

UNIVERSIDADE CATÓLICA DOM BOSCO
PROGRAMA DE PÓS-GRADUAÇÃO *STRICTO SENSU* EM
CIÊNCIAS AMBIENTAIS E SUSTENTABILIDADE AGROPECUÁRIA

Classificação Automática de Espécies de Peixe Utilizando
Reconhecimento Sintático de Padrões

Autor: Marcelo Rafael Borth
Orientador: Dr. Hemerson Pistori
Coorientador: Dr. Wesley Nunes Gonçalves

Campo Grande – Mato Grosso do Sul
Março – 2017

UNIVERSIDADE CATÓLICA DOM BOSCO
PROGRAMA DE PÓS-GRADUAÇÃO STRICTO SENSU EM
CIÊNCIAS AMBIENTAIS E SUSTENTABILIDADE AGROPECUÁRIA

Classificação Automática de Espécies de Peixe Utilizando
Reconhecimento Sintático de Padrões

Autor: Marcelo Rafael Borth
Orientador: Dr. Hemerson Pistori
Coorientador: Dr. Wesley Nunes Gonçalves

Tese apresentada como parte das exigências para obtenção do título de DOUTOR EM CIÊNCIAS AMBIENTAIS E SUSTENTABILIDADE AGROPECUÁRIA, no Programa de Pós-Graduação *Stricto Sensu* em Ciências Ambientais e Sustentabilidade Agropecuária da Universidade Católica Dom Bosco - Área de concentração: Sustentabilidade Ambiental e Produtiva Aplicada ao Agronegócio e Produção Sustentável.

Campo Grande – Mato Grosso do Sul
Março – 2017

Dados Internacionais de Catalogação na Publicação (CIP)
(Biblioteca do Instituto Federal do Paraná – Campus Umuarama)

B739c Borth, Marcelo Rafael
Classificação automática de espécies de peixe utilizando reconhecimento sintático de padrões/ Marcelo Rafael Borth. – Campo Grande, 2017.
85f. il.

Orientador: Prof. Dr. Hemerson Pistori

Tese (doutorado) - Universidade Católica Dom Bosco, Departamento de Ciências Agrárias e Sustentabilidade Agropecuária. – Programa de Pós-Graduação em Ciências Ambientais e Sustentabilidade Agropecuária, 2017.

1. Limitação de pontos de interesse. 2. Classificação de Imagens. 3. Reconhecimento Sintático de Padrões. 4. Inferência Gramatical. 5. Visão Computacional. I. Pistori, Hemerson. II. Universidade Católica Dom Bosco. III Título.

CDD 21. ed. 639.3



UNIVERSIDADE CATÓLICA DOM BOSCO
Inspira o futuro

Classificação Automática de Espécies de Peixe Utilizando Reconhecimento Sintático de Padrões

Autor(a): Marcelo Rafael Borth

Orientador: Prof. Dr. Hemerson Pistori

Coorientador: Prof. Dr. Wesley Nunes Gonçalves

TITULAÇÃO: Doutor em Ciências Ambientais e Sustentabilidade Agropecuária

Área de Concentração: Sustentabilidade Ambiental e Produtiva.

APROVADO em 10 de março de 2017.

Prof. Dr. Hemerson Pistori - UCDB

(Orientador)

Prof. Dr. Wesley Nunes Gonçalves - UFMS

(Coorientador)

Prof. Dr. Marco Hiroshi Naka - IFMS

Profa. Dra. Milena Wolff Ferreira - UCDB

Prof. Dr. Jonathan de Andrade Silva - UFMS

Prof. Dr. Willian Paraguassu Amorim - UFGD

AGRADECIMENTOS

Não podemos dar passos grandes sem contar com o apoio e o estímulo de pessoas maravilhosas que conhecemos e mantemos em nosso caminho. Por isso, quero expressar minha gratidão a todos que, direta ou indiretamente, me ajudaram a realizar este trabalho.

Sobre tudo e todos, agradeço a **Deus**. Foi Dele que recebi saúde, sabedoria, discernimento, determinação e muita capacitação nos momentos em que precisei. Foi em Suas mãos que entreguei o doutorado para que ele fosse concluído com sucesso. Toda honra e glória ao Senhor;

À **esposa** mais amada do mundo, **Niara Borth**, por todo amor, carinho, cumplicidade, orações e dedicação que teve por mim nesse longo período. Tive que abdicar sua companhia muitas e muitas vezes para estudar. Obrigado por compreender, me apoiar e mesmo assim me amar cada vez mais;

Aos meus **pais**, **Valdecir Oneide Borth** e **Fátima A. de Caldas Borth** por toda educação que me deram. Obrigado por sempre me incentivar a estudar, mesmo em situações adversas ou de dificuldades. À minha **irmã Nathália** pelo amor e carinho de sempre. Obrigado família;

Ao meu orientador, **Hemerson Pistori**, pela confiança e por ter me guiado nessa jornada desafiadora e a meu coorientador, **Wesley Nunes Gonçalves**, por toda ajuda durante esse período. Vocês fizeram a diferença nessa jornada e, com certeza, farão também na minha vida;

Ao apoio financeiro da **CAPES** e **UCDB**; e,

Aos colaboradores do Programa de Pós-Graduação em Ciências Ambientais e Sustentabilidade Agropecuária.

