
Feijoada	1	1	1	1	1	
Frango Assado com Batata e Arroz	6	3	2	2	1	
Massa com Atum	1	1	1	2	1	
Massa Bolonhesa	1	1	1	1	1	
Massa de Carne	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	
Panados com Arroz	1	1	2	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	
Prego no Prato	1	1	1	1	1	
Rojões à Minhota	6	3	2	5	4	
Salada de Alface e Tomate	---	1	1	1	1	
Salada de Feijão Frade	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	
Corretamente Classificado (%)	70,6	82,4	82,4	82,4	94,1	70,6
Nas 2 classes mais favoráveis (%)	82,4	88,2	100,0	94,1	94,1	82,4
Nas 3 classes mais favoráveis (%)	82,4	100,0	100,0	94,1	94,1	88,2
Nas 5 classes mais favoráveis (%)	82,4	100,0	100,0	100,0	100,0	88,2
Não Identificado (%)	5,9	0,0	0,0	0,0	0,0	11,8

Tabela 12: SURF e BoW de 50 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF e BoW de 50 palavras por classe, sem informação de cor.

Imagem de Consulta	SURF e BoW de 100 Palavras por Classe					Refv
	DCC	CHC				
		8R ₁ G ₁ B ₁	32R ₂ G ₂ B ₁	16H ₂ S ₁ I ₁	32H ₃ S ₁ I ₁	
Batatas com Bacalhau	2	1	1	1	1	
Caldo Verde	7	3	1	1	1	
Carne Assada com Batata e Arroz	1	1	2	1	1	
Entrecosto Grelhado com Arroz	1	1	1	1	1	
Feijoada	1	1	1	1	1	
Frango Assado com Batata e Arroz	5	7	3	2	1	
Massa com Atum	1	1	1	1	1	
Massa Bolonhesa	1	1	1	1	1	
Massa de Carne	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	
Panados com Arroz	1	1	2	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	
Prego no Prato	1	1	1	1	1	
Rojões à Minhota	4	1	1	2	2	
Salada de Alface e Tomate	---	2	1	1	1	
Salada de Feijão Frade	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	
Corretamente Classificado (%)	70,6	82,4	82,4	88,2	94,1	58,8
Nas 2 classes mais favoráveis (%)	76,5	88,2	94,1	100,0	100,0	76,5
Nas 3 classes mais favoráveis (%)	76,5	94,1	100,0	100,0	100,0	76,5
Nas 5 classes mais favoráveis (%)	88,2	94,1	100,0	100,0	100,0	76,5
Não Identificado (%)	5,9	0,00	0,0	0,0	0,0	0,0

Tabela 13: SURF e BoW de 100 palavras por classe com Informação Local de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF e BoW de 100 palavras por classe, sem informação de cor.

Na Tabela 14 temos um sumário dos resultados obtidos com a introdução de informação de cor nos descritores.

			= 1	≤ 2	≤ 3	≤ 5	NI
SIFT	BoW 50 Palavras por Classe	DCC	76,5	88,2	88,2	94,1	0,0
		8R ₁ G ₁ B ₁	70,6	76,5	88,2	94,1	0,0
		32R ₂ G ₂ B ₁	70,6	94,1	100,0	100,0	0,0
		16H ₂ S ₁ I ₁	58,8	76,5	88,2	94,1	5,9
		32H ₃ S ₁ I ₁	58,8	82,4	88,2	94,1	5,9
	BoW 100 Palavras por Classe	DCC	52,9	76,5	82,4	94,1	5,9
		8R ₁ G ₁ B ₁	64,7	76,5	82,4	94,1	0,0
		32R ₂ G ₂ B ₁	64,7	76,5	82,4	94,1	0,0
		16H ₂ S ₁ I ₁	58,8	70,6	82,4	82,4	5,9
		32H ₃ S ₁ I ₁	70,6	76,5	82,4	88,2	5,9
SURF	BoW 50 Palavras por Classe	DCC	70,6	82,4	82,4	82,4	5,9
		8R ₁ G ₁ B ₁	82,4	88,2	100,0	100,0	0,0
		32R ₂ G ₂ B ₁	82,4	100,0	100,0	100,0	0,0
		16H ₂ S ₁ I ₁	82,4	94,1	94,1	100,0	0,0
		32H ₃ S ₁ I ₁	94,1	94,1	94,1	100,0	0,0
	BoW 100 Palavras por Classe	DCC	70,6	76,5	76,5	88,2	5,9
		8R ₁ G ₁ B ₁	82,4	88,2	94,1	94,1	0,0
		32R ₂ G ₂ B ₁	82,4	94,1	100,0	100,0	0,0
		16H ₂ S ₁ I ₁	88,2	100,0	100,0	100,0	0,0
		32H ₃ S ₁ I ₁	94,1	100,0	100,0	100,0	0,0

