

Classificação de Espécies de Peixe usando Inferência Gramatical no Reconhecimento de Padrões em Problemas de Visão Computacional

M. R. Borth¹, L. C. Ribas², H. Pistori³, W. N. Gonçalves⁴, A. A. C. Junior⁵

Abstract — Este artigo apresenta uma nova abordagem para representar uma imagem como uma sentença gerada a partir de um agrupamento baseado em estrutura hierárquica, de modo que técnicas de inferência gramatical possam ser usadas em problemas de visão computacional. Para a classificação de imagens, este trabalho inicialmente detecta e descreve pontos de interesse usando o algoritmo SIFT; em seguida, constrói o alfabeto com base no conjunto de descritores dos pontos de interesse, rotulando cada um deles; na sequência, realiza a geração de uma sentença por agrupamento hierárquico; e, por fim, realiza a aprendizagem da gramática para a classificação de imagens utilizando o algoritmo de inferência gramatical *k-testable*. Dois experimentos foram realizados utilizando dois novos bancos de imagens de peixes, criados pelos autores deste trabalho, a fim de explorar parâmetros da abordagem e ter um primeiro resultado sobre o desempenho da técnica. Os resultados são encorajadores para continuar e explorar novos caminhos na área de reconhecimento sintático de padrões.

Keywords — Geração de sentença; Reconhecimento sintático de padrões; Inferência gramatical; Agrupamento hierárquico; Classificação de imagens.

I. INTRODUÇÃO

Informações que tratam sobre reconhecimento, classificação, distribuição, quantificação e migração de espécie de peixes muitas vezes são críticas para ecologistas, estudiosos do meio ambiente, órgãos governamentais e empresas de pescada [1]. Nesse contexto, a criação de softwares para monitorar o ambiente de onde os peixes estão inseridos possibilita adotar alternativas ou ações para atuarem, desde o controle preventivo até a criação de novas soluções para controlar o crescimento saudável de peixes. Além disso, sistemas de classificação de espécies de peixes podem ajudar biólogos a resolver questões de disponibilidade de alimentos, relação entre predador e sua presa [1, 2, 3], inclusão de estudos

ecológicos e ambientais das comunidades de peixes [4], projeto e colocação de escadas de peixes em represas para geração de energia hidrelétrica [5], alimentação estratégica pelos piscicultores [6] e avaliação de ações para gestão das pescas [7]. Em geral, alguns diagnósticos são realizados por peritos, onde eles usam a própria visão para detectar determinadas características ou doenças. Entretanto essas pessoas são escassas, especialmente em áreas rurais. Cada espécie de peixe possui características específicas, tais como: tamanho de nadadeiras, padrões de cores, tamanho do espécime, etc. Na classificação de peixes, por exemplo, existem 47 características diferentes por espécie para que se consiga um bom nível de exatidão na identificação [8].

Diferente do que acontece no dia a dia do trabalho que envolve a classificação de peixes, o qual se baseia em procedimentos manuais de alto custo e baixa eficiência, este trabalho apresenta uma abordagem para a classificação automática de espécies de peixes que pode ser usada nas atividades relacionadas anteriormente. Este trabalho permite criar uma solução automatizada que ajude o trabalho manual executado por seres humanos, mediante o uso de um sistema computacional baseado em técnicas de visão computacional.

Alguns trabalhos propõem o desenvolvimento de um método para classificação de formas geométricas básicas por meio de autômatos adaptativos [9, 10]. Esses trabalhos utilizam a abordagem sintática de reconhecimento de padrões, apresentando algumas vantagens sobre práticas tradicionais permitindo realizar uma identificação de propriedades estruturais das formas contidas na imagem. A característica adaptativa aperfeiçoa o processo de reconhecimento ao considerar pequenas diferenças entre bordas, geradas por distorções residuais nas fases de segmentação e pré-processamento. Entretanto, a abordagem se restringe apenas a atributos relacionados a borda do objeto, diferentemente deste trabalho que pode utilizar informações de toda a parte da imagem. Ainda assim, durante o processo de aprendizagem da gramática, a geração dos autômatos utilizados para validar as sentenças pode ser realizada a partir do uso de técnicas e conceitos de adaptatividade.

Este trabalho apresenta uma nova forma de ordenação de pontos de interesse detectados em uma imagem a fim de produzir uma sentença. A sentença, também conhecida por *string*, palavra ou cadeia de caracteres, é a representação simplificada de uma sequência finita de palavras visuais ordenadas, por exemplo, “ABACDECBCADA”. A sentença é construída com base no tamanho de um alfabeto que é definido usando o algoritmo *k*-médias (*k-means*), como no

¹ Doutorando em Ciências Ambientais e Sustentabilidade Agropecuária pela Universidade Católica Dom Bosco (UCDB). Professor do Instituto Federal do Paraná (IFPR) – Campus Umuarama. E-mail: marceloborth@gmail.com

² Mestrando em Ciência da Computação e Matemática Computacional pela Universidade de São Paulo. E-mail: lucascorreiaribas@gmail.com

³ Doutor em Engenharia Elétrica pela Universidade de São Paulo. Professor da Universidade Católica Dom Bosco (UCDB). E-mail: pistori@ucdb.br

⁴ Doutor em Física pela Universidade de São Paulo. Professor da Universidade Federal de Mato Grosso do Sul (UFMS) – Campus Ponta Porã. E-mail: wnunesgoncalves@gmail.com

⁵ Doutor em Engenharia Elétrica pela Universidade de São Paulo. Professor da Universidade Federal de Mato Grosso do Sul (UFMS) – Campus Ponta Porã. E-mail: amaury.ufms@gmail.com

histograma de palavras visuais [11, 12]. O alfabeto, também conhecido como dicionário, é um conjunto finito de símbolos, números ou caracteres. Para gerar a sentença, a ordenação é baseada no agrupamento hierárquico do rótulo de cada ponto de interesse.

O trabalho propõe a representação de uma imagem como sentença, gerada a partir de agrupamento baseado em estrutura hierárquica. Para a classificação de imagens, inicialmente são detectadas e descritos pontos de interesse usando o algoritmo *Scale Invariant Feature Transform* – SIFT [13]. Esse algoritmo permite identificar e descrever pontos de interesse relevantes de imagens para realizar o reconhecimento de padrões. Em seguida, é construído o alfabeto com base no conjunto de descritores dos pontos de interesse, rotulando cada um deles. O alfabeto é definido usando o algoritmo *k*-médias (*k-means*), como no histograma de palavras visuais (*Bag of Visual Words*). Na sequência, uma sentença para cada imagem é gerada. A sentença possuirá tamanho *n*, tal que *n* é a quantidade de pontos de interesse detectados na imagem. A ordenação gerada é baseada na técnica de agrupamento hierárquico, unindo pontos de interesse próximos a partir da posição espacial (*x*, *y*) até formar uma árvore binária completa. Por fim, é realizada a aprendizagem da gramática para a classificação de imagens utilizando o algoritmo de inferência gramatical *k-testable* [14].