SUMÁRIO

	Página
INTRODUÇÃO	1
OBJETIVOS	6
Objetivo geral	6
Objetivos específicos.....	6
REVISÃO BIBLIOGRÁFICA	7
Peixes: um panorama global	7
Peixes no Brasil	8
Peixes no Mato Grosso do Sul	9
Visão computacional	11
Linguagens formais	12
Gramáticas	13
Autômatos finitos	14
Análise sintática.....	15
Recuperação de erros em analisadores sintáticos.....	16
Inferência gramatical	16
<i>Scale Invariant Feature Transform</i> - SIFT	20
Histograma de palavras visuais (BOVW)	22
Reconhecimento sintático de padrões	23
Reconhecimento sintático de padrões aplicado em problemas de classificação de imagens	25

Abordagens propostas para o reconhecimento de espécies de peixes	26
METODOLOGIA	28
Construção dos bancos de imagens de peixe.....	28
Proposta de um algoritmo para a classificação de espécies de peixes utilizando reconhecimento sintático de padrões	32
Implementação da abordagem proposta.....	36
Validação da abordagem proposta.....	37
Organização dos próximos capítulos	38
CAPÍTULO I	40
CAPÍTULO II	51
CAPÍTULO III	58
CONSIDERAÇÕES FINAIS	65
REFERÊNCIAS BIBLIOGRÁFICAS	67

LISTA DE TABELAS

	Página
Tabela 1. Exemplo passo a passo do funcionamento do algoritmo de inferência gramatical Sequitur.....	20
Tabela 2. Características dos bancos de imagens criados.	30
Tabela 3. Espécies de peixes do banco de imagens “Aquario28e40i” com o respectivo nome popular e científico de cada espécie.	31

LISTA DE FIGURAS

	Página
Figura 1. Exemplo de detecção de pontos de interesse utilizando o algoritmo SIFT.	21
Figura 2. Etapas necessárias para um sistema de reconhecimento sintático de padrões típico (PIO, 2006).....	25
Figura 3. Aquário Municipal de Toledo - Paraná.	29
Figura 4. Exemplo de uma imagem original e uma imagem recortada para composição do banco de imagens.	30
Figura 5. Ilustração da construção do alfabeto pela aprendizagem supervisionada e não supervisionada.....	34
Figura 6. Passos da proposta de um algoritmo para a classificação de espécies de peixes utilizando reconhecimento sintático de padrões.	35
Figura 7. Interface proposta da aplicação para a classificação de espécies de peixes utilizando reconhecimento sintático de padrões.....	37

LISTA DE ABREVIATURAS

AFD	Autômatos finitos determinísticos
AFND	Autômatos finitos não-determinísticos
BOW	Histograma de palavras
BOVW	Histograma de palavras visuais
MPA	Ministério de pesca e aquicultura
OMS	Organização mundial da saúde
PAC	Programa de aceleração do crescimento
SIFT	<i>Scale invariant feature transform</i>
SURF	<i>Speeded up robust features</i>

RESUMO

A necessidade de classificar espécies de peixes muitas vezes é crítica para ecólogos, pesquisadores de campo, piscicultores e empresas de consultorias. Para tanto, a automatização desse tipo de processo pode proporcionar maior agilidade e precisão aos profissionais da área. A criação de softwares visando o aperfeiçoamento da agroecologia pode permitir a adoção de medidas alternativas instantâneas e ações pontuais. Este trabalho apresenta uma nova abordagem com testes práticos para a classificação automática de espécies de peixe em problemas de visão computacional. O trabalho propõe a representação de uma imagem como sentença, gerada a partir de agrupamento hierárquico, de modo que técnicas de inferência gramatical possam ser aplicadas. Para a classificação de imagens dessa abordagem, inicialmente é preciso ter uma base de dados com informações do objeto a ser identificado, a fim de ter o conjunto de treinamento para o classificador. Para este trabalho, são detectados os pontos de interesse usando o algoritmo *Scale Invariant Feature Transform* – SIFT e realizada uma limitação da quantidade de pontos utilizados baseado no conceito de cauda longa e Lei de Zipf's; posteriormente os pontos de interesse são descritos usando o descritor *Opponent-SIFT*; 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*); na sequência, uma sentença para cada imagem é gerada. 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; em seguida, realiza-se a aprendizagem da gramática, tendo como resultado um autômato para cada classe; por fim, a classificação é realizada utilizando o algoritmo de inferência gramatical *k-testable*. Foram executados experimentos utilizando três novos bancos de imagens de peixes, a fim de analisar o impacto da limitação de pontos de interesse com o uso de cores para descrição dos pontos. O resultado da Medida-F alcançado foi de 76,8% e uma melhora de até 1900x no tempo de processamento da classificação comparada a abordagem base analisada. As principais contribuições deste trabalho estão na nova abordagem criada para classificar imagens usando reconhecimento sintático e estrutural de padrões; no tipo de ordenação usada para a geração de sentenças, em testes com uma larga faixa de tamanho de alfabeto; criação de três bancos de imagens de peixes com fotos capturadas de celulares; e, principalmente, na limitação da quantidade de pontos de interesse usados para a classificação de imagens juntamente com o uso de informações de cores nos descritores de cada ponto de interesse.

Palavras-chave: Limitação de pontos de interesse, classificação de imagens, reconhecimento sintático de padrões, inferência gramatical, visão computacional.