Tabela 14: Resumo dos resultados com SIFT e SURF, Bag of Words de 50 e 100 palavras por Classe com Informação Local de Cor com DCC e CHC.

Globalmente a introdução de informação de cor nos descritores locais produziu melhores resultados, tanto com a utilização de *Bag of Words* padrão como com *Bag of Words* por classe, principalmente com a utilização do algoritmo SURF. Em comparação direta entre as duas abordagens, *Bag of Words* por classe leva uma clara vantagem. Novamente refira-se que os resultados obtidos com a descrição da cor nas várias camadas não superaram os resultados obtidos com a concatenação de histogramas de cor, e dentro destes uma referência para o modelo de cor HSI, o qual garantiu uma melhor descrição da informação de cor. De notar também que para *Bag of Words* por classe a utilização de um núcleo pequeno de 50 palavras produz melhores resultados. Isto pode ser explicado pelo facto de ao usarmos um pequeno núcleo de palavras para caracterizar cada classe, estarmos a seleccionar do conjunto total de descritores somente aqueles com maior peso, proporcionando uma maior distinção entre classes.

Segue-se os resultados obtidos com a introdução de informação global de cor, usando os modelos de cor RGB e HSI e histogramas de 64, 256 e 512 cores. Como base usamos a combinação das configurações que apresentaram globalmente melhores resultados (descrição local de cor com concatenação de histogramas *32HSI* e *Bag of Words* de 50 palavras por classe), e para comparação adicionamos às tabelas os resultados obtidos com a configuração base. Na Tabela 15 e Tabela 16 podemos observar a contribuição da introdução de descritores globais de cor, com SIFT e SURF respetivamente.

Imagem de Consulta	SIFT e BoW de 50 Palavras por Classe						
	64 R ₂ G ₂ B ₂	256 R ₃ G ₃ B ₂	512 R ₃ G ₃ B ₃	64 H ₃ S ₂ I ₁	256 H ₄ S ₂ I ₂	512 H ₄ S ₃ I ₂	
Batatas com Bacalhau	1	1	1	1	1	1	Refv
Caldo Verde	1	1	1	1	1	1	
Carne Assada com Batata e Arroz	2	2	2	2	2	2	
Entrecosto Grelhado com Arroz	3	1	1	2	1	2	
Feijoada	1	1	1	1	1	1	
Frango Assado com Batata e Arroz	2	2	2	3	2	2	
Massa com Atum	1	1	1	1	1	1	
Massa Bolonhesa	3	3	3	3	3	2	
Massa de Carne	1	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	1	
Panados com Arroz	1	1	1	1	1	1	
Pataniscas de Bacalhau	1	1	1	1	1	1	
Prego no Prato	1	1	1	1	1	1	
Rojões à Minhota	1	1	1	2	2	2	
Salada de Alface e Tomate	1	1	1	1	1	1	
Salada de Feijão Frade	1	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	1	
Corretamente Classificado (%)	76,5	82,4	82,4	70,6	76,5	70,6	58,8
Nas 2 classes mais favoráveis (%)	88,2	94,1	94,1	88,2	94,1	100,0	82,4
Nas 3 classes mais favoráveis (%)	100,0	100,0	100,0	100,0	100,0	100,0	88,2
Nas 5 classes mais favoráveis (%)	100,0	100,0	100,0	100,0	100,0	100,0	94,1
Não Identificado (%)	0,0	0,0	0,0	0,0	0,0	0,0	5,9

Tabela 15: SIFT e BoW de 50 palavras por classe com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SIFT com CHC-32HSI e BoW de 50 palavras por classe.