Dois experimentos foram realizados usando imagens de peixes, a fim de analisar a estratégia de ordenação dos pontos de interesse e o impacto do tamanho do alfabeto sobre o desempenho da classificação. Para esses experimentos, o algoritmo de detecção de pontos de interesse SIFT e o algoritmo de inferência gramatical *k-testable* foram utilizados. Entretanto, a técnica pode ser aplicada usando qualquer algoritmo de detecção de pontos de interesse e de inferência gramatical disponíveis.

A principal contribuição deste trabalho está na ordem empregada para a geração das sentenças, pois é baseada na técnica de agrupamento hierárquico, unindo pontos de interesse próximos entre si a partir da posição espacial (*x*, *y*) que possuem menor distância, até formar a estrutura de uma árvore binária completa. Também, contribui em testes com uma larga faixa de tamanho de alfabeto, em que o limite é a capacidade de memória do computador. Outra contribuição é a publicação de dois bancos de imagens de peixes. O aquario10e⁷ constituído de 596 imagens de peixes e o aquarioSeg10e⁸ com 100 imagens de peixes segmentadas manualmente. Ambos os bancos de imagens possuem imagens de 10 espécies de peixes e se destacam por conter apenas imagens capturadas por celulares, visto que os bancos de imagens atuais fornecem em sua maioria imagens de câmeras fotográficas e/ou em ambientes controlados.

A seguir é apresentada uma revisão teórica dos principais conceitos utilizados neste trabalho. Na sequência, são

apresentados os trabalhos relacionados, discutindo os problemas e limitações das abordagens atuais. Posteriormente, é apresentada a proposta deste trabalho, os experimentos realizados e discussões sobre os resultados. Por fim, são apresentadas as considerações finais e futuros direcionamentos.

II. REVISÃO TEÓRICA

Nesta seção, são apresentados os principais conceitos utilizados para a realização deste trabalho.

A. SIFT – SCALE INVARIANT FEATURE TRANSFORM

A extração e descrição de características em imagens têm sido extensivamente empregada na área de reconhecimento de imagens. Atualmente, o *Scale Invariant Feature Transform* – SIFT e o *Speeded Up Robust Features* – SURF [15] são os dois métodos mais populares. Esses métodos extraem da imagem uma coleção de vetores de características locais, chamados de pontos de interesse. Cada ponto de interesse é composto de uma parte pequena da imagem e a quantidade de pontos detectados pode variar de uma imagem para outra. Essas técnicas buscam ser invariantes a rotação, escala da imagem e mudança de iluminação. Esses extratores são extremamente relevantes para tarefas de reconhecimento de padrões. Sua implementação usa a função de diferença de Gaussianas. Ao serem identificados, cada ponto de interesse é representado por quatro elementos:

$$p_i = \{(x_i, y_i), \sigma_i, \theta_i, \varphi_i\}$$

onde,

- (x_i, y_i) , localização espacial na imagem;
- σ_i , escala em que ele foi detectado;
- θ_i , orientação predominante do gradiente; e,
- $\varphi_i \in \mathbb{R}^{128}$, descrição do ponto de interesse, representado por um vetor de características contendo 128 valores que descrevem a região ao redor do ponto.

B. HISTOGRAMA DE PALAVRAS VISUAIS

O histograma de palavras visuais ou, simplesmente, *Bag of Visual Words* - BOVW, é um vetor de atributos extraídos de uma imagem ou conjunto de imagens. Para a extração desses atributos é necessário o uso de um algoritmo auxiliar, como o SIFT. Assim, o BOVW pode usar os descritores extraídos pelo SIFT para construir o alfabeto e associá-los aos pontos de interesse. Após a construção do alfabeto, os pontos de interesse podem ser referenciados como palavras visuais, pois cada um deles terá uma palavra visual associada a ele que é baseada na sua descrição.

O histograma de palavras visuais é uma técnica bastante utilizada em tarefas de classificação. Basicamente, pode acontecer dela contar a quantidade de ocorrências das palavras visuais de uma determinada sentença e fazer a distribuição de frequência dessas palavras, gerando um histograma. Assim, para determinar a classe de uma nova imagem cria-se um

⁷Disponível em: <https://www.dropbox.com/sh/crql4nug14juc4/AAAY-1ul8YBQ5AtXA7M1fz1wa>

⁸Disponível em: <https://www.dropbox.com/sh/50u4z5o197bstqk/AACDF64uxVOFKTyti7smo1G-a>

histograma de palavras visuais para uma nova imagem, o qual é comparado com histogramas das imagens de treinamento. A classe determinada será a do histograma mais semelhante com as imagens de treinamento, pois imagens de uma mesma classe possui maior ocorrência de certas palavras visuais.

Apesar dos avanços recentes e dos resultados promissores, o poder descritivo dessa técnica é limitado, pois descarta informações espaciais e estruturais das palavras visuais na imagem. Essas informações podem ser características importantes em tarefas de classificação de imagens, pois quando apenas são contadas as ocorrências de uma palavra visual em uma imagem, não é considerada sua localização espacial na imagem e nem seu posicionamento com as demais palavras visuais. Logo, para algumas técnicas de classificação de imagens é importante saber o posicionamento dos pontos de interesse na imagem e manter a relação espacial dos pontos na imagem como, por exemplo, no reconhecimento sintático de padrões. Essa técnica, por sua vez, pode gerar diferentes ordens na formação de sentença a partir dos pontos de interesse detectados, o que pode influenciar no resultado da classificação.

C. K-TESTABLE

O *K-Testable* é uma técnica que tenta encontrar um autômato para reconhecer uma linguagem a partir de um tamanho de memória k . Uma linguagem *K-Testable* é uma subclasse das linguagens regulares que realiza a análise de qualquer sequência usando memória fixa de tamanho k . Esse algoritmo é capaz de inferir linguagens em tempo polinomial por meio de prefixos, sufixos e partes de sentenças que ocorrem nos dados de treinamento [16]. A principal característica é que cada caractere é dependente apenas dos $k-1$ caracteres anteriores.

D. AGRUPAMENTO HIERÁRQUICO

O agrupamento (*clustering*) é o estudo de métodos para agrupar objetos de acordo com características similares entre si, como a distância entre dois pontos. Essa é uma tarefa que separa grupos similares, com o objetivo de melhor discriminar objetos pertencentes a classes diferentes. Por exemplo, dada uma representação de N objetos, encontram-se k grupos baseando-se em uma medida de similaridade, como a distância euclidiana, tal que objetos presentes em um mesmo grupo são considerados semelhantes e objetos de grupos diferentes são considerados distintos. A similaridade indica o nível de proximidade entre dois objetos em um conjunto de dados. Existem várias aplicações que utilizam técnicas de agrupamento, por exemplo, segmentação de imagens [17, 18, 19], agrupamento de documentos [20, 21, 22], estudo de dados de genoma [23], dentre outras.

