
**IDENTIFICAÇÃO DE ESPÉCIES DE PEIXES
UTILIZANDO HISTOGRAMAS DE PALAVRAS
VISUAIS EM IMAGENS COLORIDAS**

Uéliton de Paula Freitas

Pós -GRADUAÇÃO DA UNIVERSIDADE FEDERAL DE MATO GROSSO DO SUL DA FACULDADE DE CIÊNCIAS DA COMPUTAÇÃO - FACOM

Data de Depósito:

Assinatura:_____

IDENTIFICAÇÃO DE ESPÉCIES DE PEIXES UTILIZANDO HISTOGRAMAS DE PALAVRAS VISUAIS EM IMAGENS COLORIDAS

Uéliton de Paula Freitas

Orientador: *Prof^o Dr^o Hemerson Pistori*

Dissertação apresentada ao curso de Pós Graduação em Ciências da Computação, da Universidade Federal de Mato Grosso do Sul, como requisito parcial à obtenção do título de Mestre em Ciências da Computação.

UFMS - Campo Grande
março/2015

Agradecimentos

- Agradeço a Deus por nunca ter me abandonado.
- Hemerson Pistori, por ser meu guia com paciência e compreensão.
- Ao Prof. Edson Takashi, por mais uma vez fazer parte de um acontecimento importante em minha vida.
- Aos pertencentes à banca, José Sabino e Wesley Nunes pelas contribuições.
- Aos meus amigos Edilson e Júnior, por estarem ao meu lado.
- A minha mãe, Nedina, por sempre acreditar em mim.
- A Franciele e Francieny, por me ajudarem tantas vezes nos momentos difíceis.
- A Tecsinapse, principalmente ao Paulo, Thales e Thiago pela ajuda no desenvolvimento da aplicação.
- A CAPES, pela ajuda financeira.

Resumo

Neste trabalho é apresentada uma aplicação voltada para dispositivos móveis cujo objetivo é classificar espécies de peixes por meio de técnicas de Visão Computacional e Inteligência Artificial utilizando imagens. A aplicação foi desenvolvida para *smartphones Android* e conta com o auxílio da biblioteca de Visão Computacional OpenCV tanto na fase de classificação quanto de treinamento. As técnicas empregadas na descrição das imagens são baseadas em Histogramas de Palavras Visuais aplicados em imagens coloridas. São elas: Histograma de Palavras Visuais (*Bag of Visual Words* - BoVW), Histograma de Atributos e Cores (*Bag of Features and Colors*), Histograma de Cores de Wengert (*Bag of Colors* - BoC), Histograma de Palavras Coloridas (*Bag of Colored Words* - BoCW) e Histogramas de Cores nos espaços de cores RGB e HSV. Para a classificação das espécies, foram utilizados três tipos de classificadores: Máquina de Vetores de Suporte (*Support Vector Machine* - SVM), Árvore de Decisão e os K vizinhos mais próximos (K-NN). Nos experimentos foram variados os parâmetros de todos os classificadores a fim de encontrar os melhores resultados para a classificação. Para comparar o desempenho das técnicas de extração de atributos, assim como dos classificadores, foi utilizada a métrica Medida-F (*F-Score*) como métrica principal e Área Sobre a Curva (AUC) como métrica auxiliar. A técnica com melhor resultado foi a BoC, baseada somente em informações de cores, obteve Medida-F igual a 0.9 e AUC 0.98 utilizando o classificador SVM.

Palavras Chave: Histograma de Palavras Visuais. Visão Computacional. Inteligência Artificial. Peixes. *Android*.

Abstract

This work presents an application dedicated to mobile devices whose objective is to classify species of fish concepts of Computer Vision and Artificial Intelligence using images. The application was developed to Android smartphones with the help of OpenCV Computer Vision library for classification and training phases. The techniques employed in the description of the images are based on Bag of Visual Words applied to color images. They are: HSV and RGB color histograms, Bag of Visual Words, Bag of Features and Colors, Bag of Colors and Bag of Colored Words (BoCW). For the species classification, three types of classifiers was used: Support Vector Machine (SVM), Decision Tree and K-Nearest Neighbors algorithm (K-NN). In the experiments it was tested several parameters of all the classifiers in order to find the best results for classification. To compare the performance of the feature extraction techniques, as well as the classifiers, the metric F-Score were used as the main metric and the Area Under the Curve (AUC) as an auxiliary metric. The technique with best result was BoC, based only on color information, obtained F-Score equal to 0.9 and 0.983348 for AUC using the SVM classifier.

Keywords: Bag of Words. Computer Vision. Artificial Intelligence. Fish. Android.

Sumário

Sumário	v
Lista de Abreviaturas	vi
Lista de Figuras	ix
Lista de Tabelas	x
1 Introdução	1
2 Descritores de Atributos	5
2.1 Extratores de Cores RGB e HSV	5
2.2 <i>Speeded Up Robust Features</i> (SURF)	7
2.2.1 Detecção de Pontos de Interesse	8
2.2.2 Detecção de Pontos de Interesse Utilizando Convolução	10
2.2.3 Detecção de Pontos de Interesse em Diferentes Escalas	11
2.2.4 Descrição dos Pontos de Interesse	14
2.3 Histograma de Palavras e Histograma de Palavras Visuais	16
2.4 Histograma de Atributos e Cores	17
2.5 Histograma de Palavras Coloridas	18
2.6 Histograma de Cores de Wengert	19
2.6.1 Dicionário de Cores	19
2.6.2 Assinatura de Cores	19
2.6.3 Normalizações	20
3 Materiais e Métodos	21
3.1 Classificadores	21
3.2 Banco de Imagens	23
3.3 Métricas	28
3.3.1 Característica de Operação do Receptor - (Receiver Operating Characteris- tics - ROC)	28
3.3.2 Medida-F (F-Score)	30
4 Experimentos	32
4.1 Variações dos Parâmetros dos Classificadores	32
4.2 Determinação dos Parâmetros dos Extratores de Atributos	33
4.2.1 Experimentos com o BoVW	33
4.2.2 Experimentos com o BoFC	35
4.2.3 Experimentos com o BoC	37
4.2.4 Experimentos com o BoCW	39

4.2.5 Experimentos com o Histograma de Cores HSV	42
4.2.6 Experimentos com o Histograma de Cores RGB	44
5 Resultados e Discussões	46
6 Aplicação de Reconhecimento de Espécies de Peixes - Pantanal	50
7 Conclusão	53

Lista de Abreviaturas

AUC *Area Under ROC Curve* (Área Sob a Curva ROC).

VP Verdadeiro Positivo.

FP Falso Positivo.

VN Verdadeiro Negativo.

FN Falso Negativo.

TFP Taxa de Falso Positivo.

TVP Taxa de Verdadeiro Positivo.

SVM *Support Vector Machines*.

BoW Histograma de Palavras.

BoVW Histograma de Palavras Visuais.

BoC *Bag of Colors* (Histograma de Cores de Wengert).

BoFC *Bag of Features and Colors* (Histograma de Atributos e Cores).

BoCW *Bag of Colored Words* (Histograma de Palavras Coloridas).

HHSV Histograma de Cores no Espaço HSV.

HRGB Histograma de Cores no Espaço RGB.

SURF *Speeded Up Robust Features*

SIFT *Scale-Invariant Feature Transform*

Lista de Figuras

1.1	Aquário do Pantanal sendo construído na cidade de Campo Grande, Mato Grosso do Sul.	2
2.1	Variação da cor laranja de um tom mais claro para um mais escuro no espaço RGB.	6
2.2	Variação da cor laranja para um tom escuro. Para tal fim, como ilustrado, basta variar o componente V no espaço de cores no espaço HSV.	6
2.3	Exemplo de um peixe fictício.	8
2.4	Exemplo de pontos de interesse detectados pelo SURF e utilizando FLANN (<i>Fast Library for Approximate Nearest Neighbors</i>) para comparação dos pontos de interesse.	8
2.5	Representação da imagem ilustrada na Figura 2.3 como uma função de duas variáveis.	9
2.6	Exemplo da aplicação de um filtro caixa utilizado para detectar variação nos valores dos pixels no eixo x em relação ao pixel central da sub imagem (pixel com um círculo vermelho). Para fins ilustrativos a imagem está colorida, porém, no SURF, o processo é feito utilizando a imagem em tons de cinza.	11
2.7	Núcleos de Convolução do SURF.	11
2.8	Figura ilustrando a variação de escala de um ponto de interesse. Caso o fator de escala não seja levando em consideração, o mesmo ponto não será encontrado em uma imagem maior (com uma resolução maior).	12
2.9	Pirâmides de escala.	12
2.10	Exemplo do uso de imagens integrais.	13
2.11	Oitavas variando exponencialmente em relação ao eixo y e aritmeticamente no eixo x (6 em 6, 12 em 12, 24 em 24).	14
2.12	Regiões de maior variância em diferentes escalas. As linhas amarelas indicam a direção da variação. O tamanhos dos círculos estão relacionados às escalas em que os pontos foram encontrados.	15
2.13	Núcles de convolução utilizados na detecção de variância de rotação.	15
2.14	Gráfico ilustrando a variação dos valores dos pixels de uma região em volta de um ponto de interesse. O centro da circunferência é posicionado na coordenada do ponto de interesse.	16
2.15	Distribuição de Haar Wavelets.	16

2.16	Exemplo da aplicação do BoVW em um conjunto de imagens formando um dicionário de 4 palavras correspondendo aos centroides do K-Médias. Um histograma é então formado a partir da contagem dos pontos de interesse pertencentes a cada palavra visual.	18
2.17	Dicionário de cores criado a partir do BoC ilustrando as cores mais frequentes do banco de imagens . As palhetas correspondem a $k_c = 64$ respectivamente. . . .	20
3.1	Exemplos de imagens do ambiente de captura das casas de peixes ornamentais de Campo Grande, Mato Grosso do Sul.	24
3.2	Imagem do Aquário Municipal de Toledo, Paraná.	24
3.3	Ilustração de todas as espécies de peixes com seus nome populares e nomes científicos.	25
3.4	Amostras contendo 6 imagens das 40 de 11 espécies de peixes.	26
3.5	Amostras contendo 6 imagens das 40 de 12 espécies de peixes.	27
3.6	Amostras contendo 6 imagens das 40 de 5 espécies de peixes.	28
3.7	Matriz de confusão ou de contingência.	29
4.1	Área sob a curva do dicionário de tamanho 4096 do BoVW utilizando o SVM como classificador.	34
4.2	Matriz de confusão da classificação das espécies utilizando o dicionário de tamanho 4096 do BoVW juntamente com o classificador SVM.	34
4.3	Área sob a curva do dicionário de tamanho 2048 com a técnica BoFC utilizando o SVM como classificador.	35
4.4	Matriz de confusão da classificação das espécies utilizando o dicionário de 2048 com a técnica BoFC e o classificador SVM.	36
4.5	Área sob a curva do dicionário de tamanho 2048 do BoC utilizando o SVM como classificador.	37
4.6	Matriz de confusão da classificação das espécies utilizando a técnica BoC, o dicionário de tamanho 2048, e o classificador SVM.	38
4.7	Área sob a curva do dicionário de tamanho 1024 com 32 faixas do BoCW utilizando o SVM como classificador.	40
4.8	Matriz de confusão da classificação das espécies utilizando a técnica BoCW, o dicionário de tamanho 1024, 32 faixas e o classificador SVM.	41
4.9	Área sob a curva da divisão do espaço de cor HSV em 32 faixas.	42
4.10	Matriz de confusão do histograma de cores HSV utilizando SVM como classificador e 32 faixas.	43
4.11	Área sob a curva da divisão do espaço de cor RGB em 16 faixas.	44
4.12	Matriz de confusão utilizando o SVM como classificador e o Histograma de cor RGB com 16 faixas.	45
5.1	Diagrama de caixas das técnicas utilizando Medida-F.	47
5.2	Diagrama de caixas das técnicas utilizando Medida-F por espécies.	47
5.3	Diagrama das caixas do teste <i>post hoc</i> usando Medida-F.	48
5.4	Exemplo de pontos de interesse encontrados em uma imagem do peixe Palhaço.	49
6.1	Tela inicial e do álbum vazio da aplicação.	51
6.2	Resposta de acerto ou erro de uma suposta espécie de Peixe. A Figura 6.2.1 ilustra o acerto da espécie Dourado. Já a Figura 6.2.2 ilustra um erro do jogador ao classificar a espécie Kinguio.	51
6.3	Imagem do album após a captura de um Dourado.	52

7.1	Semelhança de cores nas imagens entre as espécies Kinguio, Papagaio e Paulistinha.	55
7.2	Semelhança de cores nas imagens entre as espécies Platy Sangue e Kinguio Kor-raco.	56

Lista de Tabelas

4.1	Árvore de decisão juntamente com BoVW.	33
4.2	Comparação de Tamanhos de Dicionários do BoFC.	35
4.3	Comparação de Tamanhos de Dicionários do BoC.	37
4.4	Comparação de Tamanhos de Dicionários do BoCW utilizando 2 faixas.	39
4.5	Comparação de Tamanhos de Dicionários do BoCW utilizando 4 faixas.	39
4.6	Comparação de Tamanhos de Dicionários do BoCW utilizando 8 faixas.	39
4.7	Comparação de Tamanhos de Dicionários do BoCW utilizando 16 faixas.	40
4.8	Comparação de Tamanhos de Dicionários do BoCW utilizando 32 faixas.	40
4.9	Comparação de quantidade de faixas do Histograma de cores HSV.	42
4.10	Comparação de quantidade de faixas do Histograma de cores RGB.	44
5.1	Comparação da Medida-F de cada classificador.	46
5.2	Teste <i>post hoc</i>	48
5.3	Teste de Tukey utilizando Medida-F.	48
7.1	Matriz de confusão do BoC com dicionário de tamanho 2048 e classificador SVM.	54

Introdução

A fauna e flora brasileira destacam-se mundialmente por sua diversidade, assim, catalogar espécies de animais e plantas de forma ágil é uma tarefa difícil e árdua. Com tanta diversidade, dado um determinado animal, mesmo utilizando livros e bancos de dados digitais, leva tempo para identificá-lo. São necessários anos de estudo por um profissional para se especializar em um determinado animal. Sendo assim, áreas da ciência como a Visão Computacional buscam facilitar tarefas como a apresentada, por meio da sua autorização.

Na cidade de Campo Grande, Mato Grosso do Sul, está sendo construído o Aquário do Pantanal sendo o maior aquário nacional e o primeiro brasileiro de porte internacional com mais de 27 mil metros quadrados, 16 grandes aquários dentro do prédio e 5 na área externa, estarão presentes em cerca de 6.200.000 litros de água, 135 espécies da fauna aquática. Além de objetivos científicos, o aquário possui propósitos turísticos e educativos que serão disponíveis para toda a população. Explorando estes propósitos, este trabalho apresenta uma análise de técnicas de Visão Computacional e Aprendizado de Máquina para descrever e classificar imagens de peixes, juntamente com uma aplicação final em que, dada uma imagem de um peixe, o programa é capaz de identificar sua espécie.

Visão Computacional [20] é uma subárea de estudo da Ciência da Computação que, juntamente com o Aprendizado de Máquina [19], tem como objetivo “ensinar” o computador a interpretar imagens. Um dos principais desafios na Visão Computacional é encontrar o melhor meio de descrever uma imagem, de modo que ela seja discriminativa o suficiente para ser identificada entre várias outras. Há vários modos de descrever uma imagem dependendo da aplicação em que se está trabalhando: cor dos pixels, tamanho da imagem, vetores de atributos, entre outras. Cada método de descrição possui seus pontos fortes e fracos em relação ao tempo de execução e processamento. As técnicas devem ser escolhidas (ou utilizadas em conjunto) dependendo da aplicação e do dispositivo utilizado.

Um dos problemas encontrados na execução de técnicas de Visão Computacional em dispositivos móveis é em relação ao tempo de processamento e quantidade de memória utilizada para executar tais técnicas. Assim, dispositivos móveis eram evitados. Mas com os avanços de *hardware* encontrados nos *Tablets* e *Smartphones*, em específico, melhorias nas câmeras e capacidade de processamento, surgem cada vez mais aplicações que utilizam conceitos de visão computacional. Assim, a aplicação desenvolvida utilizando os experimentos apresen-



Figura 1.1: Aquário do Pantanal sendo construído na cidade de Campo Grande, Mato Grosso do Sul.

tados neste trabalho é voltada para plataformas móveis, em específico, para *Smartphones*. Inicialmente a aplicação está disponível apenas para aparelhos Android.

O foco principal da aplicação consiste em tirar uma foto de um determinado peixe em seu ambiente (no aquário) e informar ao usuário qual a espécie do peixe em questão. O uso da aplicação é destinado ao meio turístico e educacional, então, foram adicionadas funcionalidades tornando-a em um jogo de nome Pantanalbun a fim de atingir o público jovem. O jogo consiste em um álbum de figurinhas onde cada figurinha corresponde a uma espécie de peixe. O papel do usuário é colecionar as figuras, tirando fotos das espécies que faltam em seu álbum percorrendo os aquários do local. A aplicação foi desenvolvida com auxílio da Tecsinapse, uma empresa parceira a este projeto.

Foram encontrados na literatura estudos relativos à classificação de espécies de peixes. Nery et al. [6] relatam que a classificação de peixes não é uma tarefa fácil, segundo o mesmo, peixes possuem 47 características que os definem, tais como coloração, largura e comprimento. Além disso, as imagens em aquários são obtidas, normalmente, com diferentes iluminações dificultando ainda mais a classificação. Utilizando um classificador bayesiano e vetores de atributos providos das características citadas, os autores apresentaram uma precisão na classificação superior a 90% utilizando seis espécies de peixes.

Rodrigues [5] utilizou um sistema automático baseado na Análise de Componentes Principais [13] e o SIFT(*Scale-Invariant Feature Transform*) para a parametrização de forma, aparência e movimento das espécies. Utilizou dois sistemas imunológicos artificiais (*Artificial Immune Network* e *Adaptive Radius Immune Algorithm*) para agrupar as características das espécies. Obteve precisão de 92% utilizando PCA e *Adaptive Radius Immune Algorithm* com classificador KNN em nove espécies de peixes.

Para classificar espécies de peixes como sendo venenosas ou não, Alsmadi et al. [15] utilizou combinações de atributos juntamente com o uso de segmentação de imagem e técnicas de parâmetros geométricos. Posteriormente, ele utilizou redes neurais para definir pesos para cada extrator de atributos encontrando os extratores que mais discriminam as espécies. Como classificador foi utilizado uma árvore de decisão. Ele obteve uma precisão de 97,4% classificando sete espécies de peixes.

Huang et al. [16] utilizou a combinação de sessenta e seis tipos de atributos como cor, forma e textura da calda, cabeça, parte superior e inferior dos peixes. Utilizando um classificador hierárquico e um arcabouço baseado em uma árvore de decisão balanceada otimizada

(utilizada para diminuir o erro da classificação hierárquica) obteve uma média de precisão de 90% na classificação de dez espécies de peixes.

Hu et al. [17] classificou espécies de peixes utilizando dispositivos móveis para a piscicultura na China. Ele utilizou a câmera dos dispositivos móveis e extraiu atributos das imagens baseados na cor e textura. As imagens foram enviadas para uma aplicação em uma estação de trabalho onde eram classificadas a partir de recortes feito nas mesmas. Para cada recorte foram extraídos cinco atributos de cores correspondendo a média e variância dos pixels no espaço RGB e cinco no espaço HSV. Para extrair a textura dos recortes foram utilizados extratores de textura estatísticos, são eles: histogramas de escalas de cinza (*grayscale histogram*) e matriz de co-ocorrência em nível de cinza (*gray level co-occurrence matrices*). Na classificação foram comparadas várias máquinas de suportes de suporte (SVMs). Hu et al. obteve a melhor média de precisão das classes utilizando como classificador a máquina de vetores de suporte para multi classes baseada em votos VBMSVM (*voting based multi-class support vector machine*) com média de acurácia igual a 97.96 avaliando seis espécies.

Utilizando máquinas de vetores de suporte, e analisando a textura e forma de duas espécies de peixes, Rova et al. [4] obteve precisão de 90%. Assim, é observado que a classificação de espécies de peixes por meio de imagens (mesmo com imagens com baixa qualidade) possui uma exatidão considerável.

Este trabalho descreve experimentos feitos com técnicas de descrição de atributos em imagens de vinte e oito espécies de peixes. Relata também, o desempenho de três classificadores na classificação de estas espécies em combinação com extratores de atributos. Foram comparadas técnicas de extração de atributos que descrevem imagens coloridas ou em tons de cinza, assim como várias configurações de parâmetros de cada classificador. A combinação com melhor desempenho em relação a classificação das espécies foi implementada em uma aplicação que tem por objetivo classificar espécies peixes por meio de imagens.

Técnicas como o SURF (*Speeded-up Robust Features*) [10] e SIFT (*Scale-Invariant Feature Transform*) [9] que descrevem as principais áreas das imagens, são técnicas bastante utilizadas em diversas aplicações, contudo, descartam as informações de cores que as imagens possuem. Sendo assim, também foram experimentadas técnicas que unem as informações obtidas do SURF, e agregam características de cores, assim como, técnicas que utilizam apenas cores.

Como técnica que não utiliza informações de cor, o Histograma de Palavras Visuais (*Bag of Visual Words - BoVW*) foi utilizado, descrevendo as imagens com base nas regiões de interesse geradas pelo SURF. O BoVW é uma técnica baseada no *Bag of Words*, que por sua vez, é utilizada principalmente na descrição de textos. Já como técnicas que utilizam cores juntamente com outras informações, o Histograma de Atributos e Cores (*Bag of Features and Colors - BoFC*) utiliza conceitos baseados no BoVW acrescentando informações de cores às áreas de interesse geradas pelo SURF. Uma técnica denominada Histograma de Palavras Coloridas (*Bag of Colored Words - BoCW*) também foi implementada e une a descrição das imagens fornecidas pelo SURF juntamente com informações de cores providas de um histograma de cor no espaço HSV. Também foram estudadas técnicas que utilizam apenas cores para descrever as imagens, são elas: O Histograma de Cores de Wengert [8] (*Bag of Colors - BoC*), que quando utilizado como descritor global utiliza apenas informações de cores, criando assinaturas que descrevem as imagens, e a utilização de histogramas de cores nos espaços de cores HSV e RGB. Definidas as técnicas de descrição das imagens, são aplicados algoritmos de aprendizado automático para então classificar novas imagens em suas respectivas espécies.