Imagem de Consulta	SURF e BoW de 50 Palavras por Classe						
	64 R ₂ G ₂ B ₂	256 R ₃ G ₃ B ₂	512 R ₃ G ₃ B ₃	64 H ₃ S ₂ I ₁	256 H ₄ S ₂ I ₂	512 H ₄ S ₃ I ₂	
Batatas com Bacalhau	1	1	1	1	1	1	
Caldo Verde	1	1	1	1	1	1	
Carne Assada com Batata e Arroz	1	1	1	1	1	1	
Entrecosto Grelhado com Arroz	1	1	1	1	1	1	
Feijoada	1	1	1	1	1	1	
Frango Assado com Batata e Arroz	1	1	1	1	1	1	
Massa com Atum	1	1	1	1	1	1	
Massa Bolonhesa	1	1	1	1	1	1	
Massa de Carne	1	1	1	1	1	1	
Massa de Frango	1	1	1	1	1	1	
Panados com Arroz	2	2	2	2	2	2	
Pataniscas de Bacalhau	1	1	1	1	1	1	
Prego no Prato	1	1	1	1	1	1	
Rojões à Minhota	2	2	2	1	1	1	
Salada de Alface e Tomate	1	1	1	1	1	1	
Salada de Feijão Frade	1	1	1	1	1	1	
Sardinhas Assadas	1	1	1	1	1	1	<i>Refv</i>
Corretamente Classificado (%)	88,2	88,2	88,2	94,1	94,1	94,1	94,1
Nas 2 classes mais favoráveis (%)	100,0	100,0	100,0	100,0	100,0	100,0	94,1
Nas 3 classes mais favoráveis (%)	100,0	100,0	100,0	100,0	100,0	100,0	94,1
Nas 5 classes mais favoráveis (%)	100,0	100,0	100,0	100,0	100,0	100,0	100,0
Não Identificado (%)	0,0	0,0	0,0	0,0	0,0	0,0	0,0

Tabela 16: SURF e BoW de 50 palavras por classe com Informação Global de Cor. Na coluna de referência temos os valores obtidos com a utilização de SURF com CHC-32HSI e BoW de 50 palavras por classe.

Observando os resultados da Tabela 15 e Tabela 16, podemos confirmar que a introdução de informação global de cor proporcionou melhores resultados. Como podemos constatar em diversas ocasiões anteriormente, também nesta configuração o modelo de cor HSI proporciona uma melhor descrição da informação de cor que o modelo RGB. Ao nível do tamanho do histograma, refira-se que obtivemos um melhor resultado com 256 e 512 cores.

Para finalizar a secção procedemos à análise da proposta de realizar uma normalização do número de descritores com Nearest Neighbor Joining, reduzindo-se a 150 o número de descritores por imagem. Na Tabela 17 vemos os resultados obtidos com normalização, usando como base os algoritmos SIFT e SURF, e *Bag of Words* padrão de 50 e 100 palavras. Introduzimos na tabela os valores de referência obtidos com a configuração base para simplificar a análise posterior.