ABSTRACT

The need to classify fish species is often critical to ecologists, researchers, fish farmers, consultants, and so on. For this, automating this process can provide greater speed and precision to professionals related to this area. The creation of software aiming the improvement of agroecology may allow the adoption of some alternatives and specific actions. This work presents a new approach with practical tests for an automatic fish classification in computer vision problems. The work proposes the representation of an image as sentence, generated from a grouping based on hierarchical clustering, so that grammatical inference techniques can be applied. For image classification, initially it is necessary to have a database with information of the object to be identified, a training set for the classifier. For this work, keypoints are detected using the Scale Invariant Feature Transform (SIFT) algorithm and the keypoint limitation based on the long tail concept and Zipf's Law is performed; later keypoints are described using Opponent-SIFT descriptor; then the alphabet is built based on the set of keypoints descriptors, having as result the labeling of each one. The alphabet is defined using the k-means algorithm as in the Bag of Visual Words; as a result, a sentence labeled for each image is generated. The sentence ordering is generated based on hierarchical clustering, joining keypoints based on spatial position (x, y) to form a complete binary tree. Finally, the grammar learning is performed, resulting in an automaton for each class; finally, for the image classification is used a grammatical inference algorithm called k-testable. Experiments were carried out using three new datasets of fish images, in order to analyze the impact of the keypoints limitation with the use of colors to describe keypoints. The F-Measure achieved was 76.8% and an improvement up to 1900x in the classification processing time (disregarding the learning stage) compared to the analyzed base approach. The main contributions of this work are: the novel approach to classify images based on syntactic and structural pattern recognition; the sorting order used generating sentences, testing with a wide alphabet size range; creation of three image dataset of fish species with photos captured from cell phones; and mainly the keypoint limitation strategy used for the classification along with the use of color information in the descriptors for each keypoint.

Keywords: *Keypoint limitation, Image classification, syntactical pattern recognition, grammatical inference, computer vision.*

INTRODUÇÃO

Informações que tratam sobre reconhecimento, classificação, distribuição, quantificação e migração de espécie de peixes são críticas para ecologistas, estudiosos do meio ambiente, órgãos governamentais e empresas de pescada. Nesse contexto, a criação de softwares para monitorar peixes ou o seu ambiente possibilita adotar alternativas ou ações específicas para atuarem desde o controle preventivo de espécies de peixes até a criação de soluções para o pescada, como pode ser observado em Lima (2012) e Feitosa e da Silva (2014). Situações como, liberação de agrotóxicos em rios, esgotos de cidades, liberação de resíduos em ambientes inadequados e construção de barragens, podem afetar a distribuição, abundância e a cadeia alimentar, podendo reduzir ou até eliminar a ictiofauna local (Godinho e Godinho, 1994; Woynarovich, 1991).

Apesar do cenário ser cada vez mais positivo, o setor da piscicultura é ainda carente de tecnologias, gerando a premente necessidade por pesquisas e desenvolvimento, acompanhadas de transferência de tecnologia que preencham tais lacunas (Rodrigues *et al.*, 2013). O conceito de tecnologia nessa área passou a ocupar um crescente espaço em debates acadêmicos e empresariais (Assad e Bursztyl, 2000; Almeida, 2003). Logo, o monitoramento e desenvolvimento de pesquisas são ferramentas importantes para a sustentabilidade dessa área (Tureck e Oliveira, 2009). Portanto, é fundamental o desenvolvimento de tecnologias que possam contribuir para a competitividade desse setor (Sidonio *et al.*, 2012).

Em geral, alguns diagnósticos são realizados por peritos, em que eles usam a própria visão para detectar determinadas características ou doenças, no entanto esses profissionais são escassos. Nas empresas de pescada muitas vezes são utilizadas soluções manuais rotineiras como, receber cargas de peixes e realizar a catalogação mediante a identificação de informações referentes ao tamanho, peso,

quantidade, espécie, entre outras atividades. Para identificar a espécie de um peixe, profissionais precisam fixá-lo na formalina (solução cancerígena) e levá-lo ao laboratório para realizar a identificação. Com um software que faça a identificação da espécie de peixe é possível observar as seguintes vantagens:

- a) Preservação de espécies (peixes não seriam mortos para a identificação);
- b) Redução de tempo do profissional para identificar a espécie do peixe;
- c) Aumento da produtividade e praticidade ao profissional no processo de identificação de peixe; e,
- d) Redução do risco de vida do profissional, evitando o contato com substâncias perigosas.

Essas vantagens fortalecem a justificativa de que o desenvolvimento de softwares para classificar espécies de peixes seria importante para pesquisadores de campo e ecólogos, uma vez que seria possível identificar mais eficientemente a espécie de um peixe na margem de um rio/lago/riacho/córrego e devolvê-lo ao seu devido lugar, evitando também que ele seja morto para posterior identificação em laboratório.

Contar espécies de peixes em uma região específica de um rio é uma forma de estudar a sua saúde e viabilidade de peixes na área. Abordagens como orientar a gestão de pescas, avaliar o impacto ecológico de barragens, gerir peixes comerciais podem ser algumas das aplicações interessantes e úteis para empresas do pescado (Rova *et al.*, 2007). 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 (Rova *et al.*, 2007; Heithaus *et al.*, 2002; Zion *et al.*, 2000), inclusão de estudos ecológicos e ambientais das comunidades de peixes (Bowen *et al.*, 2006), projeto e colocação de escadas de peixes em represas para geração de energia hidrelétrica (Fernandez *et al.*, 2004), alimentação estratégica pelos piscicultores (Chan *et al.*, 1999) e avaliação de ações para gestão das pescas (Hoggarth *et al.*, 2006).

Os benefícios também podem ser observados em empresas que trabalham com consultorias e monitoramento ambiental. No Brasil, o comércio de peixes está sobreexplorado, especialmente pela falta de fiscalização e pelo uso de técnicas

predatórias de pesca (Tureck e Oliveira, 2009), causando grande impacto nas populações naturais (Val *et al.*, 2000). Para seguir os princípios de sustentabilidade na maricultura (cultivo de organismos marinhos) é preciso se basear na produção lucrativa, preservação ambiental e desenvolvimento social (Valenti *et al.*, 2000). Logo, policiais ambientais poderiam utilizar o software de reconhecimento de espécie de peixe ao abordar pescadores para verificar se foram feitas pescas de espécies proibidas naquela região e, também, se a espécie está ou não em extinção, auxiliando na identificação de possíveis infrações.