Uma técnica de agrupamento hierárquico organiza os dados em uma hierarquia de grupos. Essa solução pode ter maior vantagem em relação às abordagens planas, uma vez que divide os pontos de interesse em vários níveis de especificidade e diferentes granularidades [24]. Nesse caso, a técnica pode iniciar a tarefa considerando cada ponto de interesse como um grupo distinto e, prosseguir sucessivamente combinando pontos de interesse mais similares, até que todos

os pontos sejam alocados a um único grupo. A árvore hierárquica gerada não é apenas um conjunto de grupos, mas uma hierarquia com vários níveis, formando uma árvore binária, onde os grupos de um nível são unidos com grupos do próximo nível. Isso possibilita decidir o nível de agrupamento mais adequado para cada tipo de aplicação ou como será feita a separação dos grupos baseado no comportamento da árvore hierárquica. Os vários grupos podem ser úteis em aplicações porque os pontos de interesse de cada grupo podem representar graus diferentes de similaridade.

O processo de agrupamento possui 5 fases [25]: pré-processamento, seleção da medida para o cálculo de similaridade, agrupamento dos dados, avaliação dos resultados e interpretação dos resultados. Após o cálculo de similaridade entre os objetos, na fase de agrupamento é aplicado um algoritmo de agrupamento dos objetos como, o k -médias, onde k é o número de grupos (*clusters*). Basicamente, o centro de cada grupo é definido como a média de todos os objetos que pertencem ao grupo. Assim, ao aplicar o método k -médias os objetos são separados em vários grupos baseado no número k , conforme pode ser visto na Figura 1 (a). Nessa figura, k é 3, logo são formados 3 grupos diferentes. Por outro lado, existe o agrupamento hierárquico que permite maior flexibilidade em um agrupamento, conforme ilustrado na Figura 1 (b). É possível ver a diferença entre ambas as técnicas, tal que, pelo agrupamento hierárquico é possível agrupar informações que pelo k -médias não seriam possíveis, pois o k -médias impõe a restrição de ter grupos esféricos. Detalhes e outros trabalhos sobre k -médias podem ser vistos em [26].

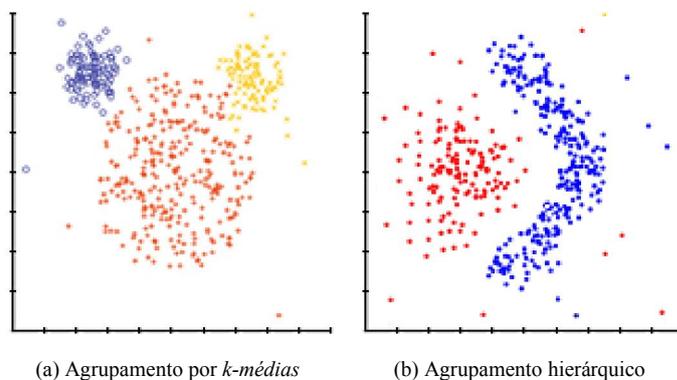


Figura 1. (a) Exemplo de agrupamento pelo algoritmo k -médias e (b) agrupamento hierárquico.

III. TRABALHOS CORRELATOS

Na visão computacional é comum usar o BOVW para classificação de imagens. Com essa técnica, cada imagem é tratada similarmente como um documento na classificação de textos, em que se obtém a frequência das palavras em um documento, gerando um histograma de palavras baseado na sua repetição. Para a classificação de imagens é necessário seguir algumas etapas: extrair suas características; descrevê-las; gerar o histograma de palavras visuais; aplicar um método de agrupamento nos descritores dos pontos de interesse detectados, que geralmente é o k -médias; e, por fim, realizar a classificação usando um algoritmo como, J48 (C4.5), *k-Nearest Neighbors* (KNN) e *Support Vectors Machine* –

Sequential Minimal Optimization (SMO). Embora grandes avanços tenham sido publicados nos últimos anos, uma das limitações dos histogramas é desconsiderar totalmente a relação espacial/estrutural dos pontos de interesse na imagem, o que acarreta na perda de informação da localização dos pontos de interesse.

Para superar a limitação do BOVW na classificação de imagens e permitir o uso de informação espacial dos pontos de interesse, alguns trabalhos foram propostos [27, 28, 29, 30, 31, 32]. Em geral, após a identificação dos pontos de interesse são geradas as palavras visuais com base em um tamanho de alfabeto utilizando o BOVW. Assim, é comum na área de reconhecimento sintático de padrões, criar uma sentença com uma determinada ordem a partir das palavras visuais extraídas da imagem. Dentre os trabalhos citados [27] propõem o uso de inferência gramatical, trazendo resultados interessantes no problema de classificação de imagens. O principal problema no uso de gramáticas em imagens está na representação da imagem a partir de uma sentença que represente suas características relevantes sem ignorar a relação espacial das palavras visuais na imagem [27]. Nesse trabalho, os autores apresentam uma nova estratégia para representar uma imagem através de uma sentença, de modo que algoritmos sejam usados no reconhecimento de padrões em imagens. A proposta explora o padrão textual da sentença e estratégias que visam preservar a informação espacial dos pontos de interesse e testa o tamanho do alfabeto e o tipo de ordenação dos pontos de interesse para a geração da gramática.

As abordagens de ordenação de palavras visuais para a formação da sentença propostas por [27] possuem limitações quanto à sua ordenação. A ordem randômica (cria a sentença obtendo as palavras visuais de forma aleatória), radial (inicia pelo ponto de interesse central da imagem e, segue sucessivamente concatenando os pontos mais próximos usando a distância euclidiana), ordem de leitura (concatena os pontos de interesse iniciando de cima para baixo e da esquerda para a direita), etc., mantém a estrutura espacial da imagem, porém não mantém pontos de interesse mais próximos um do outro juntos na sentença, o que pode prejudicar o processo de classificação. Logo, se as características da imagem não estiverem na ordem correta na sentença, o desempenho na classificação será prejudicado. Nas 4 imagens de peixe da Figura 2, é possível perceber que a orientação espacial se mantém, porém a sentença gerada para o mesmo peixe não é invariante a rotação, pois a ordem das palavras visuais é diferente para cada uma das quatro sentenças geradas. Logo, a taxa de classificação correta pode ser prejudicada se aplicada uma abordagem que mantém a orientação espacial e não é invariante a rotação para uma mesma imagem.

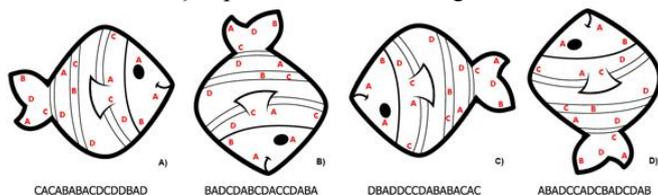


Figura 2. Quatro imagens de peixe com variação de 90° na rotação de uma para outra. Abaixo a sentença gerada de cada imagem a partir da ordenação dos pontos de interesse pela ordem de leitura ocidental (iniciando da parte superior e concatenando da esquerda para a direita).