Imagem de Consulta	SIFT				SURF			
	50		100		50		100	
Batatas com Bacalhau	2		2		1		----	
Caldo Verde	---		---		4		---	
Carne Assada com Batata e Arroz	1		2		2		4	
Entrecosto Grelhado com Arroz	2		3		1		1	
Feijoadada	3		1		2		2	
Frango Assado com Batata e Arroz	6		6		---		---	
Massa com Atum	1		1		1		1	
Massa Bolonhesa	1		1		1		1	
Massa de Carne	1		1		1		2	
Massa de Frango	1		2		1		1	
Panados com Arroz	4		2		1		1	
Pataniscas de Bacalhau	1		1		1		1	
Prego no Prato	4		4		5		4	
Rojões à Minhota	2		3		4		5	
Salada de Alface e Tomate	3		---		8		7	
Salada de Feijão Frade	1		1		1		1	
Sardinhas Assadas	1	Refv	1	Refv	1	Refv	1	Refv
Corretamente Classificado (%)	47,1	47,1	41,2	47,1	58,8	58,8	47,1	52,9
Nas 2 classes mais favoráveis (%)	64,7	64,7	64,7	64,7	70,6	70,6	58,8	64,7
Nas 3 classes mais favoráveis (%)	76,5	76,5	76,5	76,5	70,6	76,5	58,8	70,6
Nas 5 classes mais favoráveis (%)	88,2	82,4	82,4	82,4	88,2	82,4	76,5	76,5
Não Identificado (%)	5,9	5,9	11,8	11,8	5,9	0,0	17,6	0,0

Tabela 17: Amostragem de Descritores com NNJ. Nas colunas de referência temos respetivamente os valores obtidos utilizando SIFT com BoW de 50 palavras e 100 palavras, e SURF com BoW de 50 palavras e 100 palavras, em ambos os casos sem introdução de informação de cor.

Por análise dos resultados da Tabela 17 podemos constatar que de uma forma geral a abordagem proposta não permitiu obter resultados positivos. Especificamente com a utilização do algoritmo SURF verificamos que tanto o número de imagens não identificadas aumentou, como a precisão no reconhecimento da classe da imagem de consulta diminuiu. Embora seja expectável que uma normalização do número de descritores por imagem contribua para a obtenção de melhores resultados, podemos afirmar que este não é de todo o melhor caminho.

Em suma,

- ✓ Algoritmo SURF apresentou um melhor desempenho que o SIFT;
- ✓ Introdução de informação de cor local e global melhorou a classificação;
- ✓ CHC foi globalmente melhor que o DCC na descrição local de cor;
- ✓ Modelo de Cor HSI foi na generalidade melhor na descrição de cor que o RGB;
- ✓ Abordagem BoW por classe conseguiu ser superior à de BoW padrão;
- ✓ Um núcleo reduzido de palavras garantiu uma maior discriminação;
- χ Normalização do número de descritores com NNJ piorou a classificação;

Refira-se que os resultados obtidos **não** são integralmente determinísticos devido à forma como é construído o Bag of Words. Assim, uma vez que na construção do BoW é utilizado K-Means, com uma seleção aleatória dos clusters iniciais, é expectável que na construção de sucessivos Bag of Words para o mesmo conjunto de imagens exista uma ligeira variação nas palavras selecionadas. Uma vez tendo o BoW, a pesquisa é absolutamente determinística.

6.3. Estimação Calórica

Por forma a proceder a uma análise mais refinada do método proposto para estimação calórica, para cada um dos 17 conjuntos de imagens selecionou-se o par de imagens com a proporção e quantidade de alimentos mais semelhante. Em seguida procedeu-se à aplicação dos algoritmos de extração de características locais SIFT e SUFT e comparação do número de descritores entre cada par.

Podemos verificar na Tabela 18 e Tabela 19 os resultados obtidos com SIFT e SURF respetivamente. Nas duas primeiras colunas temos o número de descritores locais extraídos para cada imagem (I_1 e I_2) e nas últimas duas colunas a diferença entre o número de descritores.