As técnicas de extração de atributos foram aplicadas em 28 espécies de peixes. Cada espécie foi representada com 40 imagens totalizando 1200 imagens. As imagens foram obtidas por meio de câmeras fotográficas e câmeras de celulares. Inicialmente, haviam vários peixes

nas imagens, contudo, para a extração das características de cada espécie, as imagens foram recortadas de modo a haver apenas um peixe por imagem. Este mesmo tratamento na imagem é feito na aplicação final, onde o usuário encaixa o peixe em uma quadrado e o restante do fundo da imagem é descartado, diminuindo a quantidade de ruído.

Para identificar as espécies de peixes, foram utilizados os classificadores baseados em árvore de decisão, Máquina de Vetor de Suporte (*Support Vector Machine*) e os K vizinhos mais próximos. A árvore de decisão foi escolhida devido à simplicidade do modelo utilizado na classificação. Os K vizinhos mais próximos são amplamente utilizados para classificar imagens, e foram utilizados nas avaliações de desempenho dos trabalhos BoC [8] e BoFC [7]. Por fim, máquina de vetores de suporte também são populares na classificação de objetos, e um exemplo de uso é encontrado em Rova et al. [4].

Foram utilizadas duas métricas para medir o desempenho dos classificadores e extratores de atributos. Como métrica principal, a Medida-F (*F-Scores*) foi utilizada, e com base na mesma, os melhores parâmetros dos classificadores foram escolhidos. Também foi utilizada a Área Sob a Curva (*Area Under the Curve* - AUC) como métrica auxiliar e critério de desempate na escolha dos parâmetros. Utilizando as técnicas de extração de atributos descritas, assim como os classificadores, a técnica com melhor desempenho foi o BoC utilizando a Máquina de Vetores de Suporte como classificador.

O Capítulo 2 descreve os descritores utilizados nos experimentos. O Capítulo 3 descreve os classificadores utilizados e os ajustes de parâmetros dos mesmos. Também no Capítulo 3 é descrito o banco de imagens e as métricas utilizadas para a comparação das técnicas. Os experimentos são relatados no Capítulo 4. Os resultados e discussões são apresentados no Capítulo 5 e uma conclusão no Capítulo 7. A aplicação criada a partir dos experimentos mostrados neste trabalho é apresentada no Capítulo 6.

Descritores de Atributos

Para identificar características de uma imagem, descritores são utilizados para reunir informações sobre as mesmas. A forma como uma imagem será descrita depende da aplicação, e pode ser com descritores baseados em cor, pontos de interesse, detectores de bordas, histogramas, entre outros. Alguns descritores como o Histograma de Cores de Wengert [8] (*Bag of Colors* - BoC) e histogramas de cor nos espaços de cores RGB e HSV utilizam apenas cores para descrever imagens. Contudo, juntamente com *Speeded Up Robust Features* [10] (SURF) e *Scale-Invariant Feature Transform* [9] (SIFT), descritores como o Histograma de Palavras Visuais (*Bag of Visual Words* - BoVW), Histograma de Atributos e Cores [7] (*Bag of Features and Colors* - BoFC) e Histograma de Palavras Coloridas (*Bag of Colored Words* - BoCW) buscam pontos de interesse nas imagens a fim de encontrar e descrever regiões que podem ser de maior relevância.

A Seção 2.1 descreve os extratores de atributos RGB e HSV que utilizam apenas informações de cor para descrever as imagens por meio de histogramas de cores. A Seção 2.2 descreve o SURF, um dos extratores de atributos mais utilizados para descrição de características de imagens. O SURF é utilizado em três outros extratores como extrator local, o BoVW, BoFC e BoCW descritos nas Seções 2.3, 2.4 e 2.5, respectivamente. Assim como os extratores de cor HSV e RGB, o BoC, detalhado na Seção 2.6, também utiliza apenas informações de cor.

2.1 Extratores de Cores RGB e HSV

Uma das características que diferem espécies de peixes é a cor, assim, extratores de atributos baseados somente em cor também foram utilizados, sendo eles: histogramas de cores nos espaços de cores HSV e histogramas de cores nos espaços de cores RGB.

Amplamente utilizado em dispositivos eletrônicos, o espaço de cor RGB foi um dos escolhidos para representar cor nos experimentos. A representação neste espaço de cor se dá por valores que variam de 0 a 255 para as cores vermelho (*Red*), verde (*Green*) e azul (*blue*). Assim, a combinação destas três cores primárias formam todas as outras. Mas no espaço de cor RGB é difícil analisar variações de cores e suas tonalidades. Por exemplo: seja a cor formada pelos valores RGB, 255, 139 e 39 correspondendo a um tom da cor laranja (Figura 2.1), deseja-se determinar uma cor que também seja laranja, mas um laranja “escuro” (supondo

que a iluminação sobre a cor tenha diminuído). Para tal fim, é necessário alterar os valores de R, G e B para 169, 92 e 26 respectivamente. Note que foram necessárias alterações em todos os componentes do espaço de cor.

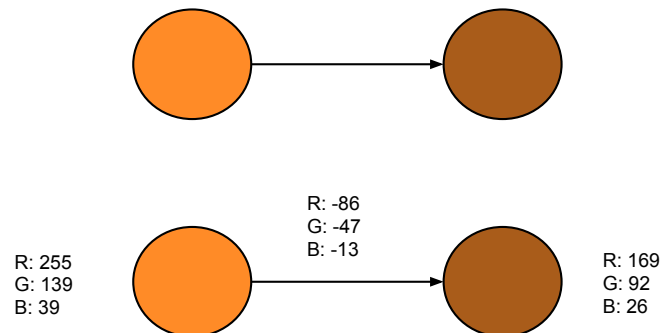


Figura 2.1: Variação da cor laranja de um tom mais claro para um mais escuro no espaço RGB.

O espaço de cores HSV (*Hue*, *Saturation* e *Value*) ou HSB (*Hue*, *Saturation* e *Brightness*) apresenta uma abordagem mais simples em relação a variações de tonalização de cores, isto é, a Matiz (*Hue*) varia a cor e o restante das componentes variam as tonalidades e pureza das cores. A Matiz corresponde à variação de todas as cores do espaço, e possui um valor em graus variando de 0° a 360° . A *Saturation* (Saturação) varia a pureza da cor variando de 0 a 100. Já o *Value* (Valor) ou *Brightness* (Brilho), dependendo de seu valor (variando de 0 a 100), torna a cor mais escura. As cores no espaço de cor HSV são principalmente formadas pelos componentes H e S. Para a mesma cor citada no exemplo anterior temos os seguintes valores HSV respectivamente: 28, 85 e 100 (Figura 2.2). Para tornar o laranja mais escuro basta variar o componente *Value*, alterando os valores de apenas uma componente.

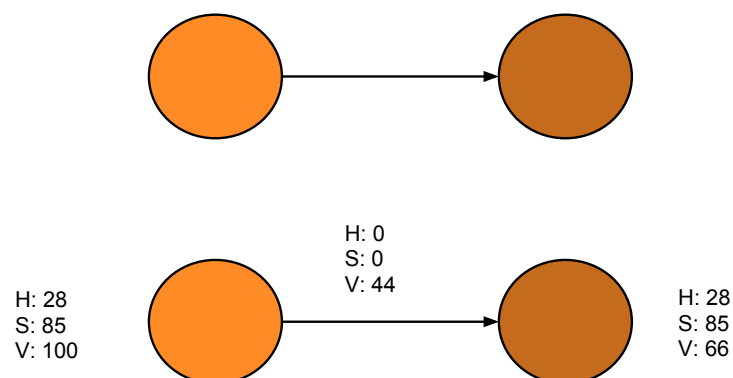


Figura 2.2: Variação da cor laranja para um tom escuro. Para tal fim, como ilustrado, basta variar o componente V no espaço de cores no espaço HSV.

Portanto, suponha um problema em que a cor é um atributo fundamental na descrição de

um conjunto de imagens, e há interferência de iluminação, a influência de uma luz branca por exemplo. Todas as cores ficarão mais claras, isto é, se o espaço de cor for o HSV, os componentes H e S serão os mesmos e apenas o componente V variará. Sendo assim, pode ser interessante ignorar o componente V e diminuir o impacto da iluminação no problema. Já no espaço de cores RGB, amenizar este tipo de interferência não é trivial. Devido ao modo de como o espaço de cor HSV representa a variação das tonalidades de cores, em especial a influência da iluminação, e ao fato de que os aquários possuem vários tipos de iluminação, ele foi um dos espaços de cores escolhidos.

Um método relativamente simples para representar as cores de uma imagem é a contagem das ocorrências das mesmas. Mas para cada imagem, o vetor de atributos será imenso utilizando este método. No espaço de cor RGB, por exemplo, o vetor terá um tamanho de 256^3 , ou seja, 16.777.216 posições, sendo que, grande parte das cores podem não possuir ocorrência. Assim, para tentar agrupar cores e diminuir o tamanho do vetor de atributos, a quantidade total de cores é dividida em intervalos. Para ambos os espaços de cores HSV e RGB, cada canal (componente) de cor é dividido em intervalos denominados faixas. Se o canal de cor R (espaço de cores RGB), por exemplo, for dividido em 2 faixas, valores que variam de 0 a 255 serão transpostos para uma variação de 0 a 1. Assim, os 256 valores possíveis do canal R serão reduzidos a apenas 2. Esta discretização de informação de cor agrupa cores que possuem valores semelhantes em suas componentes. Este método é aplicado para todos os canais de cores de um espaço de cor. Seguindo o exemplo, 2 faixas para R, G e B, formará um histograma tridimensional de 8 posições ($2 \times 2 \times 2$).

Para cada imagem, foi feito um histograma no espaço de cores RGB e outro no espaço de cores HSV. As quantidades de faixas, neste trabalho, variam entre 8, 16 e 32 para cada canal de cor. Sendo assim, o tamanho do vetor de atributos de cada imagem varia entre 512 ($8 \times 8 \times 8$), 4096 ($16 \times 16 \times 16$) e 32.768 ($32 \times 32 \times 32$). Mas técnicas mais robustas, como o SURF, também foram utilizadas para descrever as imagens dos peixes.

2.2 *Speeded Up Robust Features (SURF)*

O SURF (*Speeded Up Robust Features*) [10] é um algoritmo baseado no SIFT [9] que busca caracterizar unicamente o conteúdo de uma imagem encontrando regiões de interesse denominadas pontos de interesse (*Key Points*). Mesmo que a imagem à qual o SURF é aplicado seja colorida, ele a converte em tons de cinza para encontrar as regiões mais interessantes, tornando a cor um atributo irrelevante para a busca. Outra característica importante é a tolerância dos pontos de interesse a variações de escala e rotação. Assim, mesmo sofrendo rotações e alterações de tamanho, o mesmo conteúdo da imagem pode ser encontrado em outras imagens por meio da comparação dos pontos de interesse.

A Figura 2.4 ilustra a detecção e comparação dos pontos de interesse entre duas imagens de um mesmo peixe. A imagem original é colorida e está ilustrada na Figura 2.3. Inicialmente, os pontos foram detectados na imagem mais à esquerda da Figura 2.4, e posteriormente, a imagem sofreu uma rotação (imagem mais à direita da Figura 2.4). É possível notar que alguns pontos, em ambas as imagens, foram os mesmos, evidenciando a semelhança entre elas.

O SURF possui duas fases principais, detecção e descrição de pontos de interesse. A primeira fase consiste em detectar os pontos utilizando conceitos de imagens integrais e matriz Hessiana. Também na detecção dos pontos, é utilizada a supressão de não máximo em 3D, que por sua vez, encontra os pontos mesmo eles estando em outras escalas. Após a detecção, inicia-se o processo de descrição dos pontos utilizando o cálculo de orientação dominante,

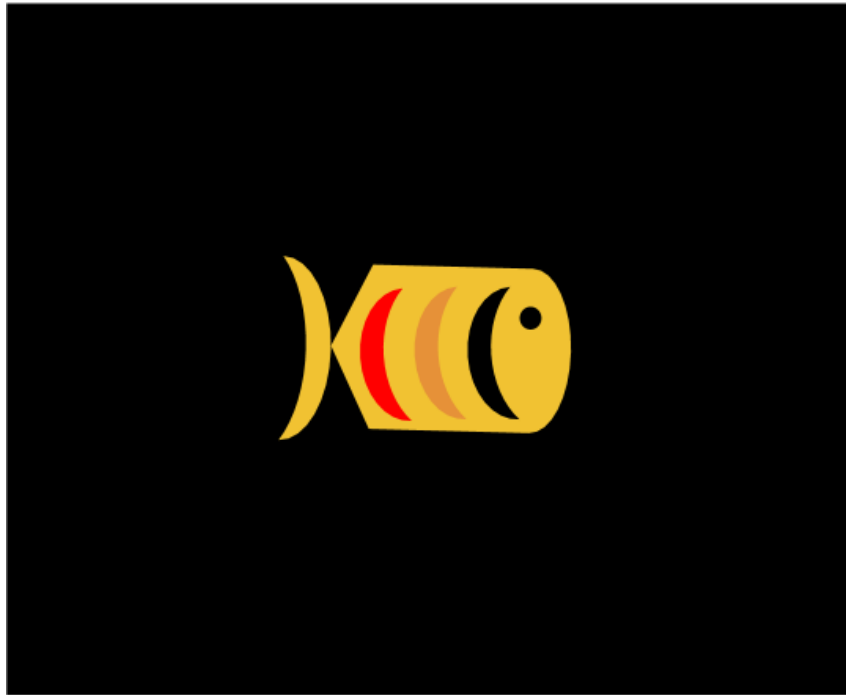


Figura 2.3: Exemplo de um peixe fictício.

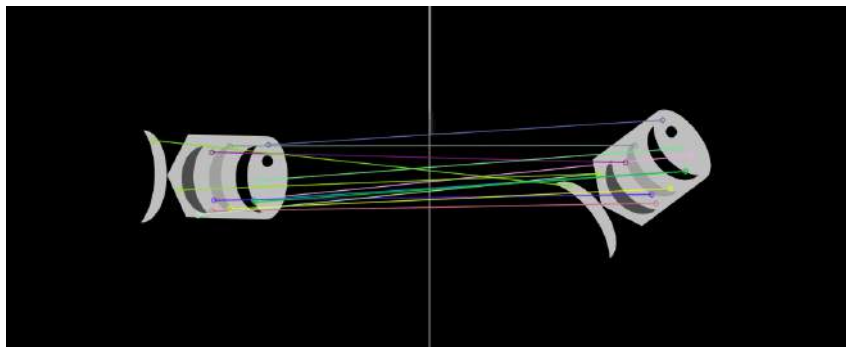


Figura 2.4: Exemplo de pontos de interesse detectados pelo SURF e utilizando FLANN (*Fast Library for Approximate Nearest Neighbors*) para comparação dos pontos de interesse.

tornando o ponto resistente a variações de rotação. Em seguida é aplicada uma aproximação da distribuição de respostas de Haar Wavelets gerando um vetor de 64 atributos para cada ponto de interesse. A seguir estão detalhadas as principais fases do SURF.

2.2.1 Detecção de Pontos de Interesse

Pontos de interesse no SURF são regiões onde há variações de valores nos pixels da imagem, e para detectar tais variações, é utilizada a matriz Hessiana juntamente com a função Gaussiana da Laplaciana. Os pontos podem ser encontrados em diferentes escalas, e o SURF utiliza um conceito denominado oitavas para tal fim.

Para entender melhor as variações nos valores dos pixels, a Figura 2.5 ilustra uma representação de uma imagem em escalas de cinza no espaço 3D da Figura 2.3. Os eixos x e y representam as coordenadas dos pixels da imagem, já o eixo z , representa o valor do pixel na escala de cinza, isto é, o valor $z = 0$ representa a cor preta, e o valor $z = 255$ a cor branca.

Observando a representação, nota-se que há maior variação nos valores dos pixels que correspondem as bordas do peixe e as limitações de suas listras. Assim, analisando os valores,

é possível determinar bordas as quais podem delimitar o peixe do fundo da imagem, e também, as listras que o compõem. O SURF busca detectar e determinar as direções das variações nas direções dos eixos x , y e diagonais xy e yx . Note que os pontos de interesse ilustrados na Figura 2.4 estão localizados em alguma borda onde as variações ocorrem com maior intensidade, e como não há variação nos valores dos pixels no fundo da imagem, não há pontos de interesse. Os pontos estão concentrados nas bordas que delimitam o peixe do fundo da imagem e nas listras que o compõem, representando algumas das regiões que melhor o caracterizam.

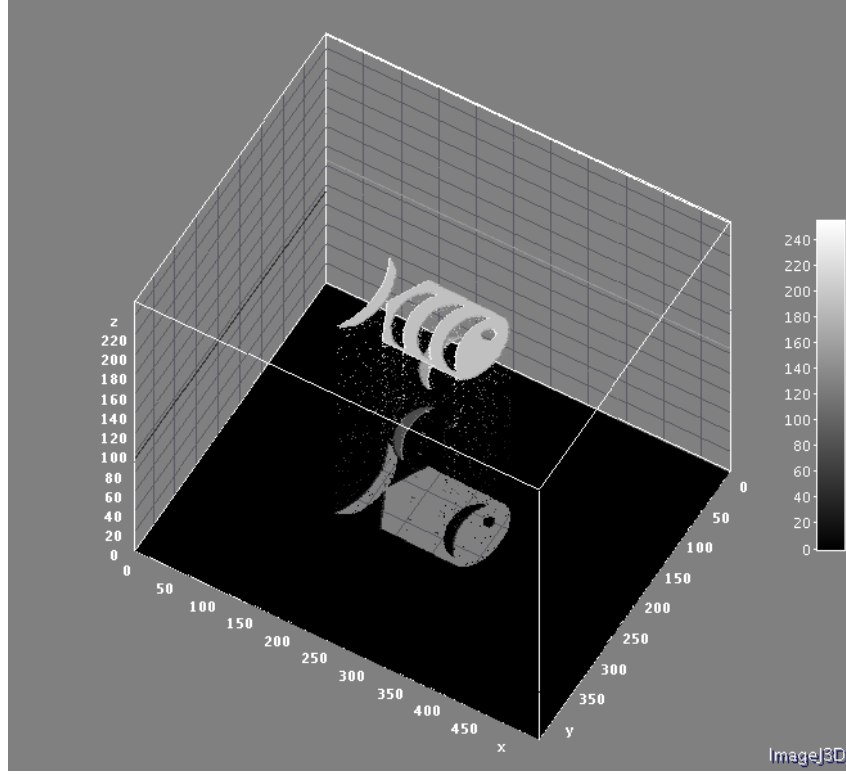


Figura 2.5: Representação da imagem ilustrada na Figura 2.3 como uma função de duas variáveis.

Dada uma representação de uma imagem em tons de cinza utilizando três variáveis (Figura 2.5), o SURF a percorre utilizando como base a matriz Hessiana, que por sua vez, é composta por derivadas de segunda ordem utilizadas para detectarem valores de máximos e mínimos locais. Cada posição da matriz (Equação 2.1) corresponde a variação de valores dos pixels em uma determinada orientação, sendo obtida por meio de convolução. A convolução é uma aproximação das derivadas de segunda ordem da Gaussiana responsáveis pela detecção da variação. A posição $\mathcal{H}_{1,1}$ ($\mathcal{L}_{xx}$) corresponde as variações dos valores na direção do eixo x e a posição $\mathcal{H}_{1,2}$ ($\mathcal{L}_{xy}$) na diagonal, isto é, variando nos eixos x e y simultaneamente. A posição $\mathcal{H}_{2,1}$ ($\mathcal{L}_{yx}$) é análoga a posição $\mathcal{H}_{1,2}$. E por fim, a posição $\mathcal{H}_{2,2}$ ($\mathcal{L}_{yy}$) corresponde a variações no eixo y . Como os pontos podem ser encontrados em diferentes escalas, a matriz Hessiana também é aplicada na imagens em diferentes escalas. Assim, dado um ponto $P = (x, y)$ em uma imagem I , a matriz Hessiana $\mathcal{H}(P, \sigma)$ em P , na escala σ , é definida na Equação 2.1:

$$\mathcal{H}(P, \sigma) = \begin{bmatrix} \mathcal{L}_{xx}(P, \sigma) & \mathcal{L}_{xy}(P, \sigma) \\ \mathcal{L}_{xy}(P, \sigma) & \mathcal{L}_{yy}(P, \sigma) \end{bmatrix} \quad (2.1)$$

Onde $\mathcal{L}_{xx}(P, \sigma)$ é a convolução da derivada de segunda ordem da Gaussiana $\frac{\partial^2}{\partial x^2} g(\sigma)$ com a imagem I no ponto P , e similarmente para $\mathcal{L}_{xy}(P, \sigma)$ e $\mathcal{L}_{yy}(P, \sigma)$.

O uso da Gaussiana mostrou-se eficaz para análise de escala e espaço segundo Koenderink

[18]. Contudo, segundo H.Bay et al. [10], na prática é necessário discretizá-la e recortá-la diminuindo o desempenho do algoritmo. Sendo assim, foi proposto a aproximação às derivadas de segunda ordem utilizando convolução com filtros caixa e imagens integrais. Ambas obtiveram performance semelhantes (H.Bay et al. [10]), e o uso das aproximações tornam a fase de detecção mais rápida. A subseção seguinte descreve a convolução aplicada ao SURF.

2.2.2 Detecção de Pontos de Interesse Utilizando Convolução

Uma definição simples de convolução pode ser uma operação entre duas funções f e g , onde $f * g$ ($*$ significa convolução) produz uma terceira função k que pode ser interpretada como uma modificação de f (filtro). A operação de convolução pode ser comparada à uma multiplicação de funções, $f \cdot g = k$.

Contudo, devido as imagens serem formadas por matrizes de pixels, uma melhor abordagem é utilizar filtros que também são representados por matrizes denominados núcleos da convolução. A Figura 2.6, por exemplo, ilustra um núcleo utilizado para verificar variações nos valores dos pixels no eixo x da imagem. Várias operações sobre imagens podem ser feitas utilizando filtros, tais como: suavização, detecção de bordas, deixar as cores da imagem mais vivas, entre outros.