Outro problema é destacado na Figura 3, pois quando realizada a ordenação dos pontos de interesse pelo método radial a geração da sentença é invariante a rotação mas não a translação. Mesmo sendo invariante a rotação, não há qualquer relação de um ponto de interesse anterior ou posterior na sentença gerada, o que pode prejudicar a classificação. Por exemplo, a palavra visual P_x não tem qualquer relação com a palavra visual P_{x-1} ou P_{x+1} na sentença. Dessa forma, uma palavra visual do olho do peixe pode ser a próxima de uma que representa a cauda e, assim, sucessivamente, misturando várias palavras visuais que podem ser totalmente distintas e sem proximidade uma da outra na imagem.



Figura 3. Distribuição das palavras visuais e marcação radial em uma imagem de peixe. Sentença resultante pela ordenação radial: HFFMGAABNMBKJ.

No processamento sintático de padrões, a ordem das palavras visuais pode caracterizar a importância ou o quanto cada palavra visual está conectada ou tem relação uma com a outra. Nesse contexto, a ordem das palavras visuais em uma sentença pode ser levada em consideração para identificar características semelhantes entre imagens, situação que não ocorre utilizando o *Bag of Words*, pois a informação estrutural é perdida ou, até mesmo, pelas abordagens mais recentes citadas nos trabalhos relacionados, pois não são invariantes a rotação ou não mantém um agrupamento dos pontos de interesse semelhantes.

No processo de formação da sentença, entendemos que é preciso manter a relação espacial entre as palavras visuais encontradas na imagem e ser invariante a rotação e translação, pois é importante manter uma estrutura organizacional das características de uma imagem. Logo, é preciso ter uma alternativa para organizar a informação estrutural da imagem como, por exemplo, a taxonomia. Essa é uma classificação de entidades de informação em forma de hierarquia de acordo com relacionamentos estabelecidos com entidades do mundo real que a representam [33]. As taxonomias possuem relacionamentos de generalização e especialização, e podem ser utilizadas para classificar informação de forma hierárquica. É em cima de toda a problemática apresentada nessa seção que a abordagem proposta tem sua maior contribuição, conforme detalhada a seguir.

IV. ABORDAGEM PROPOSTA

Neste trabalho é proposta uma nova forma de ordenação de palavras visuais em imagens baseada na técnica de agrupamento hierárquico, com objetivo de manter a estrutura das palavras visuais em uma sentença. A ideia principal da proposta é impor uma ordem para os n pontos de interesse e produzir uma sequência de tamanho n . O classificador pode ser dividido em 5 etapas: 1) extração de pontos de interesse; 2) construção do vocabulário; 3) rotulação de pontos de interesse; 4) geração da sentença; e, 5) aprendizagem da gramática.

A. EXTRAÇÃO DE PONTOS DE INTERESSE

Primeiramente, são detectados os pontos de interesse nas imagens e gerado um vetor de descrição para cada um deles usando o SIFT.

B. CONSTRUÇÃO DO ALFABETO

Este trabalho apresenta duas abordagens para a construção do alfabeto, não supervisionada e supervisionada, conforme usado por [34]. Na abordagem não-supervisionada, o conjunto de descritores é composto por todos os descritores de todas as imagens. Assim, o alfabeto é construído desconsiderando a classe de cada ponto de interesse, como se todas as classes fossem a mesma, deixando o descritor responsável para determinar a que grupo do alfabeto cada ponto de interesse pertencerá. A construção do alfabeto é realizada com base no conjunto de descritores D dos pontos de interesse, dado por:

$$D = [\varphi_i^j], 1 \leq i \leq M_j, 1 \leq j \leq N$$

onde, N é o número de imagens de treinamento e M_j é o número de pontos extraídos da imagem j .

O conjunto de descritores são agrupados usando o algoritmo k-médias e o conjunto de k centroides C é obtido a partir de:

$$C = k\text{-médias}(D)$$

O conjunto de centroides C é descrito como o alfabeto aprendido para as imagens de treinamento. Cada centroide possui a mesma dimensão dos descritores dos pontos de interesse, em que $C \in R^{128}$.

Na abordagem supervisionada, por sua vez, o alfabeto é construído separadamente para cada classe. Logo, a quantidade distinta de palavras visuais possíveis P_V de um alfabeto para cada classe é dada por:

$$P_V = T / n_c$$

onde, T é o tamanho do alfabeto e n_c é o número de classes. Os centroides C_p , são obtidos usando o algoritmo k-médias, aplicado apenas em um conjunto D_p contendo somente os descritores das imagens pertencentes a classe p :

$$D_p = [\varphi_i^j], \text{ se } j \text{ pertence à classe } p$$

$$C_p = k\text{-médias}(D_p)$$

Considerando n_c como o número total de classes, n_c conjuntos de centroides serão criados. Assim, para a construção do alfabeto supervisionado, os n_c conjuntos de centroides são concatenados. A vantagem da construção do alfabeto supervisionado em relação ao não-supervisionado é que o aprendizado das sentenças será obtido para cada classe separadamente, o que não acontece na construção não-supervisionada. Logo, com a abordagem supervisionada, cada classe pode gerar centroides mais característicos baseando-se nos seus próprios descritores.

C. ROTULAÇÃO DOS PONTOS DE INTERESSE

A rotulação de cada ponto de interesse das imagens de treinamento é feita na sequência. Dados cada ponto de interesse p_i^j para a imagem j , cada um deles é rotulado com o índice da palavra mais próxima do alfabeto, ou seja, o centroide mais próximo conforme apresentado abaixo:

$$r_i^j = \arg \min_{l=1}^k |\varphi_i^j, C_l|$$

onde r_i^j corresponde ao rótulo do ponto de interesse i da imagem j e $|\varphi_i^j, C_l|$ é a distância euclidiana.

Após essa etapa, cada ponto de interesse possuirá uma palavra visual, ou rótulo, de forma que pontos de interesse com descritores semelhantes possuirão as mesmas palavras visuais. Um exemplo pode ser visto na Figura 3, onde as palavras visuais são indicadas por letras do alfabeto

D. GERAÇÃO DA SENTENÇA

A ordenação baseada em agrupamento hierárquico foi usada para gerar a sentença a partir das palavras visuais da imagem. Essa é uma técnica invariante a rotação e translação. O algoritmo de agrupamento hierárquico segue 2 passos:

1. Encontrar a similaridade entre cada par de pontos de interesse no conjunto de dados: nessa etapa, é calculada a distância entre ponto de interesse usando a distância Euclidiana; e,
2. Agrupar os objetos em uma árvore hierárquica binária: nessa etapa, são relacionados pares de pontos de interesse que estão em proximidade usando o algoritmo *linkage*, implementado no Matlab. Uma vez que a proximidade entre os objetos no conjunto de dados já foi calculada, é possível determinar como os objetos no conjunto de dados devem ser agrupados. Essa função usa a informação de distância gerada na etapa anterior a fim de determinar a proximidade dos objetos entre si. Assim, o *linkage* pega as distâncias e relaciona pares de objetos que estejam juntos em clusters (dois a dois). Em seguida, relaciona esses grupos recém-formados para si e para outros objetos para criar grupos maiores até que todos os objetos do conjunto de dados estejam relacionados entre si, formando uma árvore hierárquica.