Classes	Imagem I_1	Imagem I_2	$ I_1 - I_2 $	$\frac{2 \times I_1 - I_2 }{ I_1 + I_2 }$ (%)
Batatas com Bacalhau	252	225	27	11,3
Caldo Verde	37	40	3	7,8
Carne Assada com Batata e Arroz	687	819	132	17,5
Entrecosto Grelhado com Arroz	954	995	41	4,2
Feijoada	1012	1033	21	2,1
Frango Assado com Batata e Arroz	418	397	21	5,2
Massa com Atum	1201	1381	180	13,9
Massa Bolonhesa	1429	1431	2	0,1
Massa de Carne	791	763	28	3,6
Massa de Frango	666	667	1	0,2
Panados com Arroz	2083	2250	167	7,7
Pataniscas de Bacalhau	908	926	18	2,0
Prego no Prato	707	583	124	19,2
Rojões à Minhota	1012	922	90	9,3
Salada de Alface e Tomate	199	211	12	5,9
Salada de Feijão Frade	1932	1908	24	1,3
Sardinhas Assadas	2304	2075	229	10,5
Média				7,2
Desvio Padrão				5,7

Tabela 18: Estimação Calórica com SIFT. Nas duas primeiras colunas temos o número de descritores locais extraídos para cada imagem e nas últimas duas colunas a diferença entre o número de descritores.

Classes	Imagem I ₁	Imagem I ₂	I ₁ - I ₂	$\frac{2 \times I_1 - I_2 }{ I_1 + I_2 }$ (%)
Batatas com Bacalhau	266	291	25	9,0
Caldo Verde	54	62	8	13,8
Carne Assada com Batata e Arroz	987	1211	224	20,4
Entrecosto Grelhado com Arroz	1408	1445	37	2,6
Feijoada	1443	1681	238	15,2
Frango Assado com Batata e Arroz	524	598	74	13,2
Massa com Atum	1752	1852	100	5,6
Massa Bolonhesa	1968	2060	92	4,6
Massa de Carne	1003	1062	59	5,7
Massa de Frango	893	896	3	0,3
Panados com Arroz	2435	2699	264	10,3
Pataniscas de Bacalhau	1350	1352	2	0,2
Prego no Prato	923	796	127	14,8
Rojões à Minhota	1264	1325	61	4,7
Salada de Alface e Tomate	267	242	25	9,8
Salada de Feijão Frade	2559	2654	95	3,6
Sardinhas Assadas	2915	2747	168	5,9
Média				8,2
Desvio Padrão				5,6

Tabela 19: Estimação Calórica com SURF. Nas duas primeiras colunas temos o número de descritores locais extraídos para cada imagem e nas últimas duas colunas a diferença entre o número de descritores.

Analisando os resultados da Tabela 18 e Tabela 19 podemos verificar que de uma forma geral seguindo esta metodologia podemos proceder a uma estimaco do valor calrico com um desvio mdio inferior a 10% em relao ao valor real de calorias (em cenrios simples). Como previsto os resultados so similares tanto com a utilizao do algoritmo SIFT como com o algoritmo SURF. Refira-se novamente que no so esperados resultados positivos em cenrios complexos, como por exemplo quando a proporo de alimentos no  mantida de forma constante ou existem ocluses.

CAPÍTULO VII: CONCLUSÕES

CAPÍTULO VII: CONCLUSÕES

7.1. Comentários Finais

O projeto desenvolvido oferece um modelo para reconhecimento automático de refeições e estimação calórica. A *pipeline* da aplicação segue uma série de etapas sequenciais na construção da base de dados e pesquisa de imagens. Na construção da base de dados passamos pela deteção e descrição dos pontos de interesse, construção de um histograma para descrição da informação de cor ao nível global para cada imagem e construção do *Bag of Words*. A pesquisa na base de dados envolve a deteção e descrição dos pontos de interesse e construção de um histograma para descrição da informação de cor ao nível global para a imagem de consulta.