Um núcleo de uma convolução é definido por uma matriz, e para cada posição é atribuído um valor sendo uma espécie de peso para o pixel. Nos exemplos da Figura 2.6, os valores variam entre 1 (posições coloridas com a cor branca) e -2 (posições em preto). As posições em cinza escuro correspondem a valores iguais à 0. O núcleo percorre a imagem pareando cada posição da matriz com um pixel, retornando uma resposta da convolução correspondendo a multiplicação do valor da posição da matriz com o valor do pixel. Como o SURF utiliza apenas imagens em escalas de cinza para detectar os pontos de interesse, os valores dos pixels da imagem variam entre 0 e 255.

No SURF, o modo como os valores são arranjados no núcleo da convolução depende de qual direção deseja-se verificar a variação. O filtro aplicado no exemplo ilustrado na Figura 2.6 arranja os valores de modo que a resposta da convolução implica se há variação na intensidade do pixel na direção do eixo x . Assim, dada uma sub imagem com as mesmas dimensões do núcleo da convolução (Figura 2.6) é feito o produto dos valores do núcleo com os pixels da imagem. Posteriormente, os valores são somados e podem ocorrer dois casos: a soma dos valores é igual a 0, então não há variação no eixo em questão, isto é, todos os pixels cobertos pelo núcleo possuem o mesmo valor (intensidade). Caso contrário, o valor será diferente de zero, indicando que há variação no eixo que corresponde ao núcleo. Fazendo o módulo da resposta da convolução, quanto maior o valor resultante da operação, mais acentuada é a variação.

O SURF utiliza os núcleos de convolução ilustrados nas Figuras 2.7.1 e 2.7.2 denominados filtros caixa. Portanto, cada filtro caixa corresponde a uma aproximação das derivadas parciais da Gaussiana ($\mathcal{L}_{xx}$, $\mathcal{L}_{xy}$, $\mathcal{L}_{yy}$), que por sua vez, compõem as posições da matriz Hessiana. Por fim, o determinante da mesma indica se há variação dos valores dos pixels em uma determinada região da imagem, indicando um ponto de interesse.

Uma aproximação do determinante da matriz Hessiana utilizada no SURF é obtida por meio da Equação 2.2. Onde D_{xx} , D_{yy} e D_{xy} são as derivadas parciais da matriz Hessiana (obtidas na convolução), e w , uma constante utilizada para compensar as diferentes escalas em que a matriz Hessiana é aplicada.

$$\det(\mathcal{H}_{aproximado}) = D_{xx}D_{yy} - (wD_{xy})^2 \quad (2.2)$$

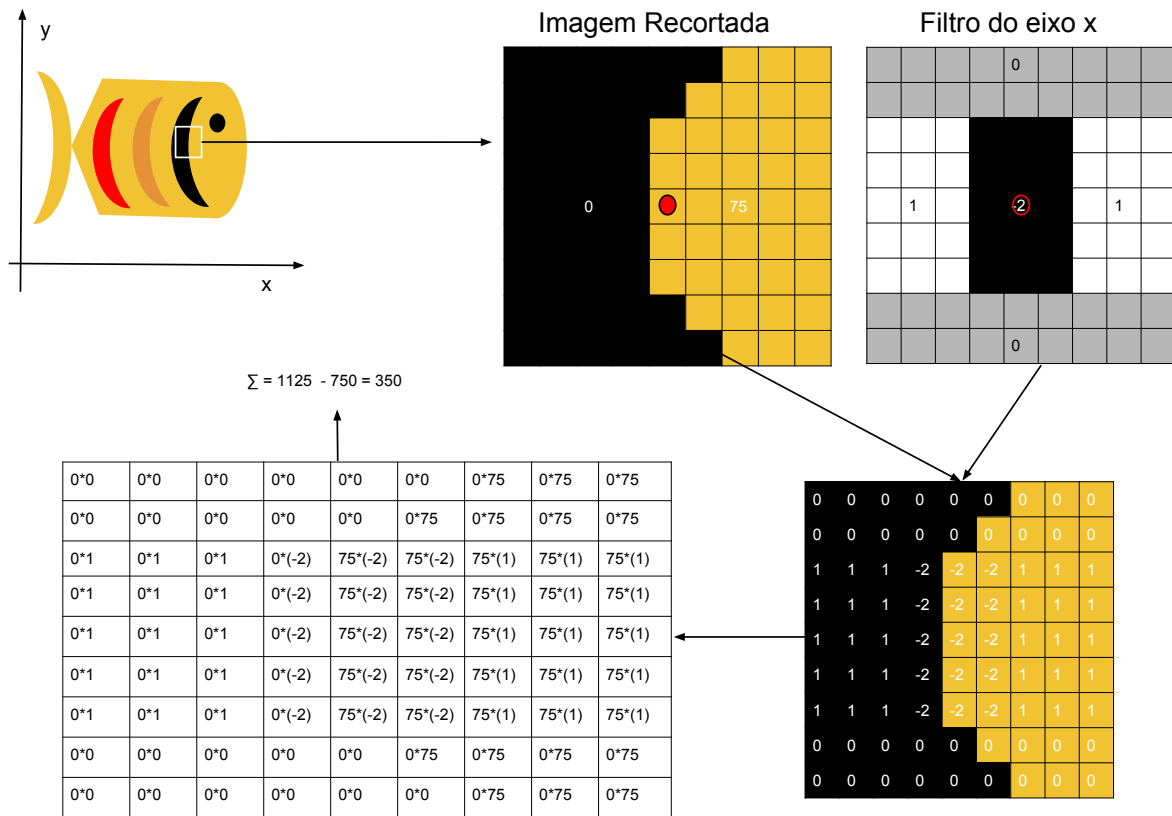


Figura 2.6: Exemplo da aplicação de um filtro caixa utilizado para detectar variação nos valores dos pixels no eixo x em relação ao pixel central da sub imagem (pixel com um círculo vermelho). Para fins ilustrativos a imagem está colorida, porém, no SURF, o processo é feito utilizando a imagem em tons de cinza.

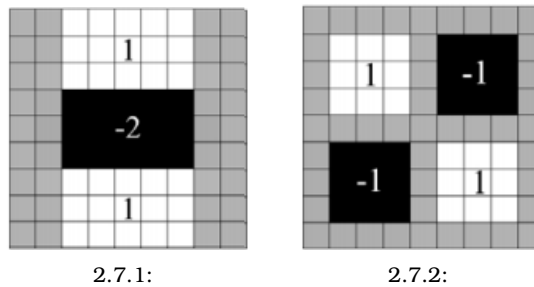


Figura 2.7: Núcleos de Convolução do SURF.

Mas como um mesmo ponto de interesse pode estar em escalas diferentes (o mesmo ponto pode possuir vários tamanhos em imagens distintas), o SURF utiliza oitavas nos núcleos da convolução juntamente com outra técnica denominada supressão de não máximo 3D.

2.2.3 Detecção de Pontos de Interesse em Diferentes Escalas

As regiões de maior relevância em uma imagem, no SURF, podem variar sua escala, isto é, a mesma região pode ser descrita por quantidades de pixels diferentes. A Figura 2.8 ilustra um exemplo em que, se o fator da escala não for levado em consideração, o mesmo ponto de interesse pode não ser identificado em uma mesma imagem com tamanhos diferentes. Este problema pode ser resolvido deixando ambas na mesma escala e aplicando os núcleos de

convolução.

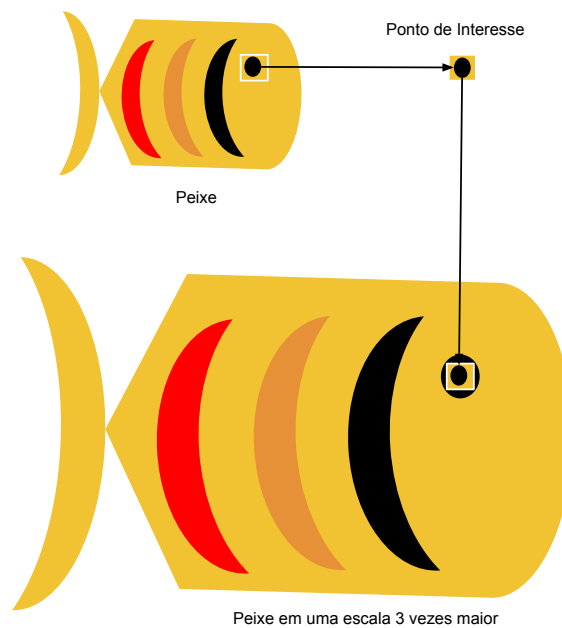
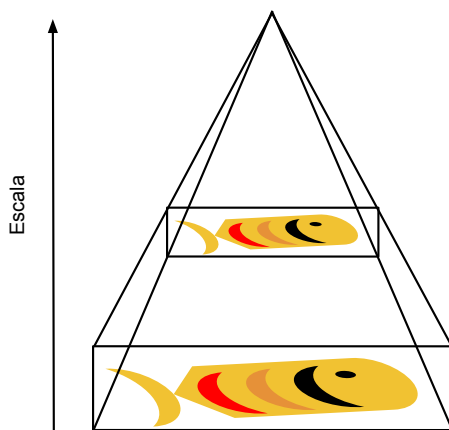
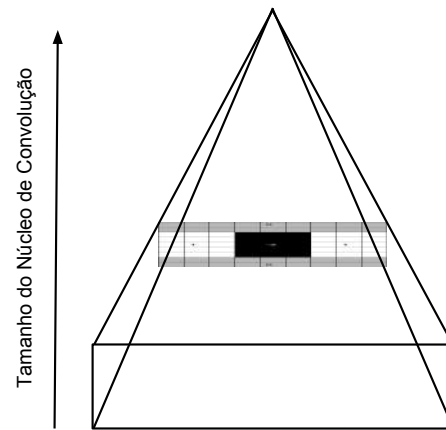


Figura 2.8: Figura ilustrando a variação de escala de um ponto de interesse. Caso o fator de escala não seja levando em consideração, o mesmo ponto não será encontrado em uma imagem maior (com uma resolução maior).

Para analisar a imagem em diferentes tamanhos é montada uma pirâmide de escalas (Figura 2.9.1), que, utilizando interpolação, escala a imagem tornando-a maior ou menor preservando ao máximo suas informações. Assim, a partir da imagem original, é formado um conjunto de imagens em que são aplicadas as convoluções para detectar os pontos de interesse.



2.9.1: Mudança de escala da imagem.



2.9.2: Mudança de escala dos núcleos de convolução.

Figura 2.9: Pirâmides de escala.

Como o SURF busca eficiência em seu processamento, em vez de variar o tamanho da imagem, ele varia a dimensão do núcleo de convolução (Figura 2.9.2). Assim, o mesmo núcleo é aplicado com tamanhos diferentes simulando as variações de escalas. Mas para grandes núcleos e uma grande variação de escalas, o tempo de processamento nos cálculos da con-

volução podem ser longos. Sendo assim, o SURF utiliza imagens integrais para efetuar as somas dos valores dos pixels e multiplicá-los pelas posições correspondente no núcleo de convolução.

Imagens integrais é um método em que todos os valores dos pixels de uma imagem (como os obtidos da representação de uma imagem em tons de cinza como uma função de três variáveis) são somados formando uma matriz, em que, cada posição contém a soma de todos os valores das posições anteriores, isto é, dada uma matriz M , a posição $M_{2,2}$ contém a soma dos valores das posições correspondente aos pixels: $I_{1,1}$, $I_{1,2}$, $I_{2,1}$ e $I_{2,2}$ de uma imagem I . Logo, o valor de uma posição $M_{x,y}$ é dado pela Equação 2.3.

$$M_{x,y} = \sum_{i=0, j=0}^{i < x, j < y} I_{i,j} \quad (2.3)$$

Onde M é a matriz integral, e I os valores dos pixels da imagem. A Figura 2.10 ilustra um exemplo do uso de imagens integrais, onde deseja-se somar os valores dos pixels que formam o quadrado de vértices A, B, C e D. Assim, para calcular a soma dos valores da região coberta pelo quadrado, é necessário aplicar apenas 3 operações básicas. A soma dos valores é dada pela Equação 2.4.

$$s = A - B - C + D \quad (2.4)$$

Note que, se um método iterativo fosse usado, seria necessário percorrer todas as posições cobertas pelo quadrado quantas vezes fosse necessário, tornando os cálculos da convolução custosos. Como há uma grande variação no número de escalas e tamanhos de núcleos da convolução, o uso de imagens integrais mostra-se eficiente, pois a soma dos valores dos pixels de uma imagem é feita apenas uma vez, assim como, a soma de qualquer região é obtida de forma simples e rápida.

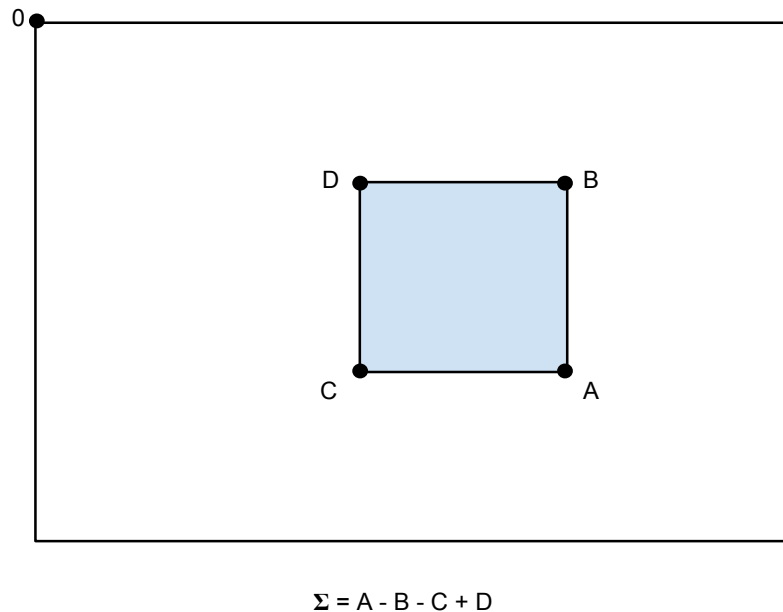


Figura 2.10: Exemplo do uso de imagens integrais.

O tamanho dos núcleos de convolução variam de acordo com o conceito de oitavas, onde ocorrem combinações de progressões geométricas, exponenciais e aritméticas. Geralmente, o tamanho das escalas variam exponencialmente, contudo, a diferença entre elas crescem

com grandes intervalos, podendo haver perda de informações entre as transições. Visando diminuir estas diferenças, o conceito de oitavas foi aplicado. A Figura 2.11 ilustra a diferença entre as escalas e oitavas. Note que na primeira oitava os tamanhos dos núcleos de convolução variam de 9 à 27, sendo variações suaves quando comparado ao crescimento exponencial da escala. O menor filtro no SURF possui tamanho 9×9 e os núcleos, seguindo as oitavas, variam nos tamanhos de acordo com os valores ilustrados na Figura 2.11.

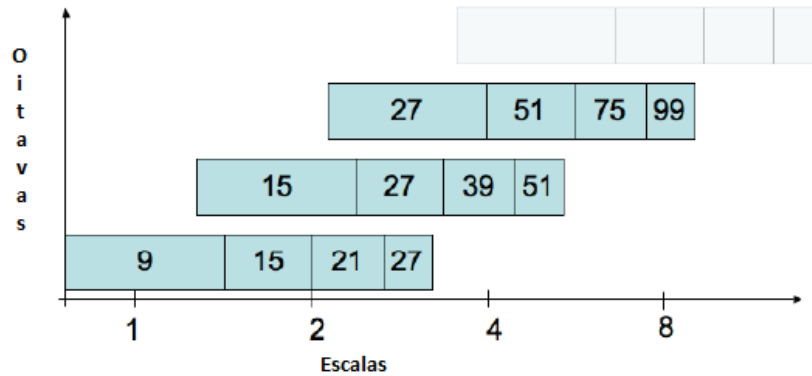


Figura 2.11: Oitavas variando exponencialmente em relação ao eixo y e aritmeticamente no eixo x(6 em 6, 12 em 12, 24 em 24).

Após a aplicação da convolução na imagem com diferentes oitavas, as respostas são comparadas para encontrar a escala com maior variação de valores. É utilizada então, uma técnica denominada supressão de não máximos 3D a qual compara as respostas das convoluções entre escalas vizinhas, retornando a com maior variação.

Utilizando os conceitos descritos, o SURF é capaz de encontrar os pontos de interesse e as escalas em que eles se encontram bastando apenas descrevê-los. A Figura 2.12 ilustra os pontos encontrados em uma imagem. Os círculos maiores indicam que foram encontradas variações em uma escala maior do que os círculos menores.

2.2.4 Descrição dos Pontos de Interesse

Na fase de descrição o SURF busca tornar os pontos de interesse resistentes a variações de rotações, isto é, mesmo sofrendo rotação o ponto é detectado. Este processo é feito utilizando convolução e encontra a direção em que a variação dos valores dos pixels ocorrem. Os pontos são então descritos com base em sua orientação principal por meio de um vetor de 64 posições.

Inicialmente, dada a posição de um ponto de interesse (coordenadas na imagem), uma área em volta do ponto é definida para a convolução ser aplicada utilizando dois núcleos: um verifica a variação dos valores no eixo x e outro na direção do eixo y em relação a imagem. A Figura 2.13 ilustra os núcleos utilizados. O tamanho da área que envolve a posição do ponto de interesse é proporcional a escala em que ele se encontra. Para cada pixel pertencente a região, é obtida a resposta da convolução do eixo x e do eixo y , assim, estes dois valores são utilizados de modo que representam pontos no círculo trigonométrico (Figura 2.14). A circunferência é então dividida em setores, e o setor com maior número de pontos é escolhido para definir a direção que apresenta maior variação. A Figura 2.14 ilustra um exemplo em que a circunferência foi dividida em 6 setores, e o setor que contém a maior quantidade de pontos é o $\frac{\pi}{3}$. Sendo assim, a direção onde há maior variação será a indicado pela seta vermelha, correspondendo ao vetor formado pelos pontos $(0,0)$ e $(\frac{\sqrt{3}}{2}, \frac{\pi}{6})$.

Definida a orientação principal dos pontos de interesse, são formadas novas janelas ao redor dos pontos, as quais possuem 20 vezes o tamanho da escala em que o ponto foi encon-

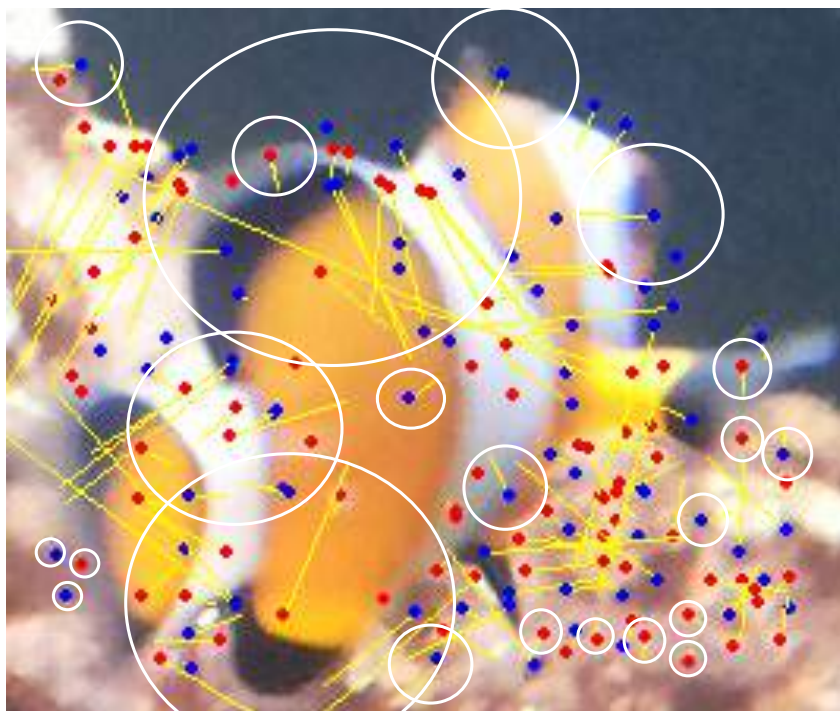


Figura 2.12: Regiões de maior variância em diferentes escalas. As linhas amarelas indicam a direção da variação. O tamanhos dos círculos estão relacionados às escalas em que os pontos foram encontrados.

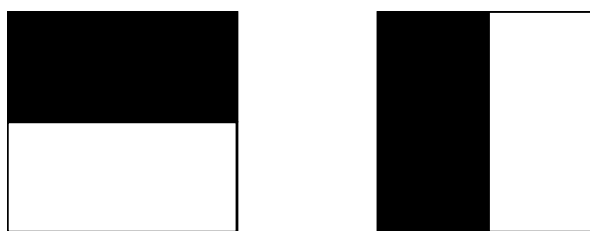


Figura 2.13: Núcles de convolução utilizados na detecção de variância de rotação.

trado. Em seguida, a janela é dividida 64 partes onde o agrupamento de 4 partes forma uma sub janela. A convolução utilizando os núcleos ilustrados na Figura 2.13 é novamente aplicada em todas 64 partes, mas desta vez, os eixos não serão mais alinhados com os eixos da imagem, mas sim em relação a orientação principal de cada ponto de interesse (em relação a direção da seta vermelha da Figura 2.14), definindo uma direção de variação para cada parte da janela. A Figura 2.15 ilustra um exemplo da orientação de um ponto e a divisão das sub janelas. É possível observar que as orientações de todas as partes da janela tendem a mesma orientação principal do ponto.

A partir de cada sub janela são obtidos 4 valores formando um total de 64 valores que correspondem a cada posição do vetor de descrição. Os valores de cada sub janela correspondem à: 1) soma dos valores variantes na direção do eixo x , 2) módulo da soma dos valores variantes na direção do eixo x , 3) soma dos valores variantes na direção do eixo y , 4) módulo da soma dos valores variantes na direção do eixo y .