Por conta dessas problemáticas e motivações apresentadas, esta tese apresenta uma abordagem para a classificação de espécies de peixes. Foi desenvolvido um software para a classificação de espécies de peixe, sustentado por técnicas de visão computacional. As técnicas aplicadas no processamento e análise de imagens possibilitam extrair informações de padrões visuais com o objetivo de identificar características presentes em imagens para utilizar em sistemas de reconhecimento de padrões.

Em geral, esse tipo de aplicação pode ajudar na expansão da capacidade produtiva como estratégia de negócio, uma vez que essa área se encaixa como um dos eixos de desenvolvimento do estado de Mato Grosso do Sul. A visão computacional tem se mostrado uma alternativa para racionalizar custos, implantando o modelo de gestão chamado agricultura de precisão, o qual está associado à utilização de equipamentos de alta tecnologia para avaliar ou monitorar condições específicas em uma determinada propriedade. Com isso, é possível fornecer alternativas aos produtores que maximize o rendimento e a lucratividade, enquanto que ao mesmo tempo poderá reduzir o impacto ambiental (Bramley, 2009; Fountas *et al.*, 2005; Griffin, *et al.*, 2004; Zhang *et al.*, 2000). Por se tratar de uma abordagem que analisa imagens, ela pode ser mais precisa, uma vez que a análise visual pode ser passível de erro. Assim, softwares inteligentes podem auxiliar na tomada de decisão de maneira que sejam adotadas novas estratégias ou adaptações junto a modelos de produção tradicional vigente.

O objetivo desta tese foi criar uma nova abordagem para o problema de classificação automática de espécies de peixe. Esta pesquisa resgata técnicas de reconhecimento sintático de padrões muito exploradas nas décadas de 60 e 70,

todavia utiliza como motivação os avanços recentes da computação, visão computacional e aprendizagem automática, tais como: aumento do poder computacional e novas técnicas de extração e descrição de atributos em imagens. Trabalhos recentes que abordam reconhecimento sintático de padrões podem ser vistos em Pistori *et al.* (2013) e Ribas *et al.* (2014).

Esta pesquisa explora a abordagem da representação de imagens como sentenças usada por Pistori *et al.* (2013). É feita a limitação de pontos de interesse baseada no conceito de *long tail* (Anderson, 1995) e Lei de Zipf's (Newman, 2006), e gera as sentenças baseada em estrutura hierárquica, ambas as técnicas não usadas na literatura até o momento. Para tanto, este trabalho inicialmente (i) detecta pontos de interesse utilizando o algoritmo *Scale Invariant Feature Transform* – SIFT (Lowe, 2004); (ii) realiza a limitação da quantidade máxima de pontos de interesse de cada imagem para remover ruídos e não influenciar na classificação de a ter predominância de informação entre classes; (iii) descreve os pontos usando informação de cor utilizando o algoritmo Opponent-SIFT (Van de Sande *et al.*, 2008; Van de Sande *et al.*, 2010); (iv) constrói 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. O alfabeto é construído pela forma supervisionada, conforme usado por Ribas *et al.* (2014). Na construção supervisionada, os centroides são gerados separadamente para cada classe, enquanto que para a não supervisionada, não é considerada a separação de classes para gerar os centroides; (v) realiza-se a geração da sentença por agrupamento hierárquico de modo que a informação espacial não seja perdida. Esse agrupamento une pontos de interesse próximos a partir da sua posição espacial do eixo (x, y), até formar uma árvore binária completa. Ele é usado com o objetivo de não perder a informação espacial entre os pontos de interesse, visto que no histograma de palavras visuais é realizada apenas uma contagem de pontos e a relação de localização entre os pontos é perdida. Em virtude disso, os pontos de interesse não têm mudança da ordem na geração de sentença quando as imagens são rotacionadas. Para isso, é usado o algoritmo *linkage* implementado no Matlab. Esse algoritmo usa a informação de distância da posição espacial (x, y) do ponto de interesse, a fim de determinar a proximidade de objetos entre si; por fim, (vi) realiza-se a aprendizagem da gramática

para a classificação de imagens utilizando o algoritmo de inferência gramatical *k-testable* (Akram *et al.*, 2010).

As principais contribuições desta tese estão na abordagem criada resgatando a classificação de imagens baseando-se no reconhecimento sintático de padrões utilizando os avanços recentes da computação; na ordenação hierárquica para a geração de sentenças sem que as informações espaciais entre os pontos de interesse sejam perdidas; em testes com uma larga faixa de tamanho de alfabeto; na criação de três bancos de imagens de peixes com fotos capturadas por celulares; e, principalmente, em uma nova estratégia para a limitação de pontos de interesse, na qual usa cores para a descrição de pontos de interesse.

Esta tese está vinculada ao projeto *FISHCV* do grupo de pesquisa INOVISAO¹, que visa criar um aplicativo para reconhecimento de espécies de peixes para dispositivos móveis. Pretende-se utilizar a nova abordagem em problemas de importância para o estado de Mato Grosso do Sul, os quais estão relacionados com a produção sustentável em agropecuária. A aplicação poderá gerar contribuição aos ecólogos, biólogos, estudantes, pesquisadores, engenheiros, instituições governamentais e empresas de pescado, entre outros.