O estado de arte de visão computacional providencia um conjunto de ferramentas interessantes para pesquisa de imagens com base no conteúdo, as quais adaptamos para proceder ao reconhecimento automático de refeições e estimação calórica, em concreto, os algoritmos de deteção e descrição de características locais SIFT e SURF. Os algoritmos SIFT e SURF proporcionam uma descrição sólida dos pontos-chave das imagens, com especial ênfase para o algoritmo SURF, o qual garantiu os melhores resultados nos testes realizados, porém nenhum destes contempla a introdução de informação de cor tanto na deteção como na descrição dos pontos de interesse. Sendo a cor uma propriedade importante nas imagens de refeições apresentamos duas propostas para compensar a falta de informação de cor na descrição dos pontos-chave, nomeadamente descrição dos pontos-chave para cada camada de cor RGB da imagem e concatenação de um histograma de cor a cada descritor base. Dentro destas duas abordagens propostas, destaque para a concatenação de um histograma de cor ao descritor base que para além de garantir o melhor desempenho na distinção entre descritores também é consideravelmente menos custosa computacionalmente.

Ainda na descrição da cor, refira-se os histogramas de cor que descrevem a informação de cor de forma global para cada imagem, e que de uma forma geral contribuíram para aumentar a precisão no reconhecimento das refeições. Dentro do que era expectável, foi possível confirmar que o modelo de cor HSI proporciona globalmente uma melhor descrição da informação de cor, e por sua vez melhores resultados em comparação com o modelo RGB. Na manipulação dos descritores locais, uma abordagem de *Bag of Words* por classe foi claramente superior à abordagem padrão, facto resultante da manutenção da distinção entre as várias classes de refeições. Em ambas as abordagens, a utilização de um núcleo reduzido de palavras conduziu à aquisição de melhores resultados, possivelmente explicado pelo facto de ao usar menos palavras, a distinção entre palavras aumentar, conduzindo a uma maior discriminação. O modelo proposto para amostragem e normalização do número de descritores por imagem com Nearest Neighbor Joining não foi de encontro ao esperado, tendo-se obtido resultados globalmente negativos. Embora seja expectável que uma normalização do número de descritores por imagem contribua diretamente para o aumento da precisão no reconhecimento das imagens, esta abordagem mostrou não ser a mais apropriada. Globalmente e para concluir, a metodologia proposta para reconhecimento automático de refeições apresentou um bom desempenho. Note-se que a variabilidade existente em termos de diferentes refeições é elevada, e mesmo para duas refeições pertencentes à mesma classe, quando confeccionadas sobre a influência de diferentes culturas ou simplesmente diferentes pessoas, a probabilidade de estas divergirem tanto no modo de confeção como nos alimentos incluídos ou mesmo nas proporções entre alimentos, não é de todo desprezável e pode induzir em falhas nas tentativas de reconhecimento. Assim e de um ponto de vista aplicacional, para um maior controlo destes problemas, idealmente deve ser o utilizador da aplicação a criar a sua base de dados de refeições.

Ao nível da estimacão calórica a metodologia proposta, baseada na comparação entre o número de descritores locais da imagem de consulta e o número médio de descritores locais das imagens da base de dados pertencentes à mesma classe da imagem de consulta, consegue ser satisfatória em cenários relativamente simples. Porém em situações onde a proporção entre os alimentos que compõem a refeição não é mantida de acordo com a proporção padrão assim como nos casos em que existem muitas oclusões de alimentos, espera-se que os valores obtidos na estimativa se desviem muito do valor calórico real.

7.2. Trabalho Futuro

Como vimos anteriormente, a cor é sem dúvida uma característica importante nas imagens de refeições, e a introdução de informação de cor contribui de forma considerável para um melhoramento no reconhecimento das refeições, porém nem os algoritmos base adotados (SIFT e SURF) usam informação de cor nem as propostas sugeridas compensam totalmente esta lacuna uma vez que mesmo aplicando a metodologia proposta a informação de cor é desconsiderada na altura da detecção dos pontos de interesse. Assim é importante encontrar um meio alternativo que permita usar informação de cor na detecção e descrição dos pontos de interesse, quer isto implique a adaptação de um algoritmo existente ou a construção de raiz de um algoritmo que contemple informação de cor.