Devido ao SURF descrever cada ponto de interesse utilizando um vetor de 64 posições, é possível comparar dois pontos utilizando seus vetores, e portanto, comparando os pontos de duas imagens é possível chegar a conclusão se elas são semelhantes. Mas cada imagem pode possuir uma quantidade diferentes de pontos, e a comparação deixa de ser ponto a ponto,

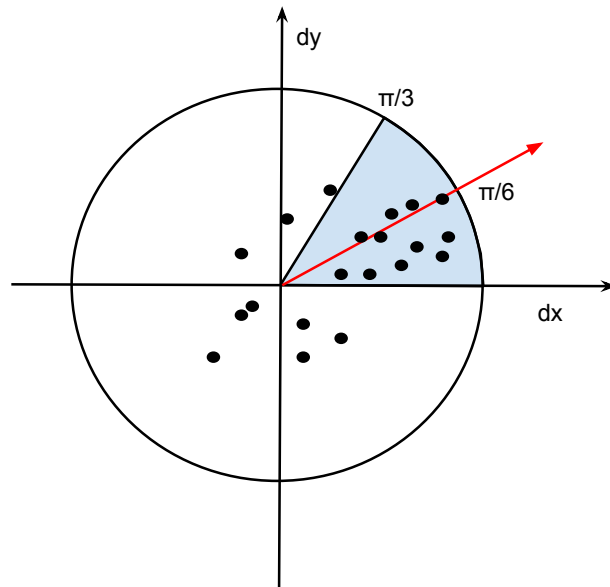


Figura 2.14: Gráfico ilustrando a variação dos valores dos pixels de uma região em volta de um ponto de interesse. O centro da circunferência é posicionado na coordenada do ponto de interesse.

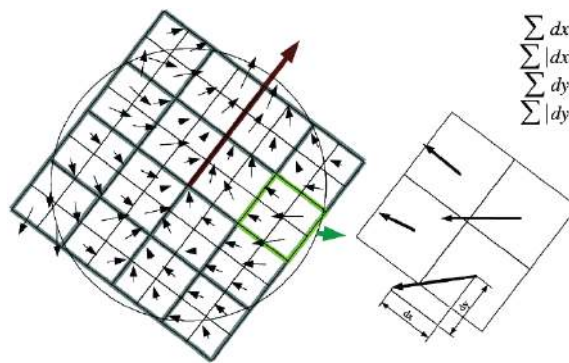


Figura 2.15: Distribuição de Haar Wavelets.

pois alguns pontos serão ignorados. Assim, técnicas como o BoVW e BoFC tornam possíveis as comparações de imagens utilizando pontos de interesse.

2.3 Histograma de Palavras e Histograma de Palavras Visuais

Uma das técnicas mais utilizadas na classificação de texto é o Histograma de Palavras (*Bag of Words* - BoW) em que são contadas as ocorrências das palavras de um texto, e entre outras finalidades, determinando qual assunto o mesmo trata. Assim, um texto de esportes, por exemplo, poderá ter uma grande quantidade de ocorrências das palavras jogo, pontos, sets, falta, entre outras. O Histograma de Palavras Visuais (*Bag of Words*), seguindo o mesmo conceito do Histograma de Palavras, utiliza pontos de interesse para formar palavras visuais. Para cada palavra visual encontrada na imagem, é obtida sua ocorrência e o conjunto das ocorrências de todas as palavras visuais descreve a imagem.

No BoVW o conjunto de palavras possíveis de se encontrar em um determinado problema, assim como no *Bag of Words*, é denominado dicionário. Contudo, no *Bag of Words* a quantidade de palavras é bem definida e determinada pela linguagem do texto. Já no BoVW, as palavras são formadas aplicando algoritmos de agrupamento nos pontos de interesse, e a sua

quantidade pode ser tão grande quanto se queira.

Para descrever as imagens, o BoVW utiliza pontos de interesse encontrados pelo SURF [10] ou SIFT [9]. Neste trabalho o SURF foi utilizado por ser um algoritmo que necessita de menor tempo e processamento quando comparado ao SIFT. Uma característica necessária para dispositivos móveis pois possuem um poder computacional limitado.

Inicialmente, para todo o banco de imagens, o SURF é aplicado e os pontos de interesse de todas as imagens são detectados e descritos. Posteriormente o algoritmo de agrupamento K-Médias forma K centroides ou grupos de pontos de interesse. Cada centroide corresponde a uma palavra visual e pertence ao dicionário do BoVW. Um histograma é então criado a partir do dicionário e cada posição corresponde a um centroide, isto é, cada posição do histograma corresponde a uma palavra do dicionário. Para haver um incremento na ocorrência de uma palavra visual, basta haver na imagem, um ponto de interesse que pertença ao centroide correspondente a palavra visual.

A Figura 2.16 ilustra um exemplo da aplicação do BoVW em um conjunto de imagens. A Figura 2.16.1 ilustra a aplicação do SURF em três imagens e a representação dos pontos no espaço. O algoritmo de agrupamento K-Médias foi aplicado com quatro grupos formando um dicionário de quatro palavras visuais, p1, p2, p3 e p4. Posteriormente, as imagens são descritas pela quantidade de pontos pertencentes a cada grupo, assim, a primeira imagem de peixe da Figura 2.16.1, por exemplo, será descrita por um vetor contendo os valores dois, um, zero e zero correspondendo as palavras visuais p1, p2, p3 e p4, respectivamente (Figura 2.16.2). As imagens descritas podem então ser treinadas por um classificador e, dada uma imagem de teste, como a ilustrada na Figura 2.16.3, os pontos são novamente identificados. Usando o mesmo dicionário do treinamento, é feito um histograma da imagem de teste (Figura 2.16.4), que por sua vez, será classificada pelo classificador.

Como o BoVW baseia-se apenas no extrator de atributos SURF, e o SURF manipula apenas imagens em tons de cinza, as informações de cores não são utilizadas. Para enriquecer as informações dos pontos de interesse, técnicas como o Histograma de Atributos e Cores foram analisadas.

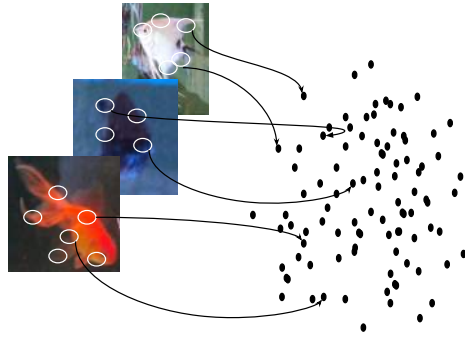
2.4 Histograma de Atributos e Cores

O Histograma de Atributos e Cores (*Bag of Features and Colors* - BoFC) [7] é uma técnica baseada no BoVW que incrementa a descrição das imagens adicionando informações de cor aos pontos de interesse encontrados pelo SURF.

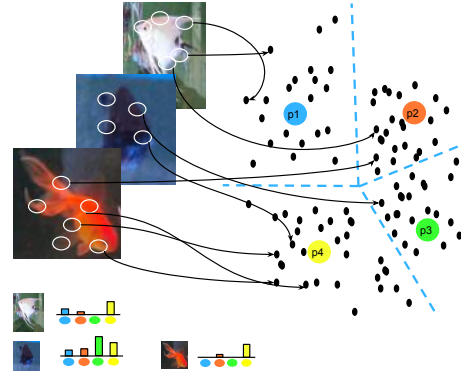
Inicialmente, dadas as imagens de treinamento, os mesmos processos aplicados no BoVW são aplicados no BoFC para encontrar e descrever os pontos de interesse, assim como a formação das palavras visuais. Contudo, o BoFC adiciona informação de cor nos vetores que descrevem cada ponto. Note que a cor possui uma influência local na descrição da imagem, limitada a cada ponto de interesse.

As informações de cores são adicionadas aos pontos de interesse da seguinte forma: para cada ponto de interesse é calculada a média e variância de cada canal de cor em uma área de 5 x 5 em volta do centro do ponto. Assim, se uma imagem possuir três canais de cores, é calculada a média e variância de cada canal de cor aumentando o vetor que descreve o ponto em seis posições.

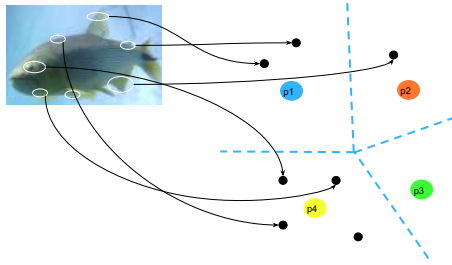
Neste trabalho, foram utilizados os espaços de cores RGB com o BoFC. Portanto o vetor de descrição dos pontos de interesse passam a possuir 70 posições, 64 posições providas do SURF mais 6 posições relativas as médias e variâncias dos canais de cores.



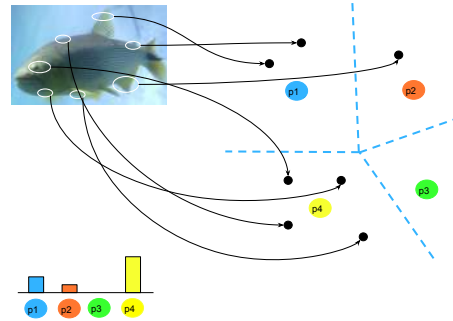
2.16.1: Pontos de interesse detectados pelo SURF e representados no espaço.



2.16.2: Aplicação do algoritmo de agrupamento K-Médias nos pontos de interesse detectados e a contagem das ocorrências dos pontos de interesse em cada centroide.



2.16.3: Aplicação do SURF em uma imagem de teste.



2.16.4: Criação do histograma de palavras visuais da imagem de teste.

Figura 2.16: Exemplo da aplicação do BoVW em um conjunto de imagens formando um dicionário de 4 palavras correspondendo aos centroides do K-Médias. Um histograma é então formado a partir da contagem dos pontos de interesse pertencentes a cada palavra visual.

As médias e variâncias de cada canal de cor são obtidas por meio das Equações 2.5 e 2.6.

$$\bar{E}_i = \frac{1}{N} \sum_{j=1}^N P_{ij} \quad (2.5)$$

$$\alpha_i = \left[\frac{1}{N} \sum_{j=1}^N (P_{ij} - \bar{E}_i)^2 \right]^{1/2} \quad (2.6)$$

Onde i é o valor do i -ésimo canal de cor, e j é o j -ésimo pixel da imagem. Assim é denotado por P_{ij} o pixel j do canal de cor i , N é o número de pixels na imagem, no caso 25 (5x5).

2.5 Histograma de Palavras Coloridas

Um técnica denominada histograma de palavras coloridas também foi utilizada neste trabalho. A base da mesma é na junção do BoVW com o histograma de cores HSV, e será descrita na seção seguinte.

O Histograma de Palavras coloridas (*Bag of Colored Words* - BoCW) é uma técnica derivada do BoVW e do histograma de cores HSV a qual une as características de ambos. Basicamente, o vetor de atributos que descreve cada imagem gerado pelo BoVW é concatenado com um histograma de cores HSV da imagem.

Cada imagem de treinamento é descrita utilizando o SURF juntamente com o BoVW, posteriormente, as informações de cores são obtidas utilizando um histograma de cores no espaço de cores HSV e o vetor resultante é concatenado com o descritor do BoVW das imagens. O tamanho do vetor corresponde a quantidade de palavras no dicionário do BoVW somado ao número de faixas do histograma ao cubo.

2.6 Histograma de Cores de Wengert

As técnicas apresentadas até o momento possuem como base da descrições de imagens o SURF e o histograma de palavras. O BoC (Histograma de Cores de Wengert) também utiliza histograma de palavras visuais, mas as mesmas são formadas apenas por cores e não pontos de interesse.

O Histograma de Cores de Wengert (*Bag of Colors* - BoC)[8] é uma técnica baseada no histograma de palavras e utiliza apenas informações de cores das imagens. Cada imagem é descrita por um histograma obtido a partir das cores mais comuns do conjunto das imagens de treinamento. O agrupamento das cores formam um dicionário utilizado para gerar assinaturas de cores para cada imagem.

Segundo Christian et al.[8], o espaço de cores que melhor representa as imagens no BoC é o CIE-Lab pois apresenta maior consistência em relação a distância Euclidiana, em particular, na criação do dicionário de cores. Portanto, para o dicionário de cores ser formado, todas as imagens são inseridas no espaço CIE-Lab.

2.6.1 Dicionário de Cores

O dicionário de cores do BoC é formado por um conjunto de k_c cores criando um dicionário $C = \{c_1, \dots, c_{k_c}\}$, também referenciado como palheta de cores. Para reduzir a gama de cores possíveis e ganhar eficiência no tempo de processamento, cada componente L, a e b do espaço de cores é separado em faixas, tornando $k_c = 64$ para quatro bins ($4 \times 4 \times 4$), $k_c = 512$ para oito bins ($8 \times 8 \times 8$), até a quantidade que se queira.

Para criar um dicionário adaptável para imagens reais, dado um conjunto de treinamento constituído de N imagens, os passos a seguir são executados para diminuir o impacto de imagens que contenham grandes áreas de cores uniformes.

1. Cada imagem é redimensionada para as dimensões 256×256 e convertidas para o espaço de cor CIE-Lab. A imagem é então dividida em 256 blocos de 16×16 .
2. Para cada bloco, a cor de maior ocorrência é encontrada. Empates são resolvidos aleatoriamente. Caso uma cor não corresponda a no mínimo 5 ocorrências, uma cor aleatória é escolhida.
3. Nesta fase há 256 cores das N imagens de treinamento, ou seja, $256 \times N$ cores ao total. O algoritmo K-Médias é então aplicado formando K cores.

Um exemplo de uma palheta de cores está ilustrados na Figura 2.17. Com o dicionário formado, é possível então produzir assinaturas de cores para cada imagem.

2.6.2 Assinatura de Cores

A assinatura de cores consiste em criar um histograma de cores de cada imagem baseado no dicionário criado C . Este procedimento é feito redimensionando cada imagem à um tamanho de 128×128 , e para cada pixel p seleciona-se a cor mais próxima do dicionário C



Figura 2.17: Dicionário de cores criado a partir do BoC ilustrando as cores mais frequentes do banco de imagens . As palhetas correspondem a $k_c = 64$ respectivamente.

utilizando a distância Euclidiana. Encontrada a cor mais próxima, a posição do histograma correspondendo a cor é incrementada. Em seguida, normalizações são aplicadas para melhorar a descrição das imagens.

2.6.3 Normalizações

Uma das técnicas de normalização bastante usada no histograma de palavras é o inverso da frequência do documento (*Inverse Document Frequency* - *idf*) a qual penaliza palavras com grande ocorrência nos documentos. No caso das palavras visuais, penalizando palavras mais frequentes e aumenta o peso das raras. Os valores do histograma de cada imagem, correspondendo às palavras, são multiplicados pelos seus respectivos *idf*.

Também é aplicada a lei da potência (*Power Law*) que, similarmente a *idf*, penaliza as palavras mais frequentes mas sem levar em consideração a estatística das outras imagens do banco de imagens. A distribuição empírica das cores dá pouca importância a cores menos frequentes, sendo que, para pequenos objetos, a lei da potência regulariza a contribuição de cada cor no descritor final da imagem. Para um histograma X formado pelos valores $X = (x_1, x_2, \dots, x_d)$, a lei da potência consiste em atualizar cada posição com os valores $x_i = \sqrt{x_i}$ com $1 \leq i \leq d$.

A seção seguinte apresenta os materiais e métodos utilizados juntamente com as técnicas descritas nesta seção.

Materiais e Métodos

Nos experimentos, foram comparados os extratores de atributos BoVW, BoC, BoCW, BoFC, Histograma no espaço HSV e RGB utilizando o banco de imagens descrito na Seção 3.2. Como técnica de amostragem, foi adotada a validação cruzada de 10 partições, com 10 repetições. Para cada algoritmo testado, foram analisados os desempenhos com base na Medida-F e como complemento a AUC.

Para identificar se os extratores testados diferem estatisticamente em relação ao desempenho, foi utilizado o teste não-paramétrico proposto por Friedman [12] e disponível no software estatístico R na versão 3.0.2. Se um nível de significância de 95% (valor $p \leq 0.05$) for constatado, um teste *post-hoc* será realizado para analisar os diagramas de caixa resultantes. O teste *post-hoc*, também disponível no R, tem como base o teste de Wilcoxon com correção para FWER (*Family-wise Error Rate*) descrito por Hollander e Wolf Hollander [3].

A Seção 3.1 descreve os classificadores utilizados neste trabalho. A Seção 3.2 descreve o banco de imagens utilizado. Na Seção 3.3 são detalhadas as métricas utilizadas na comparação.

3.1 Classificadores

Para classificar as espécies de peixes descritas pelos extratores de atributos foram analisados três tipos de classificadores: Árvore de decisão, SVM e KNN. Cada classe, referente ao problema tratado neste trabalho, corresponde a uma espécie de peixe.

- Árvore de Decisão [2, 11] : a árvore de decisão é um algoritmo de aprendizado de máquina pertencente a classe dos classificadores que utilizam aprendizado supervisionado e não paramétrico. É utilizada em regressões e classificações. A sua classificação dá-se em criar um modelo e prever a classe de um exemplo utilizando regras de decisão inferidas do conjunto de treinamento.

A estrutura de uma árvore resume-se aos nós folhas em que as classes do problema são atribuídos, e também os nós de decisão que testam um determinado atributo a fim de definir a qual subárvore o mesmo pertencerá. Assim, para cada novo exemplo, um

elemento percorre os ramos da árvore (nós de decisão) passando pelos testes até chegar a um nó folha, sendo classificado com a classe atribuída ao nó.

Um fator importante neste classificador é o modo como os nós de decisão determinam para qual sub árvore o novo exemplo pertencerá. Neste trabalho foram alternadas entre duas técnicas para medir a qualidade da escolha, são elas: entropia para o ganho de informação e impureza Gini.

Algumas vantagens [22] podem ser encontradas no uso da árvore de decisão, tais como: 1) simplicidade no entendimento de sua estrutura, 2) é requerida pouca preparação dos dados. Algumas técnicas necessitam normalizar os dados, dividir atributos em várias variáveis ou removê-los caso possuam valor nulo. 3) O custo do uso da árvore na predição de um exemplo é logarítmico em relação ao número de atributos utilizados no conjunto de treinamento e 4) é possível ser utilizada em problemas com múltiplas classes.

Contudo, são desvantagens [22]: 1) algumas árvores podem ser tão complexas sendo incapaz de generalizar os dados. Este problema é denominado *overfitting* e pode ter seu impacto reduzido podando a árvore. 2) Elas podem ser instáveis já que pequenas variações nos dados podem gerar árvores completamente diferentes. 3) O problema de inferir a árvore ótima para classificação pertence a classe NP-Completo. 4) Quando o número de exemplos de cada classe do problema é desbalanceado, a árvore pode tornar-se tendenciosa. Assim, para melhores resultados é necessário o balanceamento das classes.

- Máquina de Vetores de Suporte (Support Vector Machine - SVM) [14]: criado por Vapnik [14], a Máquina de Vetores de Suporte é um conjunto de métodos supervisionados utilizados para classificação e regressão. Classificadores SVM's baseiam-se em uma margem máxima entre as classes para classificar novos exemplos, e maximizando a margem, melhora-se a generalização na fase de classificação. A separação das classes pode ser feita linearmente ou polinomialmente por meio de núcleos. Dependendo do hiper espaço formado, funções lineares dividem melhor as classes. Contudo, há casos onde núcleos polinomiais são mais eficientes.

Algumas vantagens [22] podem ser encontradas no uso de classificadores baseados em máquinas de vetores de suporte: 1) ele é bastante efetivo em espaços dimensionais grandes, isto é, quando a quantidade de atributos abordada pelo problema é grande, 2) mesmo em casos onde a quantidade de dimensões é maior que a de exemplos ele ainda continua efetivo. 3) Diferentes núcleos podem ser especificados como funções de decisão tornando o SVM um classificador versátil. Caso seja necessário é possível definir um núcleo customizável.

As desvantagens [22] no uso do SVM incluem: 1) se a quantidade de atributos é maior que o número de exemplos, o SVM é dado com um classificador pobre. 2) O SVM não provê diretamente estimativas e probabilidades, isto é feito utilizando a escala de Platt.

Inicialmente, o SVM foi criado para problemas binários, contudo, há extensões que utilizam o método um contra todos, ou um contra um como uma abordagem para múltiplas classes. Neste trabalho a implementação do SVM é dada pelo LibSVM [1].

- K-Vizinhos mais Próximos (*K-Nearest Neighbors* - KNN): o princípio do KNN está associado em encontrar um número pré definido de vizinhos do conjunto de treinamento de um novo exemplo e classificá-lo. A classe majoritária entre os vizinhos classifica o novo exemplo. O número de vizinhos pode ser fixo ou baseado na densidade local dos pontos (aprendizado baseado em um raio em volta do exemplo). Caso o número de vizinhos seja

fixo, ele deve ser ímpar para evitar empates. Para determinar o quão perto um exemplo está de seus vizinhos, uma função que define uma distância entre dois exemplos é aplicada. Uma das mais comuns é a distância Euclidiana.

O KNN possui como uma de suas principais vantagens [22] a simplicidade, devido a sua fase de aprendizado ser rápida, além de obter bons resultados em vários problemas. Alguns dos trabalhos citados no Capítulo 1 utilizam o KNN para classificar as espécies de peixes.

A principal desvantagem [22] do KNN ocorre quando a quantidade de atributos e de exemplos é grande. Como há a necessidade do cálculo da distância de um novo exemplo com todos os outros exemplos do conjunto de treinamento para encontrar os K vizinhos mais próximos, estes cálculos podem ser demorados. Mas este problema pode ser amenizado utilizando algumas técnicas que limitam as buscas.

As imagens utilizadas na classificação são detalhadas a seguir.

3.2 *Banco de Imagens*

O Aquário do Pantanal ainda está na fase de construção e não é possível obter as imagens de seus tanques, sendo assim, os classificadores foram treinados com imagens de peixes ornamentais. O produto final será treinado com fotos obtidas do Aquário do Pantanal.

Devido à necessidade de simular o ambiente onde o aplicativo será executado, as imagens foram obtidas de vários aquários com diferentes iluminações. Assim, algumas imagens foram obtidas no Aquário Municipal de Toledo Paraná ¹ e outras em casas de venda de peixes ornamentais.

No Aquário Municipal de Toledo foram utilizados três celulares, e as fotos foram tiradas com os peixes nadando livremente a uma distância aproximada de 1 metro com iluminação natural (Figura 3.2). Foram obtidas 40 fotos de cada uma das 10 espécies de peixes. Os nomes das espécies foram obtidos por meio de placas nos aquários.

Algumas imagens de espécies de peixes foram obtidas de aquários da loja Planeta Real² em Campo Grande, Mato Grosso do Sul. As imagens foram obtidas por meio de um celular a uma distância de um metro dos aquários com iluminação natural. O restante das imagens (Figura 3.1) foram obtidas da loja Peixinho Dourado ³.