Este trabalho está organizado da seguinte maneira: a seguir é apresentada a revisão bibliográfica para melhor entendimento dos conceitos deste trabalho. Em seguida, é apresentada a metodologia. No Capítulo 1 é apresentado um artigo referente a proposta inicial deste trabalho. No Capítulo 2 é apresentado um artigo que traz a abordagem deste trabalho com duas adaptações isoladas, obtendo dois resultados que melhoraram o resultado da classificação. No Capítulo 3 é apresentado um artigo com a versão final da abordagem desta tese, na qual utiliza a limitação de pontos de interesse juntamente com o uso de cores para a geração dos descritores de cada ponto. Por fim, são apresentadas as considerações finais e direcionamentos futuros.

¹ O site do grupo INOVISAO encontra-se em <http://inovisao.weebly.com/>

OBJETIVOS

Objetivo geral

Criar uma abordagem baseada no reconhecimento sintático de padrões que seja capaz de classificar espécies de peixes, a partir de imagens capturadas por dispositivos móveis.

Objetivos específicos

Para atingir o objetivo geral definido, os seguintes objetivos específicos são necessários:

- Construir um banco de imagens com espécies de peixe;
- Propor uma abordagem que faça o reconhecimento de espécies de peixe a partir de inferência gramatical;
- Implementar a abordagem proposta; e,
- Validar a abordagem proposta com experimentos baseados nos bancos de imagens criados.

REVISÃO BIBLIOGRÁFICA

Peixes: um panorama global

A produção mundial de pescado foi de aproximadamente 158 milhões de toneladas em 2012 (Brabo *et al.*, 2016). Os maiores produtores são a China, a Indonésia e a Índia, respectivamente. Apenas a China representa cerca de 58% do total produzido no mundo (FAO, 2016). O continente responsável pela maior parte da produção de pescado em 2012 foi a Ásia com 103,6 milhões de toneladas, seguido da América com 25,5 milhões (Brabo *et al.*, 2016). O Brasil encontra-se apenas na 14ª posição na classificação mundial de produção de peixes, representando uma discreta parcela de aproximadamente 1% do total produzido (FAO, 2016).

A produção de peixes é cada vez mais valorizada como alimento, uma vez que sua carne possui baixo teor de gordura, além de possuir ácidos graxos insaturados. Além disso, essa gordura proporciona uma série de benefícios no combate e prevenção de doenças.

Muitas espécies são criadas em condições e ambientes artificiais, seja para a alimentação humana ou, até mesmo, para outras finalidades como, a habitação doméstica ou aquários comerciais para visitação. Além disso, os peixes também têm um papel respeitável na cultura mundial. Para alguns, são considerados símbolos religiosos, temas relacionados à arte, livros, filmes, entre outros. Por outro lado, existem espécies de peixes ameaçadas de extinção, seja pela prática exagerada da pesca ou deterioração de seus habitats.

Os peixes constituem aproximadamente 50% de todas as espécies de animais vertebrados conhecidas, sendo o grupo de maior diversidade dentre os vertebrados (Nelson, 2006). Cada espécie de peixe possui características específicas que as permitem ser únicas, tais como: tamanho de suas nadadeiras, padrões de cores e

tamanho do espécime. Na classificação de peixes, por exemplo, existem 47 características distintas por espécie para que se consiga um nível bom de exatidão na identificação (Nery, 2005).

Peixes no Brasil

Em 2014, a produção total da piscicultura brasileira foi de aproximadamente 474 mil toneladas, representando um aumento de 20,9% em relação ao ano anterior (IBGE, 2014). Embora o Brasil apresente uma parcela pequena da produção mundial, a piscicultura é uma das atividades da pecuária de maior expansão no país. De acordo com a FAO (2016), estima-se que o Brasil cresça em torno de 104% na produção da pesca e aquicultura até 2025, devido aos investimentos sendo realizados no setor nos últimos anos. No âmbito global, essa estimativa é em torno de 17%.

Pela evidência do alto valor nutricional do peixe, está cada vez mais comum a sua inclusão nas refeições da população brasileira. Conforme apresenta o Ministério de Pesca e Aquicultura (MPA), em média, cada brasileiro consome 9 quilos de pescado anualmente, porém o total recomendado pela Organização Mundial da Saúde (OMS) é 12 quilos (Portal Da Saúde, 2012). Por meio de ações que promovam a importância do produto, pretende-se elevar o seu consumo.

Na região Neotropical são descritas cerca de 6.000 das 13.000 espécies válidas de peixes de água doce (Reis *et al.*, 2003). O Brasil apresenta um papel muito favorável nessa região em virtude de sua grande diversidade (Buckup *et al.*, 2007), além de grande quantidade de reservatórios para o cultivo de peixes em cativeiro. O potencial hídrico e o clima favorável são outros dois motivos que compatibilizam o crescimento do Brasil nessa área. Apesar do país possuir 13% de toda a água doce do planeta (Sidonio *et al.*, 2012), ele produz uma singela parcela de 2% do pescado mundial, considerando rios e lagos (Bezerra *et al.*, 2003). Em virtude do rico manancial hídrico do Brasil, existem espécies das mais variadas adaptações morfológicas, fisiológicas e comportamentais (Rodrigues *et al.*, 2013).