Nessa ordenação, as palavras são extraídas da imagem de acordo com a distância euclidiana da posição espacial de um ponto a outro, onde as palavras visuais mais próximas entre si são agrupadas para formar a sentença. Também é usada a ordenação radial como forma de comparação, onde a sentença é gerada de acordo com o valor da distância euclidiana de cada palavra visual ao centro da imagem. Assim, as palavras visuais mais próximas ao centro são adicionadas primeiro na sentença, repetindo, sucessivamente, até a palavra visual mais distante do centro.

A Figura 4 ilustra pontos de interesse extraídos de uma imagem e a árvore hierárquica gerada baseada na construção da sentença desta abordagem. As palavras visuais agrupadas mais ao topo da hierarquia representa maior distância em relação as demais palavras visuais que possuem um agrupamento mais abaixo da hierarquia. Quanto mais alto for o agrupamento da palavra visual maior dissimilaridade terá com os grupos mais baixo, conforme ilustrado à direita, a escala de similaridade. Logo, os nós próximos das folhas representam palavras visuais que possuem maior similaridade entre si e, conseqüentemente, ficarão juntas na geração da sentença.

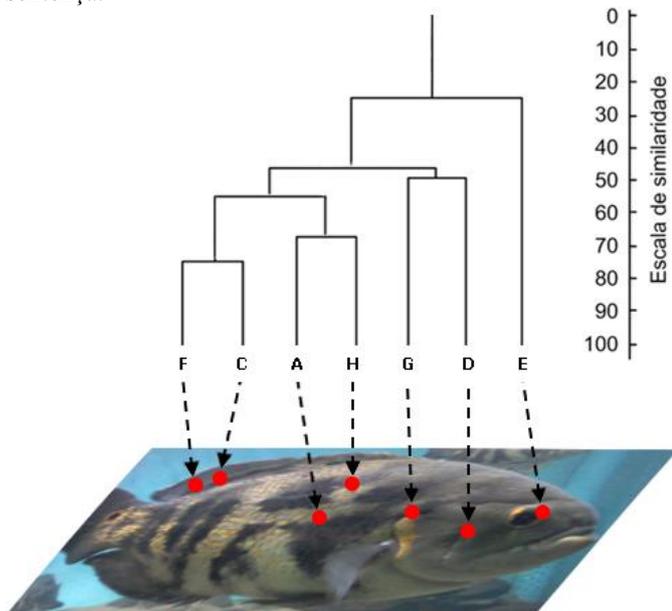


Figura 4. Ordenação dos pontos de interesse na imagem do peixe a partir do agrupamento hierárquico representado por um dendograma. A sentença gerada para essa imagem seria “F C A H G D E”.

E. APRENDIZAGEM DA GRAMÁTICA

A última etapa da abordagem é inferir uma gramática, ou regras, para cada classe a partir das sentenças geradas, resultando um autômato para cada classe. Para a inferência, o algoritmo *K-Testable* é utilizado. O *K-Testable* é uma técnica que dado um tamanho de memória k_t , tenta-se encontrar um autômato para reconhecer uma linguagem representada por um número de cadeias passadas como parâmetro. Uma linguagem *k*-testável é uma subclasse de uma linguagem regular que encontra prefixos, sufixos e sub-cadeias nos dados de treinamento [27]. A principal característica é que cada caractere é dependente apenas dos k_t-1 caracteres anteriores e

a análise de uma cadeia de caracteres pode ser feita usando uma memória de tamanho fixo k_t [27].

Nesse momento, as sentenças que representam as imagens de teste são validadas nos autômatos gerados. Para validar uma sentença é contado o número de erros para cada autômato. Os erros ocorrem quando há um caractere na sentença que não pertence a linguagem ou não existe a transição na linguagem de um autômato. O autômato que retornar o menor número de erros é a que pertence a sentença testada.

O tratamento de erros é um tópico muito importante na área de reconhecimento sintático, uma vez que pode conter erros ao validar uma sentença em uma gramática. Existem duas fases para o tratamento de erros: i) a detecção do erro e, ii) o tratamento ou recuperação do erro. Apenas detectar o erro não torna o analisador eficiente, uma vez que ele deve se recuperar do erro. Como os analisadores sintáticos analisam a gramática de linguagens, pode ser que símbolos esperados não sejam encontrados.

Um problema recorrente na área da teoria da computação é encontrar o equilíbrio entre expressividade e usabilidade na recuperação de erros, uma vez que máquinas de Turing, por exemplo, são formalismos muito expressivos, o que torna desconfortável o seu uso em problemas reais. Por outro lado, as máquinas de estados finitos são fáceis de usar, porém são mais restritas, o que também prejudica sua aplicação em problemas reais. Uma alternativa para isso é usar técnicas adaptativas [35, 36].

Os algoritmos adaptativos aumentam a expressividade de um formalismo sem afetar a usabilidade. Assim, é possível ser adaptativo sem modificar a sintaxe e a semântica original. Um algoritmo alternativo que poderia ser usado para esse tipo de problema na recuperação de erros é o AdapTree [37]. Ele é um algoritmo para indução de árvores de decisão que usa teoria de autômatos e enriquecido com tecnologia adaptativa. Uma das principais características deste algoritmo é a utilização conjunta de estratégias sintáticas e estatísticas.

V. EXPERIMENTOS

Dois experimentos foram realizados utilizando imagens de peixes de 10 espécies diferentes. O primeiro experimento traz imagens do banco de imagens aquario10e. Ao total são 596 imagens coloridas divididas em 10 classes. O segundo banco de imagens chamado aquarioSeg10e é composto de 10 imagens segmentadas (recortadas manualmente) escolhidas aleatoriamente de cada espécie do banco aquario10e, totalizando 100 imagens, 10 de cada espécie. A Figura 5 apresenta as espécies do banco aquario10e e a Figura 6 apresenta as espécies do banco aquarioSeg10e. As imagens variam umas das outras em tamanho e resolução, pois elas foram capturadas de diferentes celulares e posteriormente recortadas para deixar apenas o peixe na imagem. A distância da câmera ao peixe foi variável, de 20 centímetros a 1 metro, com iluminação natural do ambiente. Por isso, interferências como alteração na cor, iluminação desigual, reflexo, fundo diferente, sedimentos na água, plantas subaquáticas, etc., podem ser encontradas em algumas imagens. A principal diferença de um banco de imagens para outro é referente ao

primeiro banco ter o fundo em cada imagem, o que dificulta a aprendizagem para a classificação por conta de ruídos. Nesse caso, conseguimos analisar o quanto o fundo da imagem interfere na classificação da abordagem proposta. O nome e a quantidade de imagens de cada espécie podem ser obtidos na Tabela 1.



Figura 5. Imagens das espécies de peixes do banco de imagens aquario10e.



Figura 6. Imagens das espécies de peixes do banco de imagens aquarioSeg10e.

A construção do alfabeto foi realizada de duas formas, supervisionada e não-supervisionada com o tamanho do alfabeto variando entre 10 e 3000. O parâmetro do *K-Testable* utilizado nos experimentos foi $k_t = 2$, uma vez que usando valores maiores a geração da gramática não melhorou o resultado do classificador. Para a divisão do conjunto de dados de treinamento e teste foi usado o modelo de validação cruzada com 10 dobras. A Medida-F (*F-Measure*), também conhecida como *F-Score*, é adotada como parâmetro de comparação entre as abordagens testadas. Essa métrica é a média harmônica da precisão e revocação.