Como grande parte das imagens foi obtida com um elevado número de peixes, cada imagem foi analisada e foram feitos recortes em volta dos peixes restando apenas um peixe por imagem. Este recorte também é feito na aplicação final e maiores informações estão apresentadas na Seção 6. A Figura 3.3 ilustra uma foto de cada espécie que compõem o banco de imagens com seu nome popular e científico. As Figuras 3.4, 3.5 e 3.6 ilustram um conjunto de 6 imagens das 40 de cada espécie representando uma amostra do banco de imagens.

¹<http://www.toledo.pr.gov.br/portal/meio-ambiente/aquario-municipal-romolo-martinelli>.

²Avenida Afonso Pena, 1919 - Centro, Campo Grande - Mato Grosso Do Sul. Telefone: (67) 3025-4942.

³Avenida Marechal Rondon ,1338, Centro, Rondonopolis, Mato Grosso



3.1.1:



3.1.2:



3.1.3:



3.1.4:



3.1.5:

Figura 3.1: Exemplos de imagens do ambiente de captura das casas de peixes ornamentais de Campo Grande, Mato Grosso do Sul.



Figura 3.2: Imagem do Aquário Municipal de Toledo, Paraná.

Imagem	Nome Popular	Nome Científico
	Dourado	<i>Salminus brasiliensis</i> (Cuvier, 1816)
	Acará Bandeira	<i>Pterophyllum scalare</i> (Schultze, 1823)
	Tetra Negro	<i>Gymnocorymbus ternetzi</i> (Boulenger, 1895)
	Platy Ruby	<i>Xiphophorus maculatus</i> (Günther, 1866)
	Telescópio	<i>Carassius auratus</i> (Linnaeus, 1758)
	Paulistinha	<i>Danio rerio</i> (Hamilton, 1822)
	Peixe Papagaio	<i>Sparisoma chrysopteron</i> (Bloch e Schneider, 1801)
	Oscar Albino	<i>Astronotus ocellatus</i> (Agassiz, 1831)
	Mato Grosso	<i>Hyphessobrycon eques</i> (Steindachner, 1882)
	Platy Laranja	<i>Xiphophorus maculatus</i> (Günther, 1866)
	Peixe Palhaço	<i>Amphiprion frenatus</i> (Brevoort, 1856)
	Kinguio Karraco	<i>Carassius auratus</i> (Linnaeus, 1758)
	Kinguio Cometa Calico	<i>Carassius auratus</i> (Linnaeus, 1758)
	Barbus Sumatra	<i>Puntigrus tetrazona</i> (Bleeker, 1855)
	Barbus Ouro	<i>Puntius sachsii</i> (Ahl, 1923)
	Acará Disco	<i>Symphysodon aequifasciatus</i> (Pellegrin, 1904)
	Oscar	<i>Astronotus ocellatus</i> (Agassiz, 1831)
	Tricogaster	<i>Trichogaster trichopterus</i> (Pallas, 1770)
	Kinguio	<i>Carassius auratus</i> (Linnaeus, 1758)
	Platy Sangue	<i>Xiphophorus maculatus</i> (Günther, 1866)
	Molinésia Preta	<i>Poecilia shenops</i> (Valenciennes, 1846)
	Carpa	<i>Cyprinus carpio</i> (Linnaeus, 1758)
	Beta	<i>Betta splendens</i> (Regan, 1910)
	Tucunaré	<i>Cichla ocellaris</i> (Bloch e Schneider, 1801)
	Piau Três Pintas	<i>Leporinus friderici</i> (Bloch, 1794)
	Pacu	<i>Piaractus mesopotamicus</i> (Holmberg, 1887)
	Acará Bandeira Marmorizado	<i>Pterophyllum scalare</i> (Schultze, 1823)
	Carpa Média	<i>Cyprinus carpio</i> (Linnaeus, 1758)

Figura 3.3: Ilustração de todas as espécies de peixes com seus nome populares e nomes científicos.

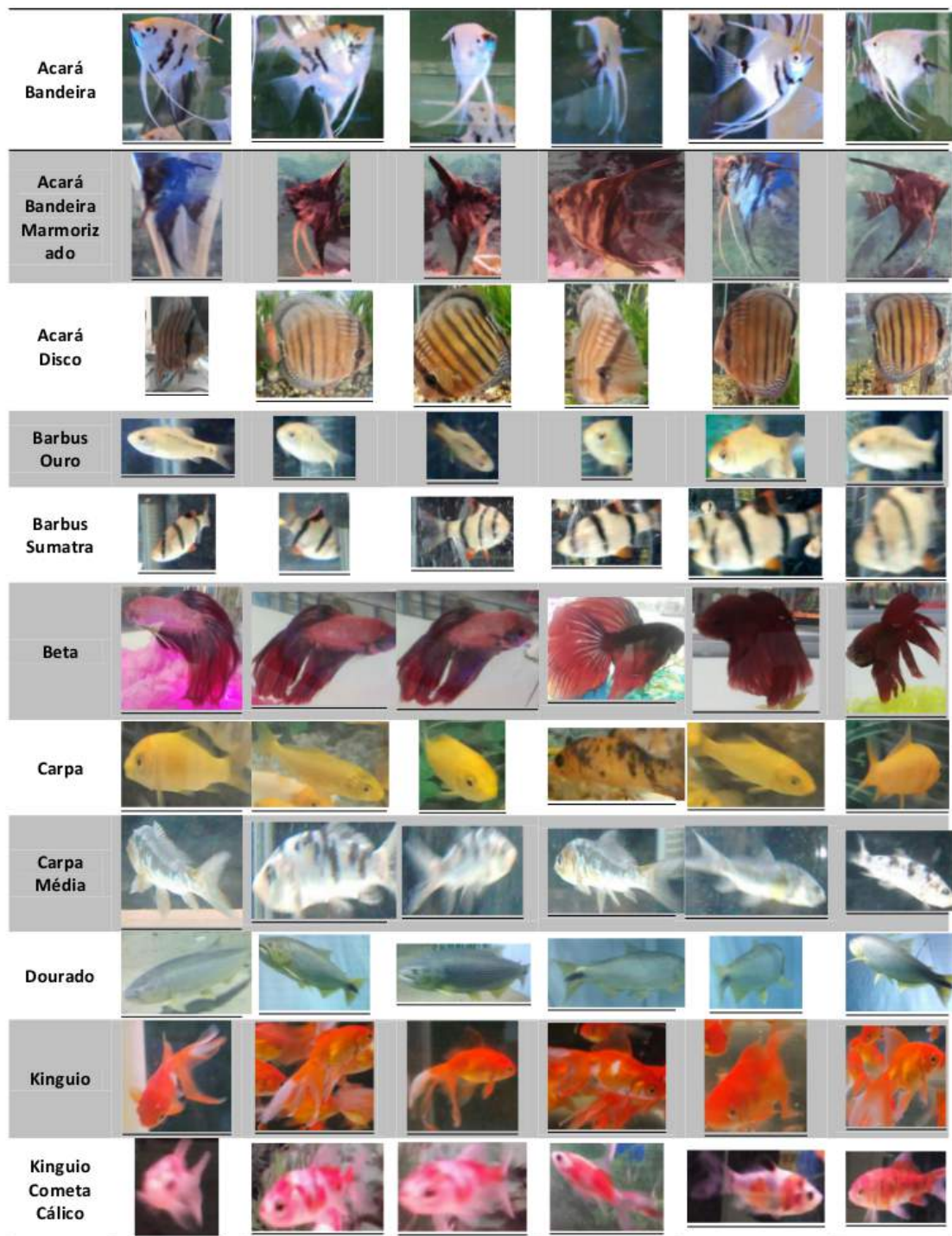


Figura 3.4: Amostras contendo 6 imagens das 40 de 11 espécies de peixes.



Figura 3.5: Amostras contendo 6 imagens das 40 de 12 espécies de peixes.



Figura 3.6: Amostras contendo 6 imagens das 40 de 5 espécies de peixes.

O banco de imagens contém 28 espécies de peixes e as mesmas então ilustradas na Figura 3.3 contendo o nome popular da espécie, nome científico e uma das imagens utilizadas nos experimentos.

Para avaliar o desempenho dos classificadores na classificação das espécies citadas, foram utilizadas as métricas descritas a seguir.

3.3 Métricas

Duas métricas foram utilizadas na avaliação dos classificadores e extratores de atributos: Área sobre a Curva (AUC) provida de uma análise ROC e a Medida-F. A seguir, ambas serão descritas.

3.3.1 Característica de Operação do Receptor - (Receiver Operating Characteristics - ROC)

As análises ROC [21] são utilizadas para visualização, ordenação e seleção de classificadores baseados em seus desempenhos. A aplicação dos gráficos ROC é vasta na comunidade médica, sendo utilizada na visualização e análise de comportamentos de sistemas de diagnósticos. Na comunidade médica os gráficos de curvas ROC possuem uma extensiva literatura e são utilizados em testes de diagnósticos.

Em inteligência artificial Spackman et al. [24] utilizam os gráficos ROC para analisar e comparar algoritmos. Esta técnica vem sendo adotada devido as outras métricas como acurácia e precisão serem pobres para comparar algoritmos (principalmente em banco de imagens desbalanceados).

Considerando inicialmente um problema binário, cada instância I mapeia um elemento de um conjunto $\{p, n\}$ representando as classes positivas e negativas do problema. Dada uma nova instância, o modelo de classificação ou classificador tenta prever em qual classe a mesma será mapeada. Para diferenciar a classe original de uma instância e a classe predita pelo classificador, as classes preditas serão denotadas da seguinte forma: $\{P, N\}$.

Dada uma nova instância, há quatro possibilidades de classificações: 1) a instância é positiva e classificada como positiva, ou seja, um verdadeiro positivo. 2) a instância é positiva e classificada como negativa sendo um falso positivo. 3) a instância é negativa e classificada como negativa, sendo assim um verdadeiro negativo ou caso contrário, 4) formando um falso negativo.

Sendo assim, dado um classificador e o conjunto de instâncias, uma matriz 2×2 denominada matriz de confusão (ou matriz de contingência) pode ser construída. A Figura 3.7 ilustra uma matriz de confusão onde a diagonal representa as classificações corretas do classificador, e o restante as incorretas.

		Classe Verdadeira	
		p - Positivo	n - Negativo
Classe Hipotética	P - Positivo	VP - Verdadeiro Positivo	FP - Falso Positivo
	N - Negativo	FN - Falso Negativo	VN - Verdadeiro Negativo

Figura 3.7: Matriz de confusão ou de contingência.

A partir da matriz de confusão, métricas como as representadas pelas Equações 3.1, 3.2, 3.3, 3.4 podem ser derivadas e, baseados nas métricas *TVP* e *TFP*, o espaço ROC é criado.

$$\text{Taxa de Verdadeiros Positivos} = TVP = \frac{VP}{p} \quad (3.1)$$

$$\text{Taxa de Falso Positivo} = TFP = \frac{FP}{n} \quad (3.2)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (3.3)$$

$$\text{Revocação} = \frac{VP}{VP + FN} \quad (3.4)$$

O espaço ROC é um formado por duas dimensões sendo que: a *TVP* representa o eixo *y* e a *TFP* no eixo *x*. O espaço ROC retrata a comparação entre os benefícios (verdadeiro positivos) e custos (falsos positivos) da classificação de um classificador. Cada ponto no espaço ROC é formado a partir de um par (*TFP*, *TVP*), obtido de um classificador discreto, isto é, tem como saída apenas uma classe.

Alguns pontos do espaço ROC requerem destaque. O ponto inferior esquerdo (0,0) indica que o classificador classifica todos os exemplos como negativos, assim não há falsos negativos e verdadeiro positivos classificados. Já o ponto (1,1) indica o oposto, isto é, dado um exemplo

o mesmo sempre será positivo. O ponto $(0,1)$ indica a classificação perfeita, ou seja, todos os exemplos positivos foram classificados como positivos. Um classificador, no espaço ROC, classifica melhor as classes que outro se o ponto correspondente a ele está mais próximo do ponto $(0,1)$. A linha diagonal $y = x$ representa que o classificador está classificando um exemplo de forma aleatória, por exemplo, se o classificador é aleatório e classifica 90% das vezes o exemplo como positivo, pode ser esperado que 90% dos verdadeiros positivos sejam encontrados, assim como a taxa de falso positivo será incrementada em 90% obtendo o ponto $(0.9,0.9)$ no espaço ROC.

Os classificadores discretos geram apenas um ponto no espaço ROC já que têm como resposta apenas classificações binárias. Alguns classificadores, ao classificar um exemplo, geram uma pontuação (probabilidades ou um ranqueamento) associados a classificação e pode variar sua pontuação gerando vários pontos no espaço ROC. A pontuação pode variar usando um limiar, onde, se o classificador produz como resultado uma pontuação maior que o limiar, a classificação é dita como positiva (P) e negativa (N) caso contrário. Deste modo é possível variar o limiar gerando vários pontos no espaço ROC e escolhendo o melhor. O limiar pode variar de $-\infty$ a $+\infty$ trançando a curva no espaço ROC.

Mas para comparar classificadores de uma forma simples, é necessário criar um escalar que represente as performances esperadas. Um método de alcançar este objetivo é calculando a área sob a curva (*Area Under ROC* - AUC). Como a AUC é uma fração do quadrado unitário que forma o espaço ROC, a área varia de 0 a 1.

A AUC e o espaço ROC como descritos tratam apenas de problemas binários, mas muitas vezes é desejável uma análise ROC de um problema com múltiplas classes. Neste trabalho, por exemplo, são 28 classes correspondendo as espécies de peixes. É possível transformar o problema de múltiplas classes em binário utilizando a técnica um versus todos, onde, define-se uma classe que será rotulada como positiva e o restante do do conjunto de treinamento como negativo. Repete-se o processo para todas as classes. Ao final, serão formadas n curvas correspondendo a n classes do problema. Uma média das curvas é feita para gerar uma curva única representando o classificador. Neste trabalho, a técnica para obter a média das curvas é a média vertical, em que, escolhe-se a TVP máxima para cada TFP , interpolando-os se necessário, e calculando a média.

3.3.2 Medida-F (*F-Score*)

A principal métrica utilizada para discriminar os classificadores e extratores foi a Medida-F (*F-Score*) obtida através da média harmônica da precisão e revocação. Ambas são calculadas utilizando os verdadeiros positivos (VP), falsos positivos (FP) e falsos negativos (FN), que podem ser obtidos por meio de uma matriz de confusão como ilustrado na Figura 3.7. A Medida-F não penaliza somente classificadores com alta taxa de falsos positivos e negativos, mas ambos. A Medida-F é obtida por meio da seguinte equação:

$$\text{Medida-F} = 2 * \frac{\text{Precisão} * \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (3.5)$$

Assim como a AUC, a Medida-F é obtida a partir de problemas binários. Contudo, pode-se generalizar a Medida-F para problema com múltiplas classes. Para tal fim, considere que exemplos pertencentes a classe são denominados positivos e exemplos pertencentes a qualquer outra classe são denominados negativos. Este processo é repetido para as n classes do problema. Seguindo o conceito citado, há dois métodos de obter-se a Medida-F, a macro Medida-F e micro Medida-F. Neste trabalho a macro Medida-F é utilizada.

Para obter o valor da Medida-F utilizando a macro Medida-F, calcula-se a média (ponderada ou aritmética) dos Verdadeiros Positivos, Falsos Positivos e Verdadeiros Positivos do problema. As Equações 3.6 e 3.7 exemplificam a média aritmética das métricas verdadeiro positivo (VP) e falsos positivos (FP) de um problema envolvendo n classes.

$$\overline{VP} = \frac{\sum_{i=1}^n VP_i}{n} \quad (3.6)$$

$$\overline{FP} = \frac{\sum_{i=1}^n FP_i}{n} \quad (3.7)$$

Após os cálculos das médias de VP e FP, a Macro Medida-F é obtida recorrendo à Equação 3.8.

$$\text{Macro Medida-F} = 2 * \frac{\overline{VP} * \overline{FP}}{\overline{VP} + \overline{FP}} \quad (3.8)$$

A seção a seguir descreve os experimentos utilizando as métricas definidas nesta seção, juntamente com os extratores citados no Capítulo 2 e os classificadores descritos na Subseção 3.1.

Experimentos

Para cada extrator foram gerados dicionários com diferentes tamanhos que descrevem cada espécie de peixes, assim como os parâmetros dos classificadores a fim de encontrar a combinação de extrator de atributos e classificador com melhor Medida-F e AUC. No caso dos histogramas HSV e RGB, o que foi variado foi a quantidade de faixas em que cada canal de cor foi dividido. Para classificar as imagens, foram utilizados os classificadores: Árvore de decisão, KNN e SVM, descritos na Seção 3.1. Os experimentos foram feitos com implementações dos classificadores fornecidos pelo Scikit-learn [22].

Os parâmetros de todos os classificadores foram variados para obter os melhores resultados na classificação utilizando validação cruzada com 10 partições e repetições. Sendo assim, a descrição das imagens fornecidas de cada técnica foram submetidas aos classificadores citados com todas as variações de parâmetros descritos a seguir.

4.1 *Variações dos Parâmetros dos Classificadores*

Os parâmetros dos classificadores foram variados da seguinte forma:

- SVM: foram utilizados os núcleos linear e RBF. Segundo Chang et al. [1] o núcleo linear possui melhor desempenho para um conjunto de treinamento com grande quantidade de atributos, se encaixando no contexto deste trabalho. O núcleo RBF pode se adaptar melhor a certos conjuntos de treinamento com atributos não lineares, sendo assim, o mesmo foi inserido nos experimentos. Os valores de C e γ foram variados no espaço logarítmico correspondendo aos valores: $\log_2^{-5}$ a $\log_2^{15}$ para C e $\log_2^{-15}$ a $\log_2^3$ para γ .
- KNN: o valor de K foi variado de 1 à 500 incrementando em uma unidade. As métricas utilizadas para calcular a distância dos pontos também variaram. Foi utilizada a métrica uniforme, mantendo a distância dos pontos de forma uniforme, e o inverso da distância, isto é, pontos mais próximos tem maior peso na classificação.
- Árvore de decisão: como estratégia de divisão de cada nó da árvore, foi utilizada a melhor divisão ou uma aleatória. O critério de divisão também foi variado entre entropia e “gini” para impuridade de Gini.

Todas as permutações de parâmetros foram aplicadas as variações de dicionários das técnicas baseadas no BoW, assim como, nas variações de faixas dos histogramas.

A Seção 4.2 relata os experimentos feitos em cada técnica de extração de atributos a fim de encontrar o extrator que, juntamente com um classificador, terá melhores resultados na classificação das espécies de peixes.

4.2 Determinação dos Parâmetros dos Extratores de Atributos

Para determinar o tamanho do dicionário que melhor represente as características de cada espécie, segundo os extratores relatados, foram gerados dicionários com tamanhos: 32, 64, 128, 256, 512, 1024, 2048, 4096 e 8192 para os extratores de atributos BoVW, BoFC, BoC e BoCW. Devido a descreverem as imagens de forma diferente quando comparado aos extratores baseados no BoVW, a quantidade de atributos que descrevem uma imagem é diferente nos histogramas HSV e RGB.

A variação ocorre devido aos histogramas dividirem cada canal do espaço de cores em uma quantidade de faixas que variam entre 8, 16, 32. O número mínimo de faixas foi definido em 8 (512 atributos), pois para valores menores, é considerável a perda de informações de cor. O número máximo de faixas é 32 (32,768 atributos) devido as limitações de *hardware* encontradas nos computadores utilizados nos experimentos. Os experimentos a seguir foram executados com validação cruzada de 10 partições e 10 repetições com todos os parâmetros citados na Seção 4.1. Apenas o resultado da melhor configuração de parâmetros são mostradas nas tabelas a seguir.

4.2.1 Experimentos com o BoVW

Observando a Tabela 4.1 é possível observar que o dicionário de tamanho 4096 obteve melhor Medida-F e AUC com o classificador SVM. Os parâmetros do SVM para alcançar este valor foram os seguintes: núcleo linear e $C = 24826.608981569752$. A Figura 4.1 ilustra a AUC média entre as classes juntamente com a AUC por cada espécie de peixe. Os parâmetros do KNN para seu melhor resultado, isto é, para o dicionário 1024 foram $k = 8$ utilizando como a distância como peso na escolha dos vizinhos. A árvore de decisão obteve melhor resultado com o dicionário 2048 utilizando entropia.

Tabela 4.1: Árvore de decisão juntamente com BoVW.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.16	0.822183	0.15	0.646302	0.18	0.571267
64	0.17	0.735667	0.14	0.714605	0.22	0.598646
128	0.21	0.803140	0.16	0.705625	0.21	0.6
256	0.31	0.833187	0.19	0.710104	0.25	0.619818
512	0.38	0.874362	0.20	0.750654	0.25	0.638622
1024	0.49	0.902782	0.23	0.739506	0.26	0.617659
2048	0.57	0.927132	0.21	0.680064	0.26	0.635244
4096	0.59	0.944135	0.20	0.650344	0.24	0.634283
8192	0.56	0.933591	0.14	0.560329	0.22	0.634658

Analisando a matriz de confusão ilustrada na Figura 4.2 é verificado que a espécie peixe Palhaço foi a mais confundida com as espécies: Platy Sangue, Platy Laranja, Kinguio e Carpa Média.

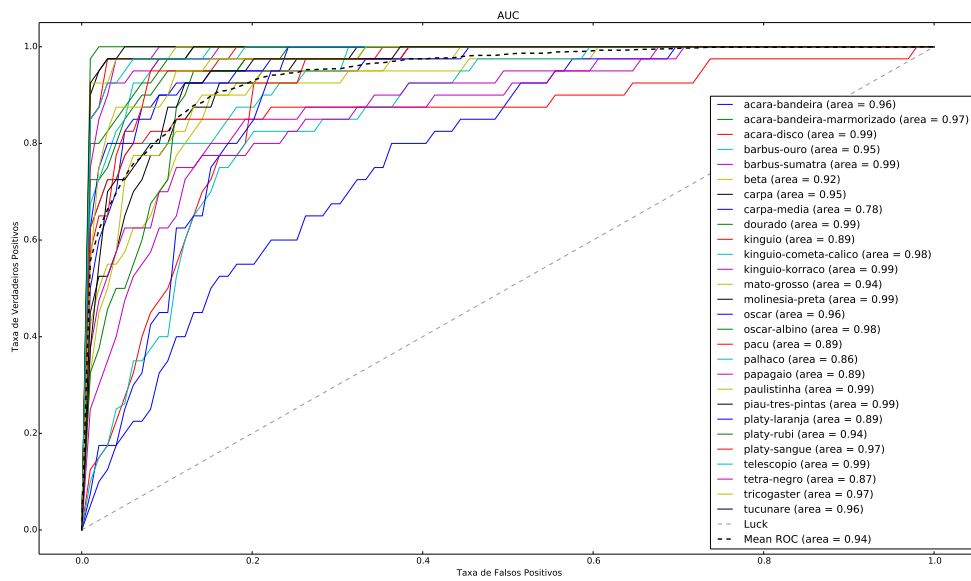


Figura 4.1: Área sob a curva do dicionário de tamanho 4096 do BoVW utilizando o SVM como classificador.