Ostrensky e Boeger (2007) apontam os três maiores problemas que afetam o setor aquícola brasileiro, sendo eles, falta de treinamento e qualificação técnica, dificuldade de acesso ao crédito e falta de políticas públicas consistentes para o

desenvolvimento da atividade. Além disso, Borghetti (2000) afirma que existem grandes problemas para que a aquicultura brasileira se desenvolva satisfatoriamente, por exemplo, a falta de organização do sistema de transferência de tecnologia; a carência de pesquisa aplicada; e, a deficiência do sistema de comercialização e distribuição de produtos pesqueiros. A sustentabilidade de recursos pesqueiros depende da adoção de alternativas que considerem os aspectos sociais, econômicos, ambientais e tecnológicos. Santos e Santos (2005) afirmam que os dois principais problemas enfrentados pelo setor pesqueiro estão nos métodos muito artesanais da própria atividade pesqueira e na tecnologia do pescado. Além disso, mediante visitas técnicas realizadas no estado de Mato Grosso do Sul, constatamos que estes problemas ainda persistem nos dias de hoje.

Peixes no Mato Grosso do Sul

A piscicultura no estado de Mato Grosso do Sul tem conquistado espaço entre as demais produções. Grande parte da área do pantanal está localizada nesse estado, caracterizando-se por ser de planície alagável e por abrigar abundantemente espécies de peixes. Cerca de 260 espécies catalogadas são mantidas devido ao cuidado e boa conservação da planície pantaneira (Satolani, *et al.*, 2008).

Em geral, a piscicultura é uma atividade extra e complementar aos produtores do Mato Grosso do Sul - Brasil, pois não se trata de uma atividade exclusiva geradora do faturamento das propriedades, devido a sazonalidade e ao potencial de escala da produção de peixes (Prochmann e Tredezini, 2004). Isso mostra que o estado precisa incentivos e estudos adicionais para auxiliar e fornecer alternativas tecnológicas, com o intuito de aumentar a produção e consolidação da atividade como fonte de renda aos produtores.

Abordando a perspectiva de entretenimento e pesquisa, em Campo Grande, capital de Mato Grosso do Sul, está prevista a construção do aquário do Pantanal^{2,3} para visitação. O aquário será o maior aquário do Brasil e terá uma estrutura de 18,6

² Detalhes da obra pode ser encontrada em: <http://g1.globo.com/mato-grosso-do-sul/noticia/2011/05/aquario-do-pantanal-sera-lancado-oficialmente-em-campo-grande.html>

³ Detalhes da obra pode ser encontrada em: <https://www.campograndenews.com.br/cidades/capital/area-do-aquario-do-pantanal-tem-o-dobro-do-tamanho-do-belmar-fidalgo>

mil metros quadrados de área e será capaz de receber até 20 mil visitantes por dia. O total da área reservada vai abrigar 16 grandes tanques dentro do prédio e 5 na área externa, em que concentrarão 6,6 milhões de litros de água, tornando-o, assim, o maior aquário de água doce do mundo. Com o volume total previsto será possível suportar aproximadamente 7 mil animais, subdivididos em torno de 260 espécies, dentre eles, peixes, crustáceos, moluscos, invertebrados, répteis e mamíferos. Será possível apresentar toda a rica biodiversidade dos rios do estado, além de uma grande variedade de animais e plantas típicas. A intenção é retratar a biodiversidade dos ambientes aquáticos do Pantanal. O manejo das espécies em exposição será objeto de estudos e pesquisas constantes. O novo aquário deve se tornar o maior referencial do País em consulta científica da fauna e flora da região pantaneira. Esse projeto integra o programa “MS Forte” do governo do estado de Mato Grosso do Sul como obra emblemática da capital. Em caráter de entretenimento para o aquário, poderia ser desenvolvido um aplicativo móvel onde o visitante tiraria uma foto de um peixe e ele mesmo identificaria qual a espécie de peixe, recebendo uma pontuação para cada acerto.

O tratamento com os peixes do pantanal também é delicado e precisa de pesquisas que ajudam na sustentabilidade. O aumento da visitação no pantanal, por exemplo, pode interferir no equilíbrio não apenas dos animais, mas também de plantas, afetando de forma negativa a diversidade de espécies, bem como suas relações ecológicas. Além disso, pode-se gerar perdas de biodiversidade em locais mais frágeis, como pode ser constatado em Sabino e Andrade (2003), a qual utiliza os peixes como indicadores. Espécies de peixes, plantas, aves e invertebrados servem como “termômetro” da saúde dos rios. Isso significa que, se houver mudança da população, para mais ou para menos, algo de errado pode estar acontecendo com os rios. Nesse caso, é preciso adotar formas de monitoramento e gestão ambiental. Logo, tornam-se necessárias pesquisas tecnológicas para auxiliar e contribuir com órgãos de fiscalização e controle, para manter ou recuperar a integridade dos ecossistemas. Nesse aspecto, a visão computacional é uma área que pode fornecer um grande auxílio com projetos e pesquisas de monitoramento para controle e recuperação ambiental.

Visão computacional

A visão computacional é uma área de pesquisa que inclui métodos de aquisição, processamento, análise e reconhecimento de imagens. Os algoritmos de visão computacional são baseados em técnicas que permitem extrair informações a partir de imagens (Gonzalez *et al.*, 2007; Trucco *et al.*, 1998; Faugeras, 1993; Hartley *et al.*, 2004; Forsyth *et al.*, 2012). Essa é uma área que desenvolve algoritmos visando automatizar tarefas realizadas normalmente pela visão humana. Com o uso da visão computacional é possível automatizar a tomada de decisões com base em imagens capturadas por câmeras fotográficas (Shapiro *et al.*, 2001), microscópios, celulares, entre outros. Essa área é usada e destaca-se em uma grande variedade de aplicações, tais como: reconhecimento de objetos, detecção e reconhecimento de face (Patel *et al.*, 2015), detecção de veículos para propósitos de direção segura, detecção de impurezas em indústria têxtil, detecção da qualidade na textura de materiais (Abdulrahman, 2014), análise de imagens aeroespaciais (Perelmuter *et al.*, 1995), dentre outras.