Um dos elementos mais críticos a exigirem trabalho futuro consiste na estimação do valor calórico das refeições. Com a metodologia apresentada apenas conseguimos uma estimativa satisfatória em cenários relativamente simples. Para uma correta estimativa em cenários complexos (e de certa forma mais realistas) é necessário proceder à divisão da refeição nos alimentos que a compõem antes de se proceder à estimação calórica. Esta ação requer naturalmente ou a utilização de métodos de segmentação seguido de nova tentativa de reconhecimento individual para cada alimento segmentado, ou uma tentativa de reconhecer com base nos descritores locais da imagem os alimentos que compõem a imagem. Aplicando o primeiro caso (segmentação), é importante ponderar sobre a necessidade de proceder a um reconhecimento prévio da refeição. Esse reconhecimento pode por uma lado favorecer a fase de segmentação e posterior reconhecimento dos alimentos, pois sabendo qual é a refeição podemos esperar encontrar à partida um determinado conjunto de alimentos, porém requer que a primeira fase de reconhecimento seja simples e fiável ou de outra forma o custo computacional começa a ficar extremamente elevado. Portanto uma possível segmentação sem reconhecimento prévio não é de descartar, embora não se espere que seja uma abordagem simples. Aplicando o segundo caso, será necessário ter em consideração que nem sempre identificar corretamente os alimentos visíveis na imagem é suficiente para deduzir a classe da refeição.

Pensando sobre um ponto de vista mais aplicacional, um elemento do qual se deve tirar partido consiste no histórico de refeições do utilizador. É expectável que na generalidade cada utilizador tenha uma dieta baseada num conjunto relativamente pequeno de classes de refeições, e que dentro deste conjunto exista algumas que sejam mais repetidas que outras. Assim como é expectável que exista uma maior probabilidade para o utilizador consumir uma determinada refeição consoante o período do dia, dia da semana, refeições recentes entre outros fatores. Desta forma, esta informação é claramente relevante e quando integrada na aplicação pode conduzir a um melhoramento do seu desempenho.

REFERÊNCIAS

- [1] Lawes, Carlene MM, et al. "High blood pressure." Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors. Geneva: World Health Organization (2004): 281-390.
 - [2] Kannel, William B., and Daniel L. McGee. "Diabetes and cardiovascular disease." JAMA: the journal of the American Medical Association 241.19 (1979): 2035-2038.
 - [3] Joint National Committee on Detection, and Evaluation, and Treatment of High Blood Pressure. The 1988 report of the Joint National Committee on detection, evaluation, and treatment of high blood pressure. No. 89. US Dept. of Health and Human Services, Public Health Service, National Institutes of Health, 1988.
 - [4] Joslin, Elliott P., et al. "The treatment of diabetes mellitus." The American Journal of the Medical Sciences 191.1 (1936): 127-130.
 - [5] Gonzalez, Rafael C., and E. Richard. "Woods, digital image processing." ed: Prentice Hall Press, ISBN 0-201-18075-8 (2002).
 - [6] Shroff, Geeta, Asim Smailagic, and Daniel P. Siewiorek. "Wearable context-aware food recognition for calorie monitoring." Wearable Computers, 2008. ISWC 2008. 12th IEEE International Symposium on. IEEE, 2008.
 - [7] Sun, Mingui, et al. "Determination of food portion size by image processing." Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE. IEEE, 2008.
 - [8] Daily Burn, Inc. Meal Snap – Calorie Counting Magic. Disponível em: <http://mealsnap.com/>. Acesso em 27 de Julho de 2013.
 - [9] Malia Frey. "Product review: Meal Snap calorie counting app for iPhone." Disponível em: <http://www.examiner.com/review/product-review-meal-snap-calorie-counting-app-for-iphone>. Acesso em 27 de Julho de 2013.
 - [10] Kato, Toshikazu. "Database architecture for content-based image retrieval." SPIE/IS&T 1992 Symposium on Electronic Imaging: Science and Technology. International Society for Optics and Photonics, 1992.
 - [11] Shyu, Chi-Ren, et al. "Local versus global features for content-based image retrieval." Content-Based Access of Image and Video Libraries, 1998. Proceedings. IEEE Workshop on. IEEE, 1998.
 - [12] Wallraven, Christian, Barbara Caputo, and Arnulf Graf. "Recognition with local features: the kernel recipe." Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on. IEEE, 2003.
-