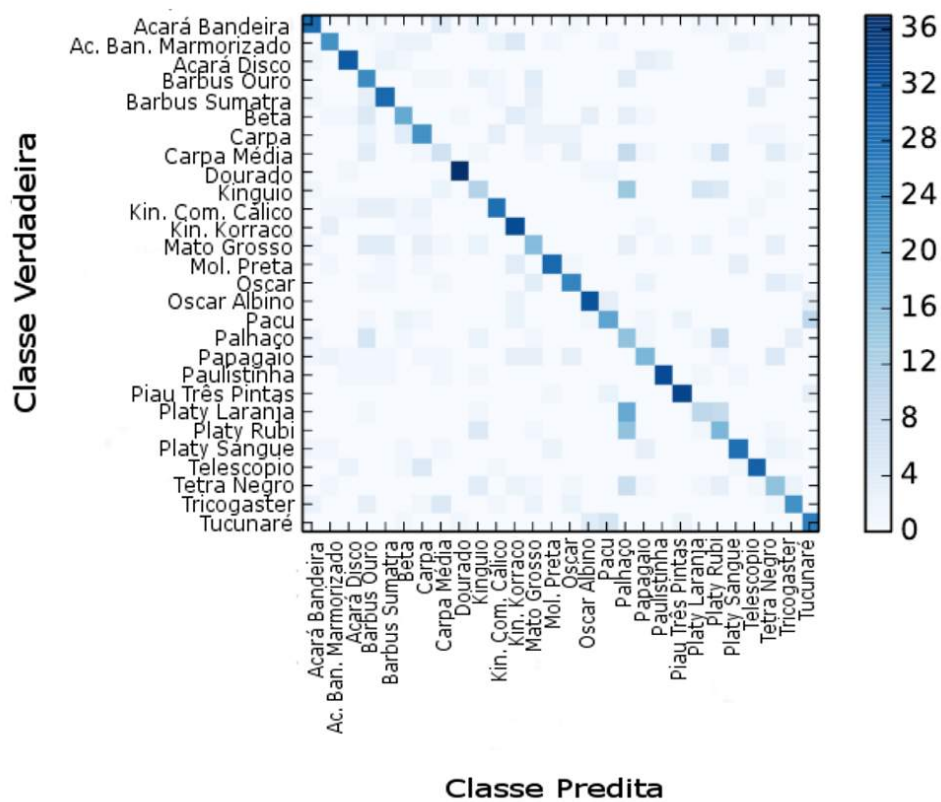


Figura 4.2: Matriz de confusão da classificação das espécies utilizando o dicionário de tamanho 4096 do BoVW juntamente com o classificador SVM.

4.2.2 Experimentos com o BoFC

O tamanho de dicionário que destacou-se sobre os demais foi o de tamanho 2048 com Medida-F igual a 0.8. O resultado foi obtido utilizando como classificador o SVM com $C = 8.86516908684$, $\gamma = 8.08386864682e - 05$ e núcleo RBF. A Figura 4.3 ilustra a AUC de cada classe e a Tabela 4.2 ilustra os resultados da Medida-F para as variações de tamanhos de dicionários. O dicionário de tamanho 2048 também obteve destaque entre os demais utilizando o classificador árvore de decisão utilizando o critério de divisão aleatório e Gini. O KNN obteve melhor resultado com o dicionário de tamanho 1024 com $K = 1$.

Tabela 4.2: Comparação de Tamanhos de Dicionários do BoFC.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.25	0.823231	0.24	0.799302	0.29	0.602282
64	0.42	0.889544	0.39	0.833239	0.40	0.661901
128	0.56	0.929800	0.53	0.908964	0.47	0.727013
256	0.68	0.955215	0.58	0.860894	0.48	0.750654
512	0.76	0.967876	0.62	0.870247	0.53	0.751483
1024	0.77	0.969725	0.63	0.784538	0.55	0.748871
2048	0.80	0.972607	0.56	0.754933	0.56	0.748999
4096	0.78	0.974841	0.49	0.727618	0.53	0.758953
8192	0.79	0.968803	0.37	0.608176	0.49	0.733589

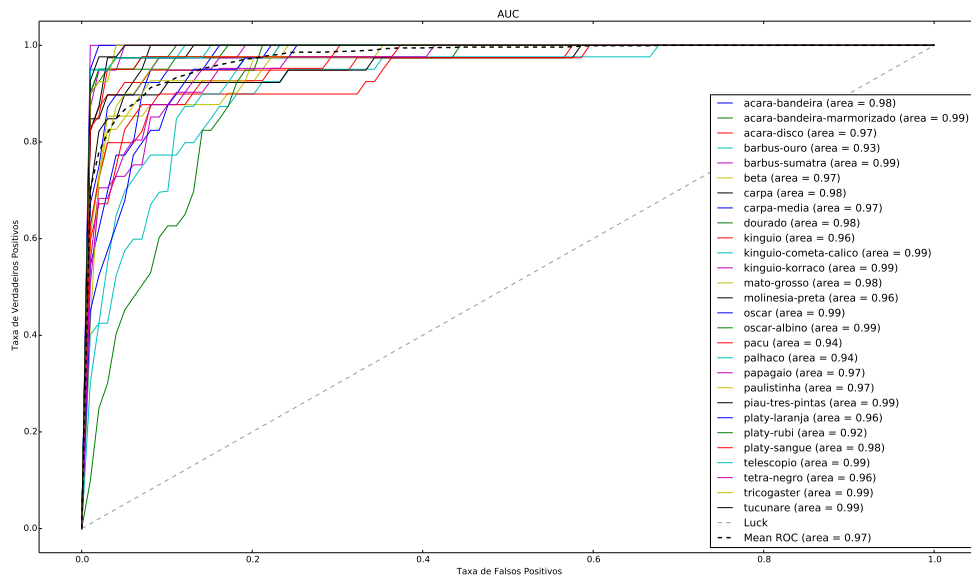


Figura 4.3: Área sob a curva do dicionário de tamanho 2048 com a técnica BoFC utilizando o SVM como classificador.

A Figura 4.4 ilustra a matriz de confusão do dicionário com tamanho 2048 utilizando o SVM.

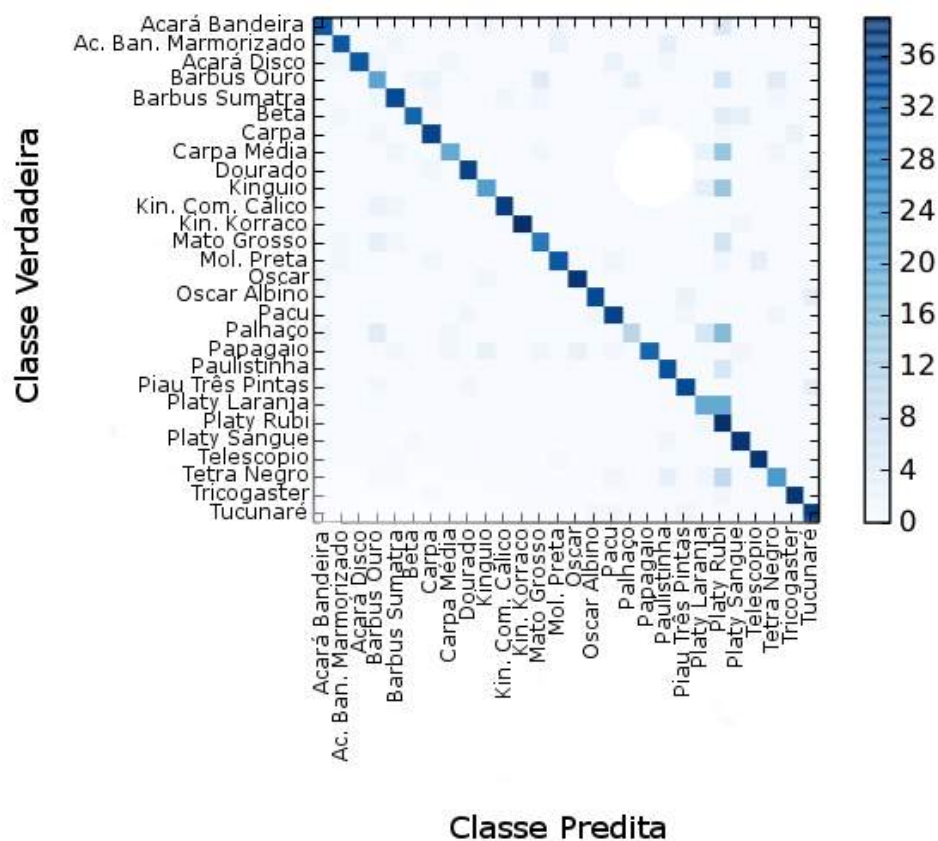


Figura 4.4: Matriz de confusão da classificação das espécies utilizando o dicionário de 2048 com a técnica BoFC e o classificador SVM.

4.2.3 Experimentos com o BoC

O dicionário com melhor desempenho em relação a Medida-F foram os de tamanhos 2048 e 8192. Utilizando como critério de desempate a AUC, o dicionário que melhor descreve as espécies de peixes é o de tamanho 2048. As configurações dos parâmetros do SVM foram as seguintes para o melhor resultado encontrado: $C = 1635.68097512$, $\gamma = 0.0471890060599$ e núcleo RBF. Com $K = 1$ e distância uniforme, o KNN obteve melhor resultado utilizando o dicionário de tamanho 8192, assim como a árvore de decisão utilizando entropia e escolhendo a melhor divisão para cada nó de decisão. A Tabela 4.3 ilustra os resultados da Medida-F para as variações de tamanhos de dicionários. A Figura 4.5 ilustra a AUC média assim como a de cada classe.

Tabela 4.3: Comparação de Tamanhos de Dicionários do BoC.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.77	0.966132	0.69	0.825876	0.59	0.753064
64	0.83	0.973133	0.75	0.857908	0.64	0.780888
128	0.84	0.974144	0.76	0.862319	0.62	0.800545
256	0.85	0.977770	0.78	0.879424	0.64	0.817035
512	0.89	0.985639	0.80	0.880192	0.65	0.835304
1024	0.88	0.983281	0.82	0.889093	0.69	0.810591
2048	0.90	0.983348	0.80	0.883555	0.68	0.840596
4096	0.88	0.979725	0.77	0.869325	0.71	0.852780
8192	0.90	0.981404	0.90	0.864816	0.77	0.857722

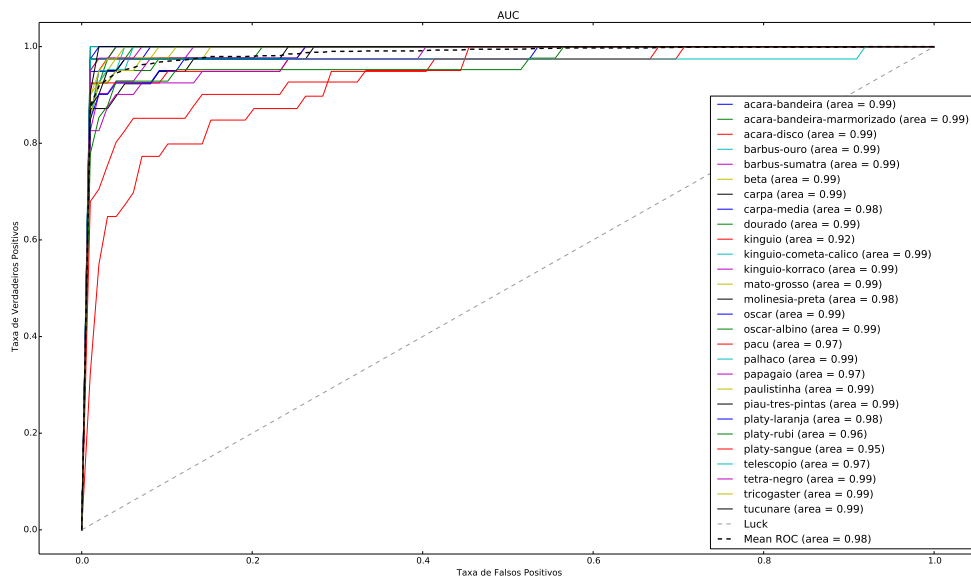


Figura 4.5: Área sob a curva do dicionário de tamanho 2048 do BoC utilizando o SVM como classificador.

Observando a Figura 4.6 é verificado que algumas imagens da espécie Kinguio foram confundidas como sendo das espécies Palhaco e Papagaio. Assim como, 4 Imagens foram classificadas como peixe Palhaco e 3 como Papagaio.

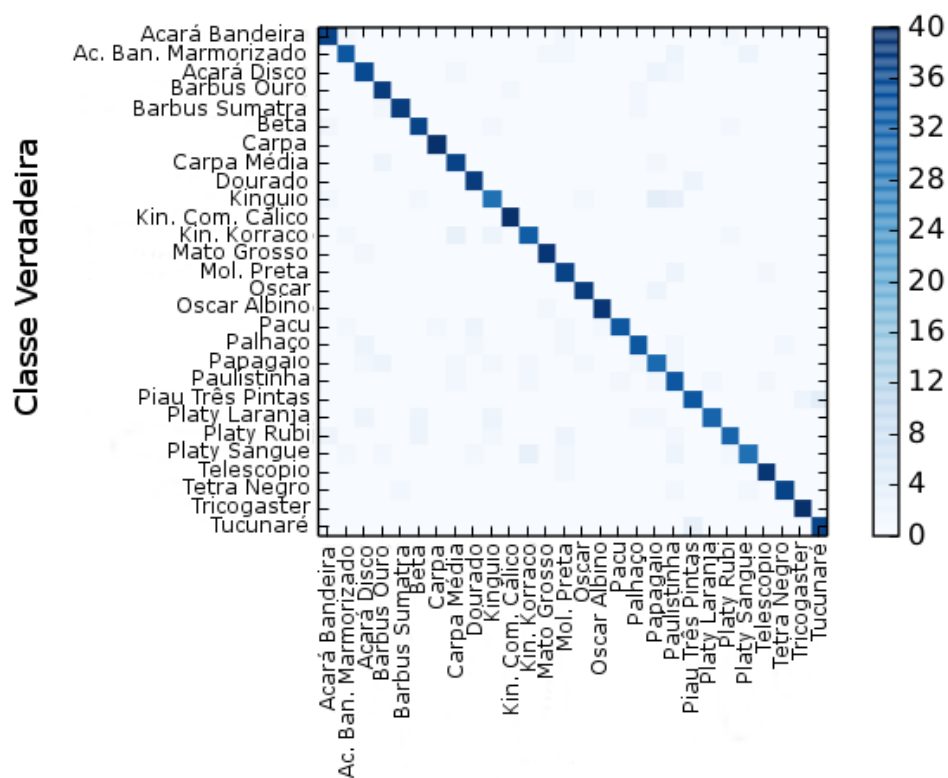


Figura 4.6: Matriz de confusão da classificação das espécies utilizando a técnica BoC, o dicionário de tamanho 2048, e o classificador SVM.

4.2.4 Experimentos com o BoCW

Devido ao BoCW ser uma técnica derivada do BoVW e do histograma de cores HSV, a quantidade de atributos que descreve cada imagem varia no tamanho do dicionário utilizado pelo BoVW e o total de faixas do histograma. Assim, para cada tamanho de dicionário, a quantidade de faixas variam entre 2, 4, 8, 16 e 32. Os dicionários de tamanho 512 e 1024, ambos com 32 faixas, obtiveram o mesmo valor de Medida-F pontuando 0.88. Mas tomando como critério de desempate a AUC, o dicionário de tamanho 1024 foi o escolhido. A Figura 4.8 ilustra a matriz de confusão do dicionário com o classificador SVM utilizando os parâmetros $C = 0.0625433801116$ e núcleo linear. Utilizando $k = 10$ e peso nas distâncias, o KNN obteve melhor resultado com o dicionário de tamanho 32 e 4 faixas. Já a árvore de decisão obteve melhor resultado com o dicionário de tamanho 32 e 8 faixas. As Tabelas 4.4, 4.5, 4.6, 4.7, 4.8 ilustram os resultados da Medida-F para as variações de tamanhos de dicionários e de faixas do BoCW. A Figura 4.7 ilustra a AUC média cada classe e a Figura 4.8 a matriz de confusão.

Tabela 4.4: Comparação de Tamanhos de Dicionários do BoCW utilizando 2 faixas.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.25	0.853761	0.18	0.666530	0.60	0.785920
64	0.19	0.777248	0.16	0.732376	0.57	0.792119
128	0.23	0.796126	0.18	0.786618	0.58	0.771811
256	0.30	0.829388	0.19	0.708550	0.58	0.778964
512	0.39	0.879270	0.20	0.693519	0.57	0.785410
1024	0.49	0.917345	0.23	0.688076	0.59	0.776050
2048	0.57	0.927044	0.20	0.690839	0.56	0.775586
4096	0.56	0.932277	0.19	0.592829	0.57	0.758796
8192	0.56	0.938762	0.15	0.570143	0.57	0.771387

Tabela 4.5: Comparação de Tamanhos de Dicionários do BoCW utilizando 4 faixas.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.37	0.910880	0.23	0.702202	0.73	0.834611
64	0.27	0.845286	0.19	0.768447	0.71	0.832797
128	0.29	0.828897	0.21	0.822038	0.71	0.822260
256	0.35	0.868241	0.20	0.702352	0.68	0.846268
512	0.43	0.888522	0.22	0.697997	0.67	0.815069
1024	0.50	0.907752	0.23	0.690047	0.69	0.800505
2048	0.58	0.933665	0.21	0.679445	0.67	0.783547
4096	0.59	0.935017	0.20	0.677200	0.67	0.774607
8192	0.57	0.929949	0.15	0.571069	0.70	0.787904

Tabela 4.6: Comparação de Tamanhos de Dicionários do BoCW utilizando 8 faixas.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.71	0.938564	0.59	0.688618	0.74	0.847931
64	0.70	0.932242	0.57	0.605226	0.73	0.851775
128	0.66	0.887275	0.56	0.751549	0.70	0.845994
256	0.67	0.896095	0.51	0.757922	0.72	0.865239
512	0.71	0.918506	0.49	0.692934	0.71	0.835991
1024	0.68	0.927345	0.43	0.608669	0.69	0.860066
2048	0.68	0.939728	0.33	0.601283	0.69	0.816168
4096	0.58	0.943263	0.19	0.644761	0.70	0.812724
8192	0.55	0.938754	0.14	0.569301	0.67	0.857327

Tabela 4.7: Comparação de Tamanhos de Dicionários do BoCW utilizando 16 faixas.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.87	0.966378	0.66	0.716208	0.71	0.845019
64	0.82	0.953373	0.67	0.695929	0.72	0.858350
128	0.78	0.937392	0.66	0.688476	0.72	0.858697
256	0.79	0.933874	0.65	0.674030	0.70	0.856317
512	0.81	0.946356	0.63	0.663213	0.72	0.858731
1024	0.79	0.949008	0.58	0.639094	0.71	0.843689
2048	0.79	0.950397	0.50	0.607123	0.71	0.832967
4096	0.63	0.951759	0.21	0.647789	0.71	0.857185
8192	0.60	0.946492	0.16	0.575614	0.72	0.859337

Tabela 4.8: Comparação de Tamanhos de Dicionários do BoCW utilizando 32 faixas.

Tamanho dos Dicionários	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
32	0.87	0.978157	0.66	0.772911	0.71	0.835395
64	0.87	0.978598	0.68	0.762312	0.72	0.825212
128	0.86	0.961626	0.66	0.738020	0.72	0.838937
256	0.86	0.971717	0.65	0.743238	0.71	0.822060
512	0.88	0.972709	0.65	0.730703	0.72	0.840050
1024	0.88	0.973656	0.63	0.697975	0.71	0.821125
2048	0.87	0.972574	0.59	0.650194	0.71	0.824874
4096	0.72	0.970887	0.27	0.669057	0.71	0.819003
8192	0.68	0.967433	0.19	0.590258	0.69	0.816841

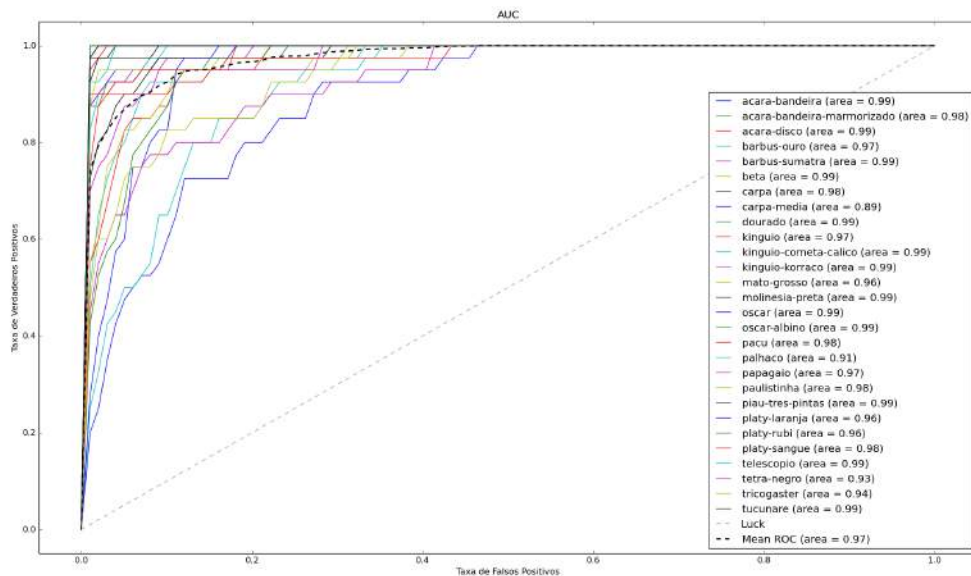


Figura 4.7: Área sob a curva do dicionário de tamanho 1024 com 32 faixas do BoCW utilizando o SVM como classificador.