Espécie/Classe	Total de imagens
1. Acará bandeira (<i>Pterophyllum scalare</i>)	60
2. Dourado (<i>Salminus brasiliensis</i>)	80
3. Kinguio (<i>Carassius auratus</i>)	41
4. Molinésia Preta (<i>Poecilia Shenops</i>)	44
5. Oscar (<i>Astronotus ocellatus</i>)	101
6. Pacu (<i>Piaractus mesopotamicus</i>)	49
7. Piau 3 Pintas (<i>Leporinus friderici</i>)	75
8. Platy Sangue (<i>Xiphophorus maculatus</i>)	43
9. Tricogaster Azul (<i>Trichogaster trichopterus</i>)	41
10. Tucunaré (<i>Cichla piquiti</i>)	62

Tabela 1. Número de imagens para cada espécie de peixe do banco de imagens aquario10e, totalizando 596 imagens.

VI. RESULTADOS E DISCUSSÃO

A Tabela 2 apresenta os resultados dos experimentos para ambos os bancos de imagens, na forma supervisionada e não-supervisionada, com diferentes tamanhos de alfabeto e dois tipos de geração de sentença, radial e hierárquica. A construção do vocabulário supervisionado forneceu resultados bem superiores quando comparado com a construção do vocabulário não-supervisionado. Esse resultado demonstra que a obtenção de palavras individuais para cada categoria separadamente é importante no processo de descrição de imagens. A Medida-F mais alta obtida em todas as configurações foi para o banco de imagens aquarioSeg10e, na abordagem supervisionada, ordenação hierárquica e com tamanho do alfabeto de 2400.

A Figura 7 ilustra um gráfico baseado nos resultados desse banco de imagens. O gráfico compara os dois processos de construção do alfabeto, pela forma supervisionada e não-supervisionada, mostrando a Medida-F para cada tamanho do alfabeto. No gráfico podemos observar que a abordagem supervisionada é superior em praticamente todos os tamanhos de alfabetos testados.

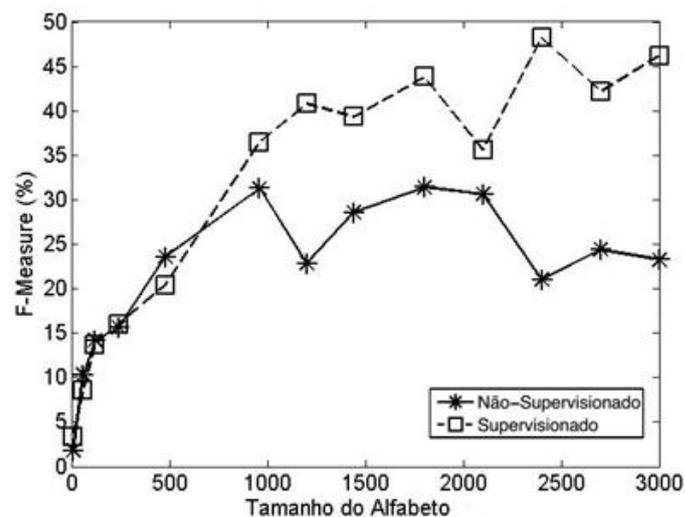


Figura 7. Gráfico comparativo das Medidas-F para a abordagem supervisionada e não supervisionada observando o tamanho de cada alfabeto com o banco de imagens aquarioSeg10e e ordenação hierárquica.

Em geral, percebe-se que o desempenho do classificador melhora conforme se aumenta o tamanho do alfabeto, pois o tamanho está diretamente relacionado ao poder descritivo das características dos pontos de interesse nas imagens. Logicamente, existe um limite para atingir sua descrição máxima que, nesse caso, foi encontrado no alfabeto de tamanho 2400, pois com um dicionário maior a classificação reduziu a Medida-F. As melhores Medidas-F, para ambos os bancos de imagem, foram obtidas usando a geração de sentença por agrupamento hierárquico e tamanho de dicionário de 2400.

O resultado mais satisfatório na classificação de espécie de peixes foi para o banco de imagens aquarioSeg10e, pois nesse caso, ao aplicar a extração de pontos de interesse na imagem, os pontos são encontrados somente na imagem do peixe, ao contrário do que ocorre com o banco de imagens aquario10e, pois além dos pontos de interesse detectados no peixe são

detectados inúmeros pontos no fundo da imagem, conforme ilustrado na Figura 8. Nesse caso, os pontos de interesse detectados no fundo da imagem são ruídos, reduzindo o desempenho do classificador. Em várias imagens, a maioria dos pontos de interesse são detectados no fundo da imagem e não no peixe.

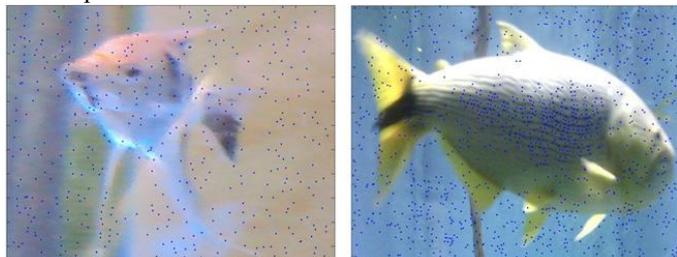


Figura 8. Exemplo de pontos de interesse detectados em imagens de duas espécies de peixe em que a maioria são ruídos

A Figura 9 apresenta a matriz de confusão para o melhor resultado obtido no experimento com a medida-F de 48,19. Percebe-se que a taxa de classificação correta se sobressai em dois pontos, na diagonal principal e na classificação da espécie Tricogaster, coluna vertical formada na matriz de confusão. Pela diagonal principal pode ser destacada a excelente taxa de recuperação da espécie Tricogaster de 100%, seguido pela Dourado com 90% e Molínesia Preta com 80%. Por outro lado, a espécie Tricogaster foi a que teve mais Falso-Negativo na classificação. É possível observar também que as espécies que tiveram as menores taxas de classificação correta foram a Kingiuo e a Platy Sangue, ambas obtiveram classificação correta de apenas 10% de suas imagens, o que levou para baixo a Medida-F geral.

A hipótese de classificação errada das espécies conforme mostrada na matriz de confusão pode estar na complexidade da base de imagens, pois as classes possuem características muitas similares entre si. Além disso, pode ser justificada pela quantidade de pontos de interesse detectados pelo algoritmo de extração, pois quanto mais pontos de interesse analisados de uma classe, melhor será sua aprendizagem e, conseqüentemente, mais imagens poderão ser classificadas devido a ter mais descrição da classe como um todo, podendo abranger até características de outras classes.