Para realizar o reconhecimento de um objeto na visão computacional (por exemplo, uma espécie de peixe⁴) é comum usar algumas etapas como: (i) captura de imagens; (ii) separação das imagens capturadas em classes distintas para compor o conjunto de exemplos; (iii) pré-processamento de imagens com o objetivo de reduzir ruídos e imperfeições geradas na fase da captura. Nesse caso, haverá a imagem original como entrada e uma imagem melhorada como saída (resultado), visando reduzir ou adequar distorções, brilho, contraste, cor, defeitos e ruídos, presentes na captura, para que etapas posteriores não sofram interferências dessas imperfeições. Também, uma etapa importante é a (iv) segmentação da imagem, que consiste na separação ou divisão da imagem em regiões baseando-se nos objetos constituintes. Esse processo determina a fronteira de regiões na imagem, ou seja, ocorre a separação do objeto (peixe) do fundo da imagem. Uma segmentação bem sucedida aumenta o percentual de acerto em problemas de identificação e classificação, logo, quanto maior a precisão da segmentação, maior a chance de sucesso do classificador (Gonzalez e Woods, 2007). Outra etapa fundamental para alcançar os objetivos do reconhecimento é a (v) extração de atributos, pois a capacidade de reconhecer

⁴ O objeto citado é o peixe. Entretanto, pode ser qualquer animal, objeto, tipos de imagem, etc.

padrões se baseia na quantidade de informações que se conhece a priori do objeto em questão (Peura *et al.*, 1997). A extração de atributos resulta em informações quantitativas de interesse ou que possam ser utilizadas para diferenciar uma classe de outra. Por fim, pode ser aplicada a (vi) aprendizagem de máquina como técnica de descoberta de conhecimento, a fim de inferir conhecimentos implícitos nos dados previamente conhecidos para prever comportamentos futuros. Embora as etapas mencionadas são apresentadas em sequência, elas podem ser independentes, sendo assim, pode haver situações em que apenas algumas das etapas resolvam o problema mediante o uso de uma metodologia diferente da apresentada.

Junto com a área da visão computacional pode ser usada técnicas de reconhecimento de padrões que possibilitam a identificação e classificação de objetos. É nesse ponto o foco principal deste projeto de pesquisa, na qual esta tese propõe uma alternativa para realizar o reconhecimento de espécies de peixe baseando-se no reconhecimento sintático de padrões, uma vez que será utilizado técnicas recentes e populares do estado da arte. Em conjunto, é possível usar sistemas com formalismos complexos, que buscam induzir modelos capazes de representar e reconhecer novos objetos, como as linguagens formais.

Linguagens formais

O ser humano usa a linguagem natural para se comunicar, porém ela possui ambiguidades que impedem seu uso em sistemas formais. Para evitar esse problema é preciso utilizar uma linguagem construída especificamente para esse propósito, definindo um conjunto de elementos (símbolos) e um conjunto de métodos (regras de derivação). A linguagem formal, por sua vez, é um conjunto de símbolos de um alfabeto finito e não-vazio. Logo, a teoria das linguagens formais estuda modelos matemáticos para especificar e reconhecer linguagens, suas classificações, estruturas, características e propriedades. Na teoria das linguagens formais, é possível gerar um conjunto de regras, as quais especificam a formação de um conjunto de elementos, caracterizando uma gramática formal.

Símbolos são os menores elementos que compõem uma linguagem. Nas linguagens formais são considerados os mnemônicos.

Alfabeto⁵ é um conjunto finito de símbolos distintos, como no alfabeto romano, tal que, $R = \{a, b, c, d, \dots, z\}$, ou no alfabeto grego, tal que, $G = \{\alpha, \beta, \gamma, \delta, \dots, \eta\}$. Vários símbolos podem ser considerados parte de um alfabeto e esses símbolos são considerados parte de uma linguagem.

Sentença⁶ é uma sequência finita de símbolos ordenados de um alfabeto, por exemplo, “abbacde”. A sentença é construída com base em um alfabeto pré-determinado. Uma sentença não pode ter um símbolo que não pertence ao alfabeto. Além disso, uma sentença pode ter símbolos repetidos e tamanhos diferentes.

Uma linguagem constituída sobre um alfabeto é um conjunto finito de sentenças formadas a partir desse mesmo alfabeto. Para reconhecer uma linguagem é preciso analisar as sentenças, dizendo se elas pertencem ou não à linguagem. Para isso é preciso gerar gramáticas a partir de sentenças.

Gramáticas

As gramáticas facilitam a definição de características de linguagens, pois elas servem para representar uma linguagem. Uma gramática define uma estrutura sobre um alfabeto de modo que apenas determinadas combinações são consideradas sentenças. Conforme afirma Menezes (2011), uma gramática é um conjunto finito de regras que, quando aplicadas sucessivamente, geram ou reconhecem sentenças e, o conjunto de todas as palavras geradas pela gramática define a linguagem.

As gramáticas permitem projetar o formalismo para definir linguagens. Esse formalismo está ligado ao modo de como são geradas as sentenças de uma linguagem. O elemento fundamental da gramática é sua regra. Em geral, uma linguagem pode ser representada por diversas regras gramaticais diferentes. A regra é obtida por um par ordenado (u, v) , geralmente escrito na forma $u \rightarrow v$ (u implica em v , onde u e v são sentenças de alfabetos). Além disso, é o conjunto de pares ordenados que fornecem a estrutura da gramática. Uma gramática G pode ser utilizada de duas maneiras:

⁵ Também conhecido como dicionário na área da visão computacional.