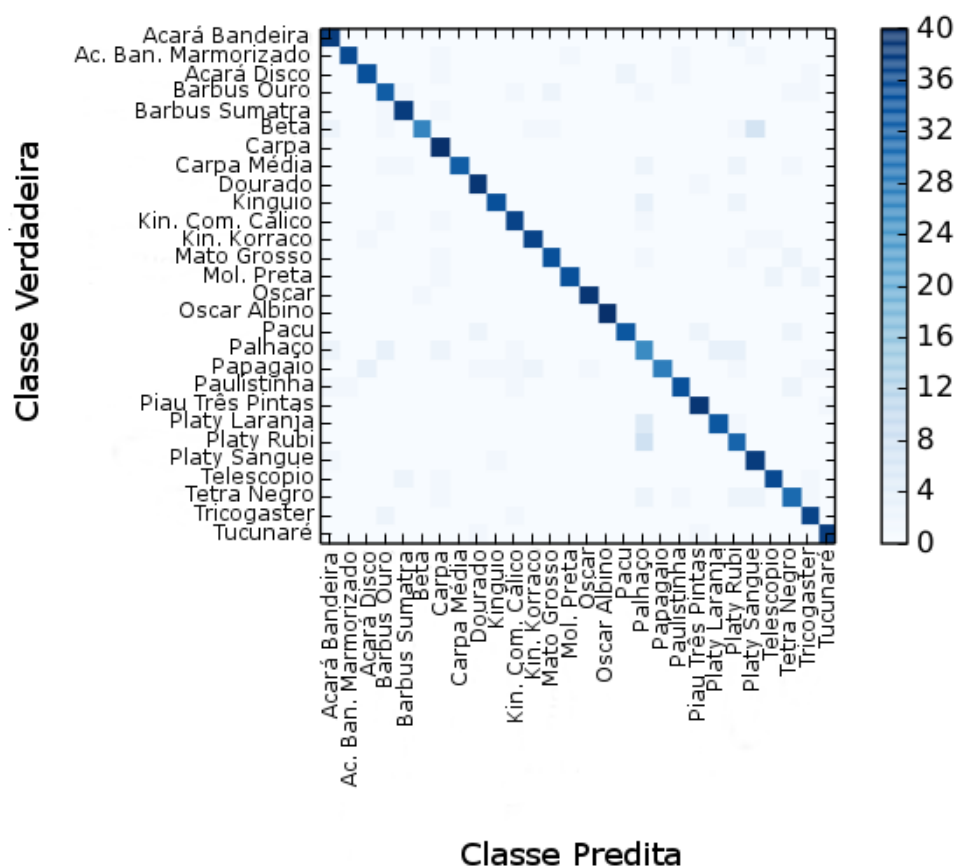


Figura 4.8: Matriz de confusão da classificação das espécies utilizando a técnica BoCW, o dicionário de tamanho 1024, 32 faixas e o classificador SVM.

4.2.5 Experimentos com o Histograma de Cores HSV

Utilizando como técnica um histograma de cor no espaço de cores HSV, juntamente com o classificador SVM com núcleo RBF, $C = 16833.4006851$ e $\gamma = 0.00210510528871$, a melhor Medida-F obtida foi 0.89 juntamente com uma AUC igual à 0.981666 (Figura 4.9) utilizando 32 faixas. A Figura 4.10 ilustra a matriz de confusão obtida no experimento e a Tabela 4.9 todos os resultados. O classificador KNN obteve melhor resultado utilizando 8 faixas com $K = 1$, assim como a árvore de decisão utilizando entropia e a melhor divisão para os nós de decisão.

Tabela 4.9: Comparação de quantidade de faixas do Histograma de cores HSV.

Faixas	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
04	0.83	0.959969	0.80	0.898065	0.74	0.859452
08	0.86	0.967860	0.86	0.913217	0.83	0.876275
16	0.87	0.974043	0.77	0.883935	0.71	0.850113
32	0.89	0.981666	0.70	0.851477	0.69	0.838484

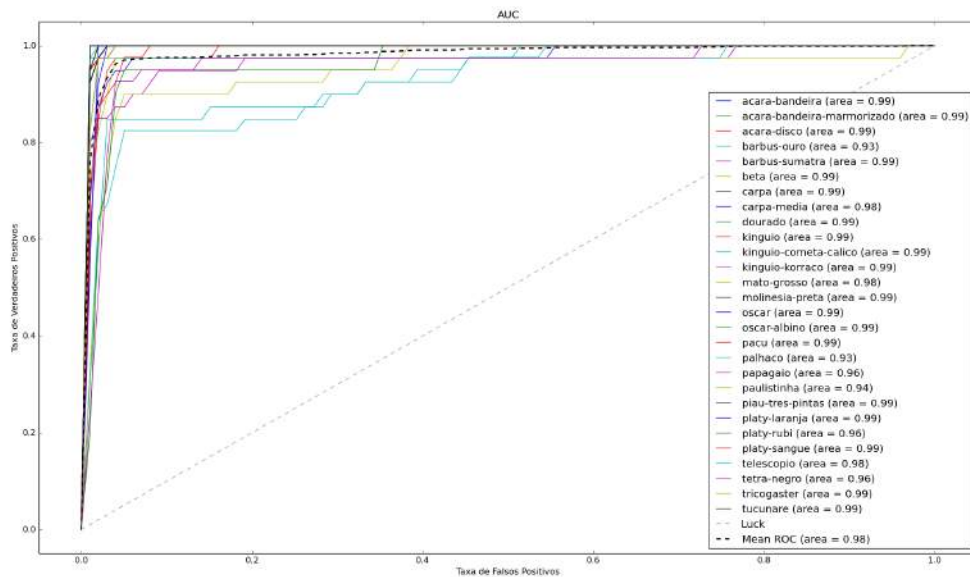


Figura 4.9: Área sob a curva da divisão do espaço de cor HSV em 32 faixas.

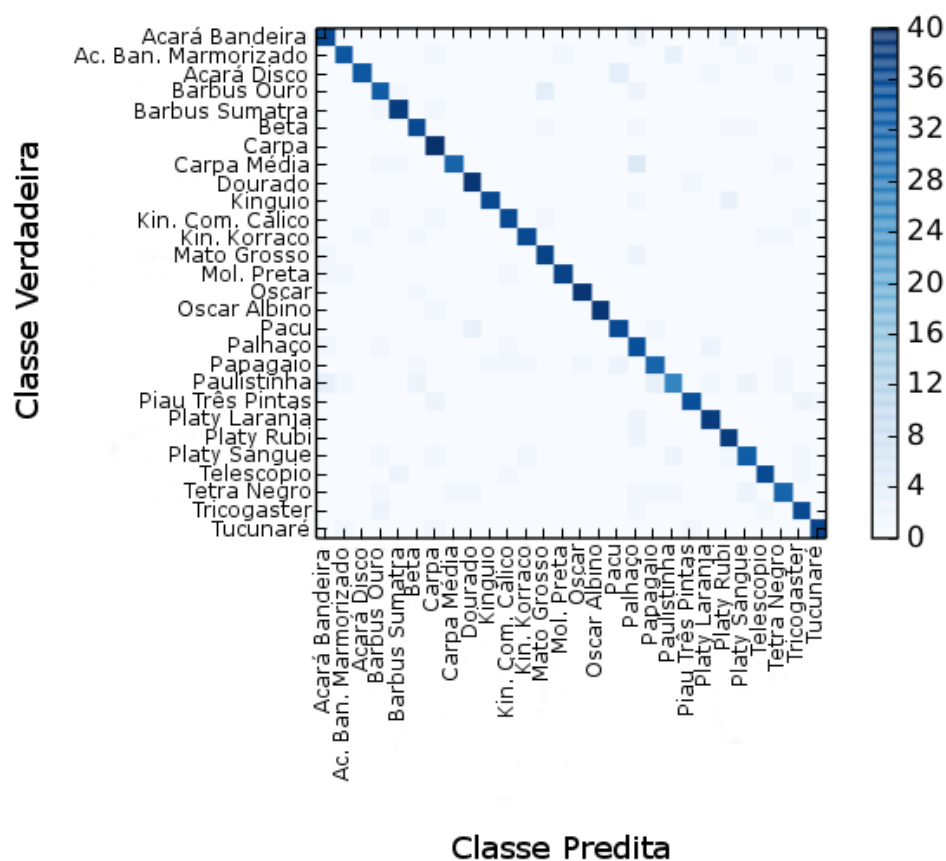


Figura 4.10: Matriz de confusão do histograma de cores HSV utilizando SVM como classificador e 32 faixas.

4.2.6 Experimentos com o Histograma de Cores RGB

A melhor Medida-F obtida nos experimentos utilizando o histograma de cor RGB foi de 0.88 com 16 faixas. Os parâmetros do classificador SVM foram os seguintes: núcleo linear e $C = 6.3540113525$. O KNN obteve melhor resultado utilizando 8 faixas com $K = 1$ juntamente com a árvore de decisão tendo como parâmetros a impuridade Gini e aplicando a melhor divisão. A Tabela 4.10 ilustra os resultados dos outros dicionários. A AUC média da classificação, assim como a AUC por classe, estão ilustradas na Figura 4.11. A matriz de confusão está ilustrada na Figura 4.12.

Tabela 4.10: Comparação de quantidade de faixas do Histograma de cores RGB.

Faixas	SVM		KNN		Árvore de Decisão	
	Medida-F	AUC	Medida-F	AUC	Medida-F	AUC
04	0.86	0.951616	0.83	0.913880	0.73	0.876630
08	0.87	0.963213	0.85	0.920641	0.76	0.863251
16	0.88	0.974300	0.81	0.908237	0.75	0.854777
32	0.87	0.986589	0.72	0.865965	0.70	0.844429

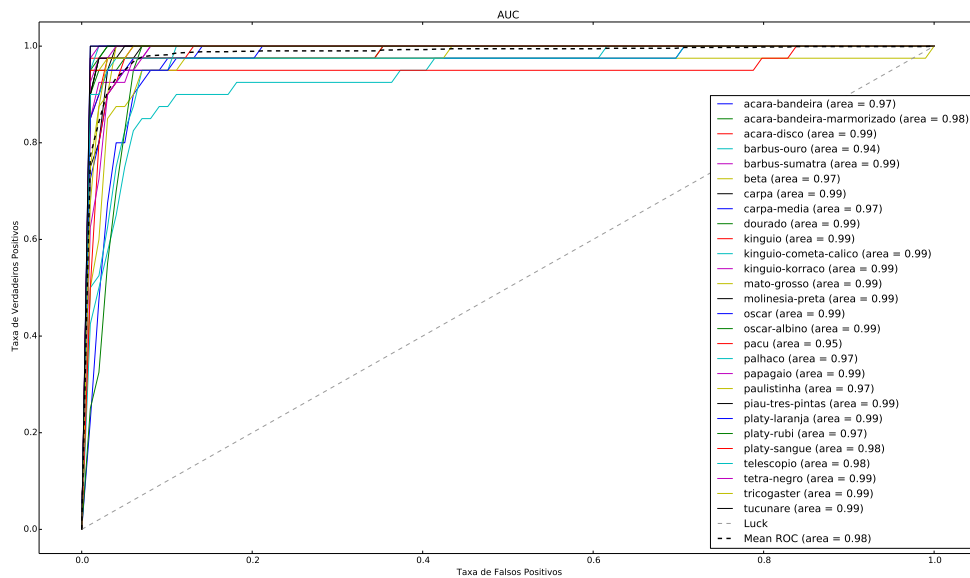


Figura 4.11: Área sob a curva da divisão do espaço de cor RGB em 16 faixas.

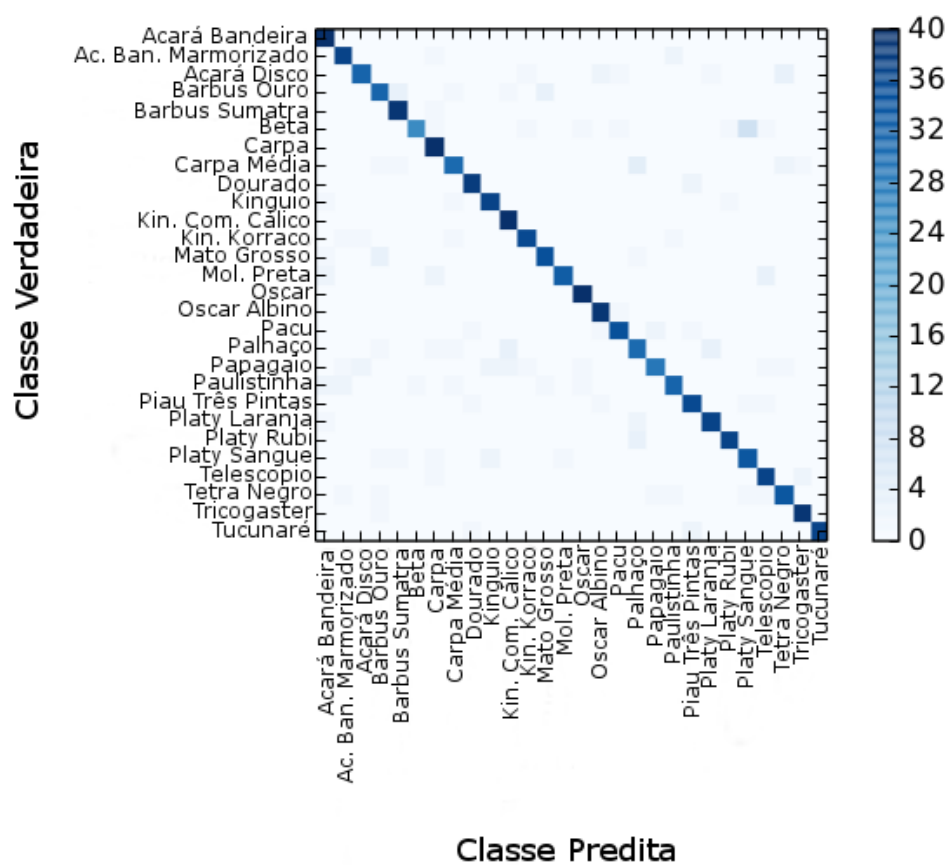


Figura 4.12: Matriz de confusão utilizando o SVM como classificador e o Histograma de cor RGB com 16 faixas.

Resultados e Discussões

Utilizando os dicionários com maior Medida-F juntamente com os classificadores otimizados, os extratores de atributos foram comparados a fim de encontrar o que melhor classifica as espécies de peixes apresentadas. Para uma maior confiabilidade dos resultados, o teste de hipótese de Friedman [12] foi aplicado para evidenciar a diferença estatística dos resultados, assim como o teste ANOVA.

A Tabela 5.1 ilustra o desempenho dos classificadores em relação a Medida-F e AUC de cada extrator de atributos. É possível observar que o classificador SVM obteve os melhores resultados para todos os extratores de atributos experimentados. Assim, os testes estatísticos foram executados utilizando a Medida-F dos dicionários ilustrados na Tabela 5.1. Analisando os extratores de atributos, o BoC obteve o melhor resultado seguido pelos histogramas de cores RGB e HSV.

Tabela 5.1: Comparação da Medida-F de cada classificador.

Extrator	Classificador	Dicionário/Faixa	Medida-F	AUC
BoVW	SVM	4096	0.59	0.944135
BoC	SVM	2048	0.90	0.983348
BoFC	SVM	2048	0.80	0.972607
BoCW	SVM	1024/16 faixas	0.88	0.973656
Histograma de Cores - HSV	SVM	32 faixas	0.89	0.981666
Histograma de Cores - RGB	SVM	16 faixas	0.88	0.974300

A Figura 5.1 ilustra o diagrama de caixas dos valores da Medida-F para cada técnica indicando as técnicas com maior estabilidade, isto é, que obtiveram baixa variação de Medida-F na classificação das espécies. Também é possível observar que os extratores de atributos BoC e histograma HSV e RGB possuem valores de Medida-F por volta de 0.9 e com pouca variação. A Figura 5.2 para cada espécie na seguinte ordem: Acará Bandeira, Acará Marmorizado, Acará Disco, Barbus Ouro, Barbus Sumatra, Beta, Carpa, Carpa Média, Dourado, Kinguio, Kinguio Cometa, Kinguio Korraco, Mato Grosso, Molinésia Preta, Oscar, Oscar Albino, Pacu, Peixe Palhaço, Papagaio, Paulistinha, Piau Três Pintas, Platy Laranja, Platy Sangue, Telescópio, Tetra Negro, Tricogaster, Tucunaré.

O teste de Friedman obteve um valor p de 1.035×10^{-08} para Medida-F e 1.275×10^{-13} para AUC. Sendo assim, o teste pós-teste [3] foi feito, e os resultados estão ilustrados na Tabela 5.2. De acordo com o pós-teste, é observado que os extratores baseados em cor possuem

acordo com as comparações do teste de Tukey, há diferença estatística entre o BoVW e todas as outras técnicas.

Tabela 5.2: Teste *post hoc*.

Extratores	BoFC	BoC	BoCW	Histograma HSV	Histograma RGB
BoVW	9.848803×10^{-03}	6.166622×10^{-08}	4.682090×10^{-08}	7.713548×10^{-09}	7.174929×10^{-07}
BoFC	•	1.578712×10^{-01}	1.029825×10^{-01}	7.800627×10^{-02}	3.238798×10^{-01}
BoC	•	•	9.999729×10^{-01}	9.997243×10^{-01}	9.991803×10^{-01}
BoCW	•	•	•	9.999979×10^{-01}	9.943114×10^{-01}
Histograma HSV	•	•	•	•	9.868104×10^{-01}

Tabela 5.3: Teste de Tukey utilizando Medida-F.

Extratores	BoFC	BoC	BoCW	Histograma HSV	Histograma RGB
BoVW	0	0	0	0	0
BoFC	•	0.0375672	0.1422833	0.0902713	0.1135536
BoC	•	•	0.9950130	0.9999287	0.9981378
BoCW	•	•	•	0.9996264	0.9999987
Histograma HSV	•	•	•	•	0.9999447

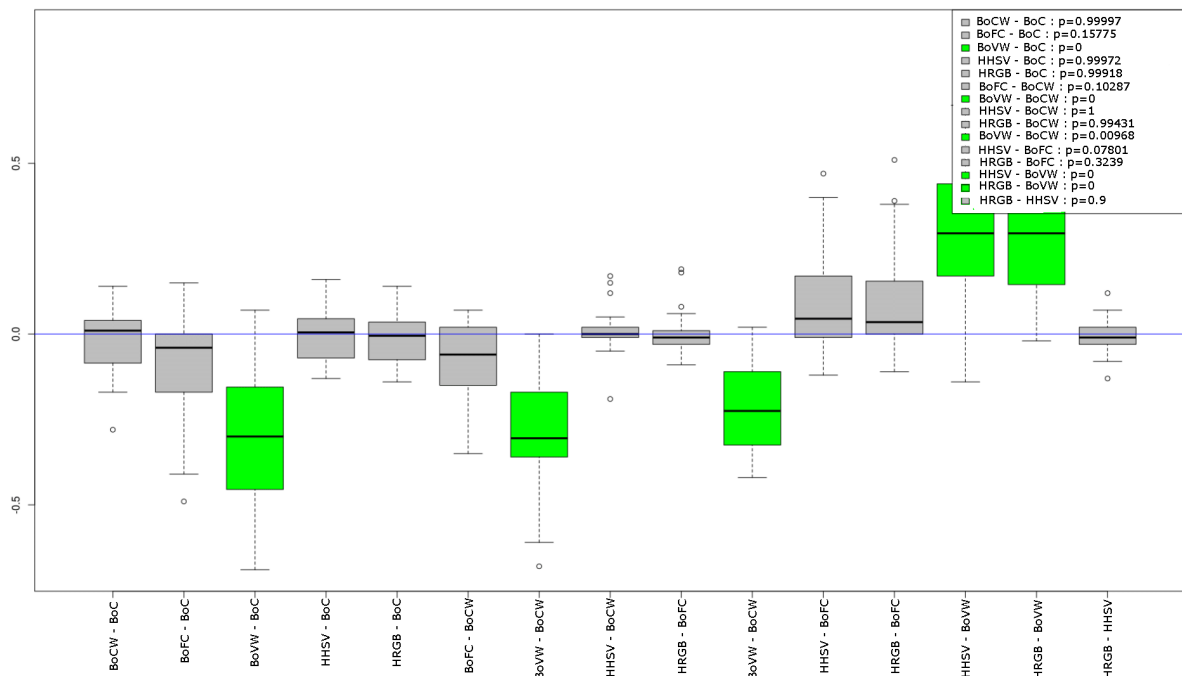


Figura 5.3: Diagrama das caixas do teste *post hoc* usando Medida-F.

A Figura 5.3 representa o diagrama de caixas do pós teste de Friedman, obtidos a partir dos dados da Tabela 5.2. Na figura é possível observar os extratores que se diferem (caixas verdes). Quanto maior a distância do eixo $y = 0$ maior é a diferença entre os extratores.

É notória a diferença de desempenho do BoVW em relação às demais técnicas. Este fato deve-se ao uso exclusivo do agrupamento de pontos de interesse sem levar em consideração as outras características do peixe. O BoVW baseia-se apenas em pontos de interesse, e eles podem ser encontrados no peixe e/ou no fundo da imagem. Um exemplo de ruído encontrado em uma das imagens utilizadas nos experimentos é ilustrado na Figura 5.4, onde é possível identificar que grande parte dos pontos de interesse foram encontrados no fundo da imagem. Assim, grande parte das palavras visuais formadas são obtidas de pontos que não pertencem ao peixe, prejudicando a descrição. Mas o problema do fundo da imagem não tem grande impacto na descrição das imagens quando a cor é utilizada na descrição.

Dado um conjunto de imagens referentes a uma espécie de peixe pertencentes a determi-

nado aquário, o fundo das imagens tendem a possuir as mesmas cores, já que a iluminação e o aquário são os mesmos. Assim, em técnicas como o BoC onde as cores são agrupadas, é provável que uma cor seja formada correspondendo a média das cores do fundo da imagem. No caso do BoC, um núcleo representando uma assinatura de cor correspondente ao fundo. E como as imagens de treinamento são do mesmo aquário, a cor média tende a ter uma distribuição uniforme na descrição das imagens, tornando-se menos discriminativa na classificação.