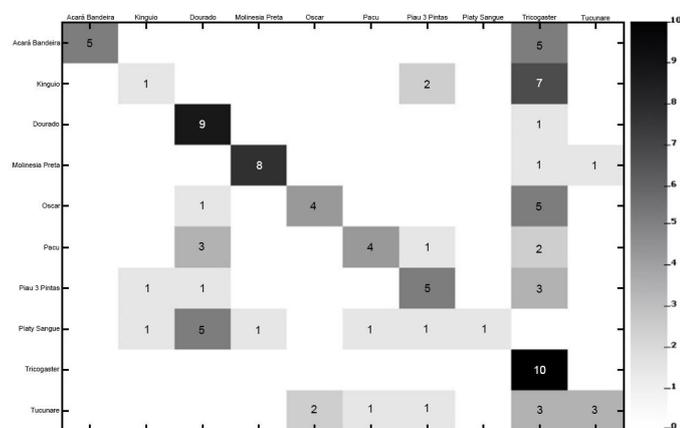


Figura 9. Matriz de confusão mostrando o número de imagens classificadas corretamente para cada espécie de peixe para o resultado de maior Medida-F

(banco de imagens aquarioSeg10e, abordagem supervisionada, geração da sentença pela ordenação hierárquica e tamanho de alfabeto 2400).

Essa discrepância de quantidade de informação para cada classe resultou em uma diferença grande de classificação de uma espécie para outra. Ao visualizar a matriz de confusão da classificação é possível notar a predominância da classe Tricogaster na classificação. Nota-se por meio da exibição da quantidade de imagens classificadas apresentadas na vertical. Essa classe possui a maior quantidade de pontos de interesse detectados, uma vez que possui 21,5% do total de pontos identificados. Conseqüentemente, para essa espécie foram classificadas 37% de todas as imagens. Logo, a espécie Tricogaster foi a classe que mais obteve extração de pontos de interesse nas imagens que, conseqüentemente, teve maior descrição de características da classe. Como resultado, obteve 100% de precisão na classificação, porém resultou em uma grande quantidade de Falso-Negativo, prejudicando a precisão de outras espécies.

Em geral, as classes de maior poder descritivo classificaram mais imagens, pois abrangeu e descreveu mais características quando comparadas às classes com menos pontos de interesse detectados. Em [38] já afirmavam em seus trabalhos que a capacidade de reconhecer padrões de uma imagem sobre um conjunto de imagens depende da quantidade de informações que se conhece a priori do objeto em questão.

O processo de reconhecimento não é uma tarefa trivial, pois os peixes podem aparecer nas imagens em diferentes tamanhos, formas, escalas, orientações, distorções, ruídos, cores diferentes (mesmo para uma mesma espécie) e diferente contexto de fundo. Nesse tipo de experimento, todo ruído nas imagens pode prejudicar na classificação. Logo, esses são desafios que dificultam o processo de classificação de imagens.

VII. CONCLUSÕES

Na visão computacional, a tarefa de classificação de peixe apresenta aos pesquisadores uma série de desafios. O ambiente natural onde vivem os peixes dificulta as abordagens como a proposta neste trabalho, a exemplo disso podemos citar a iluminação e a subtração de fundo. Esse tipo de problema abrange a teoria computacional e os algoritmos, pois treinar o algoritmo sem um conhecimento completo da classe é um desafio. Neste trabalho, abordamos uma nova proposta para converter imagens em sentenças, ou cadeias de caracteres, baseando-se na metodologia do BOVW, como proposto por [27].

Este trabalho mostrou-se encorajador para seguir com novas propostas de estudos e aprimoramentos, uma vez que houve uma melhora no desempenho conforme é usado valores altos de alfabetos, combinados com a construção do alfabeto pela forma supervisionada. Além disso, a abordagem proposta neste trabalho abre novas frentes de investigação ao permitir novas formas de representar uma imagem por meio de cadeias de caracteres quando comparadas às técnicas usadas em [9, 10].

Como trabalho futuros pretendemos estudar novas formas de geração de sentença a partir dos pontos de interesse detectados; expandir para outros algoritmos de indução gramatical; uso de autômatos adaptativos para realizar a

classificação; verificar a viabilidade do uso de gramáticas regulares e outras que usam exemplos positivos e negativos podem ser alternativas para melhorar o desempenho da classificação; usar uma estratégia de recuperação de erros e planejar um experimento para classificação de objetos a partir de um conjunto de imagens utilizados na literatura.

AGRADECIMENTOS

Os autores agradecem o apoio financeiro da CAPES – Coordenação de Aperfeiçoamento de Pessoal de Nível Superior fornecido ao aluno de doutorado Marcelo Rafael Borth, o qual este trabalho é resultado.