- [13] Tuytelaars, Tinne, and Krystian Mikolajczyk. "Local invariant feature detectors: a survey." *Foundations and Trends in Computer Graphics and Vision* 3.3 (2008): 177-280.
 - [14] Lowe, David G. "Distinctive image features from scale-invariant keypoints." *International journal of computer vision* 60.2 (2004): 91-110.
 - [15] Lowe, David G. "Object recognition from local scale-invariant features." *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Vol. 2. Ieee, 1999.
 - [16] Bay, Herbert, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features." *Computer Vision–ECCV 2006*. Springer Berlin Heidelberg, 2006. 404-417.
 - [17] Suhasini, P. S. Dr. IV Murali Krishna, "CBIR using color histogram processing". *Journal of theoretical and applied information technology* 6.1 (2009): 116-122.
 - [18] Nowak, Eric, Frédéric Jurie, and Bill Triggs. "Sampling strategies for bag-of-features image classification." *Computer Vision–ECCV 2006*. Springer Berlin Heidelberg, 2006. 490-503.
 - [19] Hartigan, John A., and Manchek A. Wong. "Algorithm AS 136: A k-means clustering algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1 (1979): 100-108.
 - [20] Bradski, Gary, and Adrian Kaehler. *Learning OpenCV: Computer vision with the OpenCV library*. O'reilly, 2008.
 - [21] Larose, Daniel T. "k-Nearest Neighbor Algorithm." *Discovering Knowledge in Data: An Introduction to Data Mining* (2005): 90-106.
-

ANEXOS

Anexos

ANEXO A: Conversão entre espaço de Cor RGB e HSI

A conversão entre o modelo RGB e o modelo HSI é efetuada da seguinte forma: (**R**, **G** e **B** os valores respetivamente de Vermelho, Verde e Azul, e **H**, **S**, **I** os valores respetivamente de Matiz, Saturação e Intensidade).

- **RGB para HSI:**

$$Matiz = \begin{cases} \theta, & G \geq B \\ 2\pi - \theta, & G < B \end{cases}, \theta = \cos^{-1} \left\{ \frac{\frac{1}{2}[(R-G) + (R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\}$$

$$Saturação = 1 - \frac{3}{(R + G + B)} [Min(R, G, B)]$$

$$Intensidade = \frac{1}{3} \times (R + G + B)$$

- **HSI para RGB:**

$$X = I \times (1 - S) \quad Y = I \times \left[1 + \frac{S \cos H}{\cos\left(\frac{\pi}{3} - H\right)} \right] \quad Z = 3 \times I - (X + Y)$$

- **Se:** $0 \leq H < \frac{2\pi}{3}$

$$\begin{cases} Vermelho = Y \\ Verde = Z \\ Azul = X \end{cases}$$

- **Se:** $\frac{2\pi}{3} \leq H < \frac{4\pi}{3} \rightarrow H = H - \frac{2\pi}{3}$

$$\begin{cases} Vermelho = X \\ Verde = Y \\ Azul = Z \end{cases}$$

- **Se:** $H \geq \frac{4\pi}{3} \rightarrow H = H - \frac{4\pi}{3}$

$$\begin{cases} Vermelho = Z \\ Verde = X \\ Azul = Y \end{cases}$$

ANEXO B: Base de Dados

Imagens usadas na Construção da Base de Dados					Imagem de Consulta
					
					
					
					
					
					
					
					



Tabela 20: Imagens utilizadas durante a análise à metodologia proposta. Os pares de imagens com bordas vermelhas foram utilizados para a estimaco calrica.