O BoFC apresentou um bom resultado quando comparado às técnicas que utilizam cor na descrição das imagens. Por mais que o mesmo problema encontrado no uso de pontos de interesse seja encontrado no BoFC, a inserção de cor na descrição amenizou o problema. O fato dos pontos de interesse possuírem informações de cor em sua descrição diminuiu o impacto dos pontos encontrados no fundo da imagem, pelo mesmo motivo citado na análise do BoC e, devido a este motivo, o BoFC obteve melhores resultados do que o BoVW.

O fato do agrupamento das cores do fundo da imagem é evidenciado principalmente na técnica BoCW, onde, observando as tabelas correspondentes às Medidas-F na Seção 4, quanto maior a informação de cor, maior é a Medida-F. Inicialmente, para uma quantidade pequena de faixas, os resultados são próximos ou inferiores ao do BoVW devido a pouca informação de cor comparado com a mesma variação do tamanho dos dicionários (32, 64,...,8192). Mas aumentando a quantidade de faixas aumenta, os valores de Medida-F também aumentam, evidenciando a importância da cor na descrição.

Os histogramas de cores também evidenciam a importância da cor na classificação, assim como o agrupamento de cores do fundo da imagem, já que a divisão do espaço de cor em faixas é um modo de agrupamento. Contudo, a grande quantidade de atributos (16^3 para o Histograma de cores RGB e 32^3 para o HSV) é várias vezes superior a quantidade de atributos das técnicas baseados em Histograma de Palavras Visuais(BoVW), representa um problema em dispositivos móveis onde a quantidade de memória disponível para processamento é mais limitada do que em computadores.

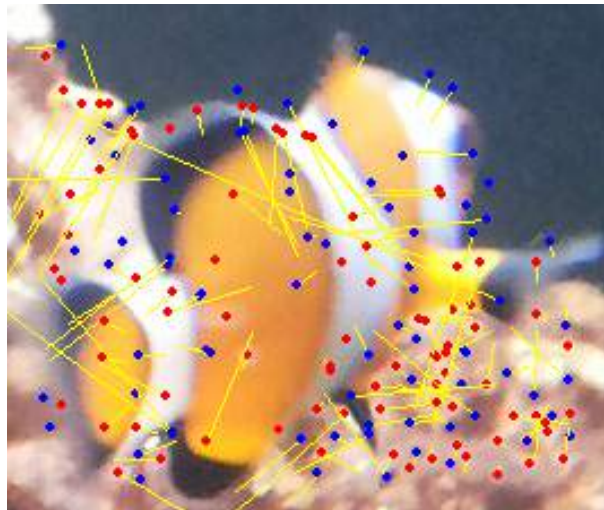


Figura 5.4: Exemplo de pontos de interesse encontrados em uma imagem do peixe Palhaço.

Aplicação de Reconhecimento de Espécies de Peixes - Pantanalbum

Utilizando o extrator de atributos que obteve melhor resultado na classificação das espécies de peixes, juntamente com o classificador que obteve melhor desempenho, uma aplicação Android foi desenvolvida implementando ambos com o objetivo de classificar imagens de peixes quanto a sua espécie. A aplicação consiste em um jogo em que o objetivo é completar um álbum de figurinhas criado a partir das imagens dos peixes. A principal interação do jogador é em tirar uma foto de um peixe, tentar predizer sua espécie, e receber duas possíveis respostas da aplicação: uma resposta positiva, indicando que o jogador acertou a espécie, ou uma resposta negativa indicando o erro.

Para manipular as imagens e aplicar os extratores de atributo necessários, foi utilizada a biblioteca de visão computacional de código livre OpenCV [23], em específico, uma versão para Android denominada OpenCV4Android. A implementação do classificador SVM, utilizado na classificação das espécies, é baseada no LibSVM [1] fornecido pelo OpenCV.

A extração dos atributos e treinamento do classificador foi feita em um computador. Posteriormente, o modelo do classificador e a técnica de extração de atributos foi migrada para a aplicação Android.

O público alvo da aplicação são crianças e portanto, o foco do aplicativo é na simplicidade e facilidade de aprendizado. Inicialmente o jogador inicia o jogo com um álbum vazio, tendo como dica apenas o nome do peixe e sua forma obtida de uma imagem do conjunto de treinamento. A Figura 6.1.2 ilustra o álbum vazio.

Quando um toque é detectado sobre a imagem do peixe, uma descrição do mesmo é mostrada para o usuário informando-o sobre várias curiosidades do animal, tais como: local de origem, tipo de alimentação, tamanho máximo, entre outras.

O botão capturar inicia a câmera utilizada para obter a imagem. A câmera possui um retângulo em seu centro indicando onde o peixe deve estar posicionado para a imagem ser capturada. O retângulo tem o objetivo de diminuir a área de fundo da imagem, minimizando assim os ruídos que possam ocorrer. Assim, cada peixe possui um botão capturar ao qual deixa implícito a suposta espécie do animal antes da classificação. Após a foto ser tirada, a mesma é descrita utilizando o extrator de atributos BoC formando um vetor de tamanho 2048



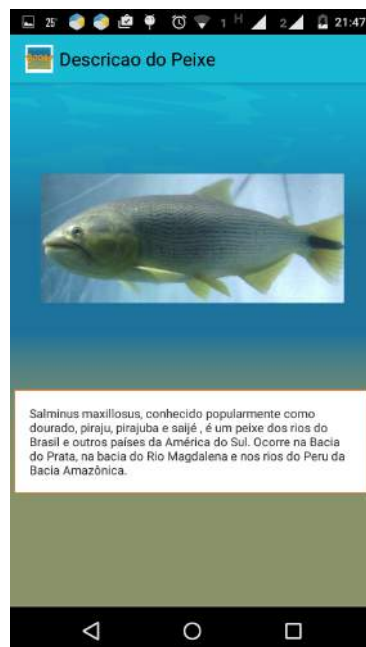
6.1.1:



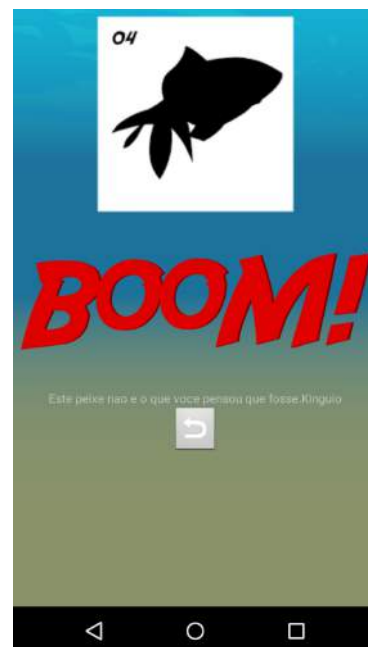
6.1.2:

Figura 6.1: Tela inicial e do álbum vazio da aplicação.

posições correspondendo as assinaturas de cores da imagem. A imagem é então classificada utilizando o classificador SVM com os parâmetros descritos na Seção 4 e caso haja acerto do usuário uma resposta positiva é retornada. A Figura 6.2 ilustra as mensagens obtidas da classificação.



6.2.1:



6.2.2:

Figura 6.2: Resposta de acerto ou erro de uma suposta espécie de Peixe. A Figura 6.2.1 ilustra o acerto da espécie Dourado. Já a Figura 6.2.2 ilustra um erro do jogador ao classificar a espécie Kinguio.

Caso o jogador acerte a espécie, uma imagem do peixe em cores substitui a contendo apenas a forma do animal em tons de cinza indicando a captura do animal. A Figura 6.3

ilustra o álbum após a captura do peixe Dourado. Após todos os peixes serem capturados é possível reiniciar o álbum para o jogo ser iniciado novamente.

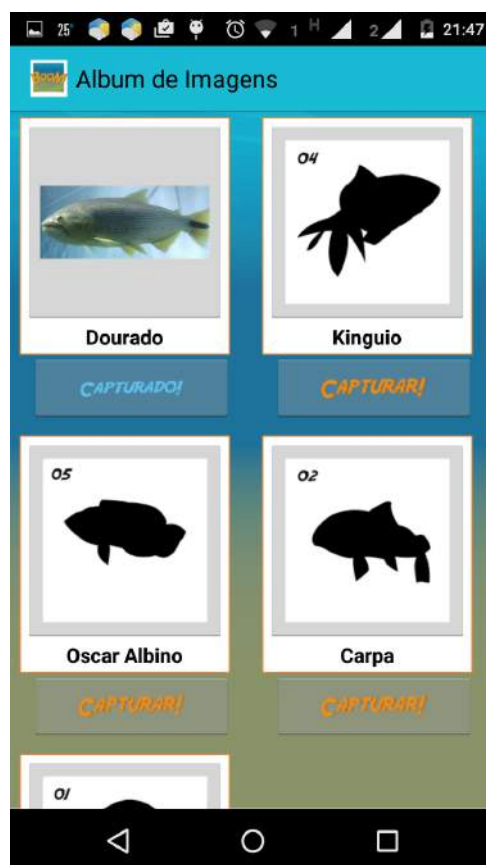


Figura 6.3: Imagem do álbum após a captura de um Dourado.

Conclusão

O uso de cores na classificação de imagens mostrou-se de grande valor para incrementar a precisão da classificação de imagens de peixes. Neste trabalho, as técnicas que utilizaram cores na descrição das imagens obtiveram resultados melhores quando comparadas a técnica que utiliza apenas imagens em tons de cinza. Comparando estatisticamente, não há diferença significativa entre as técnicas de histograma de cores e as técnicas baseadas em BoVW que utilizam cores, como BoC, BoFC e BoCW. Assim, pode-se escolher a técnica que melhor se adapte as necessidades de uma aplicação que os resultados serão semelhantes.

Em classificações onde apenas as cores descrevem as imagens, os resultados são satisfatórios, com valores de Medida-F 0.90, 0.80, 0.88, 0.89 e 0.88 para os extratores BoC, BoFC, BoCW, histograma de cor HSV e RGB, respectivamente. Mas a influência de uma fonte de iluminação diferente da encontrada no conjunto de treinamento, além dos ruídos como distorção da imagem causada pelo movimento da câmera, podem atrapalhar a classificação.

Observando a Tabela 7.1 que ilustra os valores absolutos da matriz de confusão do classificador SVM utilizando BoC, a mesma ilustrada na Figura 4.6, é possível notar que as espécies com maior número classificações erradas foram: Kinguio e Platy Sangue, ambas com 10 imagens classificadas incorretamente, seguido por Platy Laranja e Rubi com 8 imagens e kinguio-korraco com 7. Algumas das espécies que obtiveram maior quantidade de imagens classificadas de forma errada foram analisadas.

O Kingui foi confundido principalmente com o peixe Papagaio (4 imagens) e Paulistinha (3 imagens) devido as imagens possuírem um tom semelhante da cor de fundo e o laranja característico da espécie. A Figura 7.1.1 ilustra uma imagem da espécie Kinguio em que a maior parte da área dos pixels é coberta por uma cor de fundo clara e escura. A descrição desta imagem é composta, principalmente, por assinaturas de cores que pertencem ao fundo, o qual possui cores semelhantes a espécie Papagaio (Figura 7.1.2) e Paulistinha (Figura 7.1.3) confundindo o classificador. Este é um problema recorrente ao utilizar unicamente a cor na descrição das imagens. O problema deste exemplo está no modo como a imagem do peixe foi capturada. Na aplicação, é feito um quadrado na tela em que a câmera está ativada e o jogador é induzido a inserir a imagem do peixe dentro de um retângulo. A imagem é então recortada de modo que apenas o conteúdo do retângulo é analisado e classificado. Assim, as chances da imagem analisada conter, na maioria de sua extensão, apenas informações

Tabela 7.1: Matriz de confusão do BoC com dicionário de tamanho 2048 e classificador SVM.

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
0 acará-bandeira	37	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0
1 acará-bandeira-marmorizado	0	34	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	2	0	0	0	2	0	0	0	0
2 acará-disco	0	0	36	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0
3 barbus-ouro	0	0	0	38	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
4 barbus-sumatra	0	0	0	1	38	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
5 beta	1	0	0	0	0	37	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
6 carpa	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7 carpa-media	0	0	0	2	0	0	0	37	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
8 dourado	0	0	0	0	0	0	0	0	38	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0
9 kinguio	1	0	0	0	0	1	0	0	0	30	0	0	0	0	1	0	0	0	4	3	0	0	0	0	0	0	0	0
10 kinguio-cometa-calico	0	0	0	0	0	0	0	0	0	0	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
11 kinguio-korraco	0	1	0	0	0	0	0	3	0	2	0	33	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
12 mato-grosso	0	0	1	0	0	0	0	0	0	0	0	0	39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13 molínesia-preta	0	0	0	0	0	0	0	0	0	0	0	0	0	37	0	0	0	0	0	2	0	0	0	0	1	0	0	0
14 oscar	0	0	0	0	0	0	0	0	0	0	0	0	0	0	38	0	0	0	2	0	0	0	0	0	0	0	0	0
15 oscar-albino	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	39	0	0	0	0	0	0	0	0	0	0	0	0
16 pacu	0	1	0	0	0	0	1	0	2	0	0	0	0	1	0	1	34	0	0	0	0	0	0	0	0	0	0	0
17 palhaco	0	0	2	0	0	0	0	0	1	0	0	0	0	1	0	0	0	34	0	1	0	0	0	0	0	1	0	0
18 papagaio	0	0	1	2	0	0	0	1	0	1	0	1	0	0	1	0	0	1	31	1	0	0	0	0	0	0	0	0
19 paulistinha	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0	1	34	0	1	0	0	1	0	0	0
20 piau-tres-pintas	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	34	0	0	0	0	0	2	4
21 platy-laranja	0	0	2	0	0	2	0	0	0	2	0	0	0	0	0	0	0	1	1	0	0	32	0	0	0	0	0	0
22 platy-rubi	2	0	0	0	0	2	0	0	0	1	0	0	0	2	0	0	0	0	0	1	0	0	32	0	0	0	0	0
23 platy-sangue	0	1	0	1	0	0	0	0	1	0	0	3	0	1	0	0	0	0	2	0	0	1	30	0	0	0	0	0
24 telescópio	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	39	0	0	0
25 etra-negro	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	37	0	0
26 tricogaster	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	40	0
27 tucunare	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	37

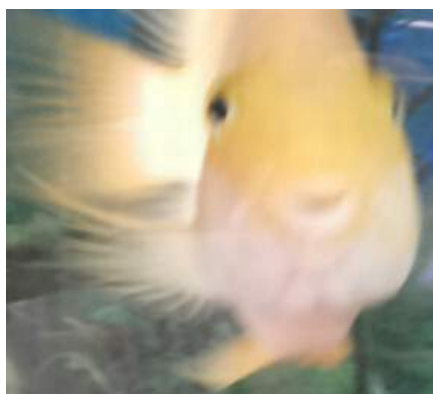
relativas ao peixe é maior.



7.1.1: Kingiuo



7.1.2: Paulistinha



7.1.3: Papagaio

Figura 7.1: Semelhança de cores nas imagens entre as espécies Kingiuo, Papagaio e Paulistinha.

Mesmo possuindo um tom de vermelho muito peculiar, há 3 imagens no conjunto de treinamento do Platy Sangue com o mesmo problema do Kingiuo, isto é, a maior parte dos pixels da imagem pertencem ao fundo, além da influência de uma luz tornando a imagem “amarelada” (Figura 7.2.1). As cores do fundo das 3 imagens formam assinaturas de cores semelhantes as do Kingiuo Korraco, sendo assim, estas três imagens foram classificadas como sendo Kingiuo Korraco mesmo pertencendo a espécie do Platy Sangue. A Figura 7.2.1 ilustra um exemplo de uma classificação errada. Note que em ambas as imagens serão formadas assinaturas em tom claro e amarelado.

O BoC mostrou-se um eficiente descritor de imagens de peixes, contudo, o fato de apenas cor ser utilizada, as imagens podem ser sensíveis a variações de iluminação levando ao classificador errar a predição da imagem, como mostrado no exemplo de classificação errado das 3 imagens do Platy Sangue. Mas, se a aplicação for destinada a um aquário específico onde não há uma variação intensa de iluminação, como em aquários em ambientes fechados com iluminação constante, um conjunto de imagens de treinamento provido deste aquário pode mostrar-se de grande eficácia na classificação de espécies de peixe utilizando o BoC. O ambiente de captura das imagens pelas pessoas no Aquário do Pantanal segue estas características. Assim, devido ao ambiente de captura das imagens ser coberto e sem grandes variações de iluminações, um novo conjunto de imagens contendo as espécies do aquário será obtido e descrito pelo BoC. Posteriormente um classificador SVM será treinado e inserido na aplicação Pantanalbun.



7.2.1: Platy Sangue



7.2.2: Kinguio Korraco

Figura 7.2: Semelhança de cores nas imagens entre as espécies Platy Sangue e Kinguio Korraco.

Referências Bibliográficas

- [1] Chang, Chih-Chung e Lin, Chih-Jen, ***LIBSVM: A Library for Support Vector Machines***, ACM Transactions on Intelligent Systems and Technology, vol. 2, p. 1-27, 2011 Citado nas páginas 22, 32, e 50.
- [2] L. Breiman, J. Friedman, R. Olshen e C. Stone, ***Classification and Regression Trees***, Chapman & Hall, New York, pg. 358, vol 1, 1984. Citado na página 21.
- [3] Hollander M, Wolfe D. A, ***Nonparametric Statistical Methods***, Wiley, Universidade de Michigan, 1973. Citado nas páginas 21 e 46.
- [4] Rova A., Mori G. e Dill L. M., ***One fish, two fish, Butterfish, Trumpeter: Recognizing Fish in Underwater Video***, Conference on Machine Vision Applications, p. 404-407, Tokyo Japão, 2007. Citado nas páginas 3 e 4.
- [5] Rodrigues, M. T. A., ***Classificação Automática de Espécies de Peixes Baseada em Técnicas Robustas para Extração de Características e Sistemas Imunológicos Artificiais***, Tese de Mestrado, Belo Horizonte : Centro Federal de Educacao Tecnologica de Minas Gerais, 2009. Citado na página 2.
- [6] Nery M. S. ,Machado A. M. ,Campos M. F. M., Pádua F. L. C., Carceroni R. e Queiroz-Neto J. P., ***Determining the Appropriate Feature Set for Fish Classification Tasks***, Graphics, Patterns and Images, SIBGRAPI Conference on, vol. 0, pg. 173-180, n. 1530-1834, 2005. Citado na página 2.
- [7] Bahri A. e Zouaki H., ***A SURF-Color Moments for Images Retrieval Based on Bag-Of-Features***, European Journal of Computer Science and Information Technology, vol. 1, pg. 11-22, n. 1, 2013. Citado nas páginas 4, 5, e 17.
- [8] Wengert C., Douze M., Jégou e H., ***Bag-of-Colors for Improved Image Search***, MM 2011 - 19th ACM International Conference on Multimedia, ACM, Scottsdale, United States, 2011. Citado nas páginas 3, 4, 5, e 19.
- [9] Lowe David G., ***Distinctive Image Features from Scale-Invariant Keypoints***, Int. J. Comput. Vision, vol. 60, pg. 91-110, n. 2, 2004. Citado nas páginas 3, 5, 7, e 17.
- [10] Bay H., Ess A., Tuytelaars T. e Van Gool L., ***Speeded-Up Robust Features (SURF)***, Comput. Vis. Image Underst., vol. 110, pg. 346-359, n. 3, 2008. Citado nas páginas 3, 5, 7, 10, e 17.
- [11] Quinlan J. Ross, ***C4.5: Programs for Machine Learning***, C4.5 - programs for machine learning / J. Ross Quinlan, Morgan Kaufmann, San Mateo CA, 1993. Citado na página 21.

- [12] Friedman Milton, **A Comparison of Alternative Tests of Significance for the Problem of m Rankings**, The Annals of Mathematical Statistics, vol. 11, pg. 86-92, n. 1, 1940. Citado nas páginas 21 e 46.
- [13] Jolliffe I.T., **Principal Component Analysis**, Springer Series in Statistics, Springer-Verlang, Universidade Cornell, 1986. Citado na página 2.
- [14] Vapnik V. N., **An Overview of Statistical Learning Theory**, Trans. Neur. Netw., vol. 10, pg. 988-999, n. 5, 1999. Citado na página 22.
- [15] M. K. Alsmadi, K. B. Omar, S. A. Noah, I. Almarashdeh, **Fish Recognition Based on the Combination Between Robust Feature Selection, Image Segmentation and Geometrical Parameter Techniques using Artificial Neural Network and Decision Tree**, International Journal of Computer Science and Information Security, vol. 6, pg. 215-221, n. 2, 2009. Citado na página 2.
- [16] Phoenix X. Huang, Bastiaan Johannes Boom, Robert B. Fishe, **ACCV 1 Lecture Notes in Computer Science**, Springer, 2012. Citado na página 2.
- [17] Hu, Jing e Li, Daoliang e Duan, Qingling e Han, Yueqi e Chen, Guifen e Si, Xiuli, **Fish species classification by color, texture and multi-class support vector machine using computer vision**, Comput. Electron. Agric., vol. 88, pg. 133-140, n. 8, 2012. Citado na página 3.
- [18] Koenderink, JanJ., **The Structure of Images**, Biological Cybernetics, Springer-Verlag, vol. 50, pg. 363-370, n. 5, 1984. Citado na página 10.
- [19] Forsyth, D. and Ponce, J., **Computer Vision: A Modern Approach**, Kluwer international series in engineering and computer science: Knowledge representation, learning, and expert systems, Springer, 2011, New York, NY, USA, 1986. Citado na página 1.
- [20] Forsyth, D. and Ponce, J., **Computer Vision: A Modern Approach**, Always learning, Pearson, 2012, vol. 2, p. 761, 2012. Citado na página 1.
- [21] Fawcett, Tom, **An Introduction to ROC Analysis**, Pattern Recogn. Lett., Elsevier Science Inc, New York, New York, United States, vol. 27, p. 861-874, n. 8, 2006. Citado na página 28.
- [22] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E., **Scikit-learn: Machine Learning in Python**, Journal of Machine Learning Research, vol. 12, 2825-2830, 2011 Citado nas páginas 22, 23, e 32.
- [23] Bradski, G., **Opencv Library**, Dr. Dobb's Journal of Software Tools, 2000 Citado na página 50.
- [24] Spackman, Kent A., **Signal Detection Theory: Valuable Tools for Evaluating Inductive Learning**, Proceedings of the Sixth International Workshop on Machine Learning, pg. 160-163 Morgan Kaufmann Publishers Inc. 1989 Citado na página 28.