REFERÊNCIAS

- [1] ROVA, A.; MORI, G.; DILL, L. M. One Fish, Two Fish, Butterfish, Trumpeter: Recognizing Fish in Underwater Video. In IAPR Conference on Machine Vision Applications, pp. 404-407, 2007.
- [2] HEITHAUS, M. R.; DILL, L. M. Food availability and tiger shark predation risk influence bottlenose dolphin habitat use. *Ecology*, 83(2), pp. 480-491, 2002.
- [3] ZION, B.; SHKLYAR, A.; KARPLUS, I. In-vivo fish sorting by computer vision. *Aquacultural Engineering* 22, pp. 165-179, 2000.
- [4] BOWEN, M.; MARQUES, S.; SILVA, L.; VONO, V.; GODINHO, H. Comparing on Site Human and Video Counts at Igarapava Fish Ladder, Southeastern Brazil. *Neotropical Ichthyology*, vol. 4, pp. 291-294, 2006.
- [5] FERNANDEZ, D. R.; AGOSTINHO, A. A.; BINI, L. M. Selection of an Experimental Fish Ladder Located at the Dam of the Itaipu Binacional, Paraná River, Brazil. *Brazilian Archives of Biology and Technology*, vol. 47, no. 4, pp. 579-586, 2004.
- [6] CHAN, D.; HOCKADAY, S.; TILLET, R.; ROSS, L. A Trainable N-Tuple Pattern Classifier and its Application for Monitoring Fish Underwater,” in International Conference on Image Processing and its Applications, pp. 255-259, 1999.
- [7] HOGGARTH, D.; ABEYASEKERA, S.; ARTHUR, R.; BEDDINGTON, J. Stock Assessment for Fishery Management: A Framework Guide to the Stock Assessment Tools of the Fisheries Management Science Programme. FAO Fisheries Technical Paper, Rome, Technical Report 487, 2006.
- [8] NERY, M. S. Determining the Appropriate Feature Set for Fish Classification Tasks. *Graphics, Patterns and Images, SIBGRAPI Conference on*, pp. 173-180, 2005.
- [9] NETO, L. C. B.; HIRAKAWA, A. R.; MASSOLA, A. M. A. Aplicação de técnicas adaptativas em reconhecimento de formas. In: Segundo Workshop de Tecnologia Adaptativa - WTA. EPUSP, 2008.
- [10] COSTA, E. R.; HIRAKAWA, A. R.; NETO, J. J. An adaptive alternative for syntactic pattern recognition. *Proceeding of 3rd International Symposium on Robotics and Automation, ISRA*, v.1, 2002.
- [11] SIVIC, J.; ZISSERMAN, A. Video Google: A text retrieval approach to object matching in videos. In *Proceedings of ICCV*, volume 2, pages 1470-1477, Nice, France, oct., 2003.
- [12] CSURKA, G.; DANCE, C. R.; FAN, L.; WILLAMOWSKI, J.; BRAY, C. Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1-22, 2004.
- [13] LOWE, D. G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
- [14] AKRAM, H. I.; DE LA HIGUERA, C. XIAO, H. ECKERT, C. Grammatical inference algorithms in matlab. In *ICGI 2010: Proceedings of the 10th International Colloquium on Grammatical Inference*. Valencia, Spain: Springer-Verlag, 2010.
- [15] BAY, H.; ESS, A.; TUYTELAARS, T.; VAN GOOL, L. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, vol.110, no. 3, pp. 346-359, 2008. Disponível em: <<http://dx.doi.org/10.1016/j.cviu.2007.09.014>>
- [16] DE LA HIGUERA, C. *Grammatical Inference: Learning Automata and Grammars*. Cambridge University Press, 2010.
- [17] FRIGUI, H.; KRISHNAPURAM, R. A robust competitive clustering algorithm with applications in computer vision. *IEEE transactions on pattern analysis and machine intelligence*, 21, 450-465, 1999.
- [18] JAIN, A. K.; FLYNN, P. *Advances in image understanding*. IEEE Computer Society Press. Chap. Image segmentation using clustering, pp. 65-83, 1996.
- [19] SHI, J.; MALIK, J. Normalized cuts and image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 22, 888-905, 2000.
- [20] IWAYAMA, M.; TOKUNAGA, T. Cluster based text categorization: a comparison of category search strategies. Pages 273-281 of: *Proceedings of the 18th ACM international conference on research and development in information retrieval*, 1995.
- [21] SAHAMI, M. Using machine learning to improve information access. Ph.D. thesis, Computer Science Department, Stanford University, 1998.
- [22] BHATIA, S.; DEOGUN, J. Conceptual clustering in information retrieval. *IEEE transactions on systems, man and cybernetics*, 28(B), 427-436, 1998.
- [23] BALDI, P.; HATFIELD, G. *DNA microarrays and gene expression*. Cambridge University Press. 2002.
- [24] SAHOO, N.; CALLAN, J.; KRISHNAN, R.; DUNCAN, G.; PADMAN, R. Incremental hierarchical clustering of text documents. In: *15th ACM international conference on Information and knowledge management*, pp. 357-366, New York, NY, USA, 2006.
- [25] JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. *ACM Computing Surveys*, v.31, n.3, p.264-323. 1999.
- [26] JAIN, A. K. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, v.31, n.8, pp. 651-666, 2010. Disponível em: <<https://s3-us-west-2.amazonaws.com/mlsurveys/45.pdf>>.
- [27] PISTORI, H.; CALWAY, A.; FLACH, P. A new strategy for applying grammatical inference to image classification problems. In: *IEEE International Conference on Industrial Technology (ICIT)*, pp.1032-1037, 2013.
- [28] ZHANG, E.; MAYO, M. Improving bag-of-words model with spatial information. In *25th International Image and Vision Computing New Zealand (IVCNZ)*, pp. 1-8, nov., 2010.
- [29] ZHANG, C.; WANG, S.; HUANG, Q.; LIU, J.; LIANG, C.; TIAN, Q. Image classification using spatial pyramid robust sparse coding. *Pattern Recognition Letters*, vol. 34, no. 9, pp. 1046 - 1052, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167865513000573>>.
- [30] PEDROSA G.; TRAINA, A. From bag-of-visual-words to bag-of-visual phrases using n-grams. In: *26th Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 304-311, aug., 2013.
- [31] LAZEBNIK, S.; SCHMID, C.; PONCE, J. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006.
- [32] YUAN, J.; WU, Y.; YANG, M. Discovery of collocation patterns: from visual words to visual phrases. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-8, jun., 2007.
- [33] DACONTA, M.; OBRST, L.; SMITH, K. *The Semantic Web*. Wiley Publishing Inc., 2003.
- [34] RIBAS, L. C.; BORTH, M. R.; CASTRO JR., A. A.; GONÇALVES, W. N.; PISTORI, H. Grammatical Inference and SIFT for Scene Recognition. *X Workshop de Visão Computacional (WVC)*, Uberlândia-MG, 2014.
- [35] Neto, J. J. Solving complex problems efficiently with adaptive automata. *Conference on Implementation and Application of Automata - CIAA*, Julho, 2000.
- [36] Neto, J. J. Adaptive rule-driven devices - general formulation and case study. *Lecture Notes in Computer Science*. Watson, B.W. and Wood, D. (Eds.): *Implementation and Application of Automata 6th International Conference, CIAA*, Vol. 2494, Pretoria, South Africa, Jul. 23-25, Springer-Verlag, pp. 234-250, 2001.
- [37] Pistori, H.; Neto, J. J. AdapTree - Proposta de um Algoritmo para Indução de Árvores de Decisão Baseado em Técnicas Adaptativas. *Anais Conferência Latino Americana de Informática - CLEI*. Montevideo, Uruguai, Novembro, 2002.
- [38] PEURA, M.; LIVARINEN, J. Efficiency of Simple Shape Descriptors. In: *Aspects of Visual Form*, pp. 443-451. World Scientific, Singapore, 1997.

Tamanho do Alfabeto (k)	Medida-F (%)							
	Aquario10e				AquarioSeg10e			
	Não Supervisionado		Supervisionado		Não Supervisionado		Supervisionado	
	Radial	Hier.	Radial	Hier.	Radial	Hier.	Radial	Hier.
10	1.82	1.82	1.82	1.82	1.81	1.81	3.29	3.39
60	7.15	6.56	9.48	9.10	8.42	10.32	10.81	8.60
120	14.06	10.17	9.46	7.46	15.31	14.21	16.09	13.68
240	10.52	10.19	5.88	6.73	15.74	15.63	18.70	15.99
480	15.93	16.44	15.05	13.49	19.40	23.59	26.36	20.43
960	18.27	17.03	19.82	17.96	31.34	31.27	38.10	36.43
1200	19.54	19.02	18.08	19.19	33.49	22.87	35.47	40.78
1440	19.79	19.09	20.11	18.19	21.66	28.53	46.91	39.31
1800	19.38	20.19	19.05	19.25	23.27	31.41	35.98	43.80
2100	20.39	18.49	20.30	20.28	40.47	30.61	39.01	35.60
2400	20.64	18.68	18.94	*22.31*	23.81	21.02	40.72	*48.19*
2700	19.07	18.43	20.60	20.29	20.96	24.39	38.29	42.14
3000	19.20	21.51	21.08	21.31	23.36	23.27	46.99	46.21

Tabela 2. Medidas-F para a abordagem proposta executada no banco de imagens aquario10e e aquarioSeg10e. Os valores mais altos para cada tamanho do alfabeto para ambos os bancos de imagens estão marcados em negrito. A Medida-F mais alta para cada banco de imagens está marcada com asteriscos.