⁶ Também conhecida por palavra, cadeia de caracteres e *string*.

- Generativo (geradores): cria sequências de símbolos terminais, onde é gerada uma sentença na linguagem da gramática; e,
- Analítico (reconhecedores): procedimento que indica se uma determinada sentença pertence ou não a uma linguagem. Dada uma sentença, deseja-se determinar:
 - i. Se a sentença foi gerada pela gramática G ; e,
 - ii. A estrutura da sentença.

Verificar se uma sentença pertence a uma linguagem é um dos principais problemas da área de linguagens formais (Menezes, 2011). Logo, para cada tipo de linguagem existe um reconhecedor associado como, por exemplo, reconhecedor baseado em autômatos.

Autômatos finitos

Um autômato finito é um reconhecedor de linguagens regulares. Os reconhecedores recebem uma sequência w como entrada e, baseando-se em uma linguagem L , responde “sim” se $w \in L$ e “não” caso contrário. Os autômatos podem ser classificados em autômatos finitos determinísticos (AFD) e, autômatos finitos não-determinísticos (AFND). Formalmente, um AFD é definido como uma 5-tupla $(\Sigma, Q, \delta, q_0, F)$, em que:

- a) Σ é um conjunto finito de símbolos, chamados de alfabeto de um autômato;
- b) Q é um conjunto finito de possíveis estados do autômato;
- c) δ são as regras ou funções de transição:

$$\delta: Q \times \Sigma \rightarrow Q$$

o qual é uma função. Supondo que a função é definida para um estado p e um símbolo “a”, resultando no estado q , então:

$$\delta(p, a) = q$$

- d) q_0 é o estado inicial do autômato, onde $q_0 \in Q$; e,

e) F é um conjunto de estados finais ou de aceitação, onde $F \subseteq Q$.

Dado um conjunto de símbolos $w = a_1, a_2, \dots, a_n$, o autômato aceita w se, e somente se, existir uma sequência de estados $q_0, q_1, \dots, q_n \in Q$ com:

- 1) $q_{i+1} = \delta(q_i, a_{i+1})$;
- 2) $q_n \in F$.

Análise sintática

A análise sintática, também conhecida por *parsing*, é o processo que analisa uma sentença para determinar sua estrutura gramatical, de acordo com regras determinadas por uma gramática. Essa análise precisa comprovar que a sequência de símbolos, retornados pelo analisador léxico⁷, cumpre regras gramaticais e pode gerar uma árvore estrutural. Em geral, a análise sintática é utilizada na computação pelos compiladores e interpretadores a fim de verificar se um código-fonte está certo ou errado. O analisador sintático determina se uma entrada de dados pode ser derivada de um símbolo inicial a partir das regras de uma gramática e, isso pode ser feito de duas maneiras:

- Descendente (*top-down*): o analisador inicia pelos maiores elementos e os divide em elementos menores, ou seja, ele tenta deduzir as produções da gramática a partir do nó raiz; e,
- Ascendente (*bottom-up*): o analisador localiza os elementos mais básicos e, na sequência, os elementos maiores que contêm os elementos mais básicos e, assim, sucessivamente. Dessa forma, é construída a árvore gramatical das folhas à raiz.

Pela análise léxica é obtido um grupo de símbolos, para que o analisador sintático receba os dados de entrada e construa uma estrutura de dados. Em geral, a estrutura é algum tipo de hierarquia, para fornecer uma representação estrutural de informação e verificar se a entrada de dados corresponde às regras gramaticais. Ao encontrar um símbolo que não pertence às regras de uma gramática é gerado um erro

⁷ Etapa precedente à análise sintática. O analisador léxico verifica a validade de determinado alfabeto, convertendo uma sentença em um conjunto de símbolos.

e o analisador sintático precisa se recuperar desse tipo de erro, que muitas vezes pode ser recorrente.

Recuperação de erros em analisadores sintáticos

A recuperação de erros (ou tratamento de erros) é um tópico relevante na área de reconhecimento sintático, uma vez que a aplicação precisa se adaptar e saber o que fazer caso analise uma sentença que contenha erros sintáticos 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. O recuperador de erros sintáticos possui metas a serem estabelecidas como, relatar a presença de erros, se recuperar de erros com a finalidade de detectar outros erros na sequência e, se possível, reparar os erros.

Como os analisadores sintáticos avaliam a gramática de linguagens, pode ser que símbolos esperados não sejam encontrados. A forma adotada neste trabalho para se recuperar de um erro sintático é descartar o símbolo ao encontrá-lo e contar como um erro na análise da sentença. Na sequência, analisa-se novamente, até encontrar outro símbolo que pertença ao subconjunto de símbolos que permita a interpretação e continuação da análise.

Inferência gramatical

A inferência gramatical, também conhecida por indução gramatical, refere-se ao processo de aprendizagem de gramáticas e linguagens a partir de um conjunto de amostras (Russel e Norvik, 2003; Sakakibara, 1995; Higuera, 2005). Inferir significa obter uma resposta baseando-se em fatos ocorridos, analisando apenas uma porção amostral dos fatos e não todo o conjunto. De modo geral, a partir de um conjunto de exemplos com características em comuns, deseja-se obter uma gramática que possa explicar como esses exemplos são gerados. A técnica é aplicada com o objetivo de identificar padrões em uma parte do conjunto de amostras com o objetivo de

representar o conjunto todo. Espera-se, assim, que a gramática obtida reconheça outros exemplos não apresentados.

A área de inferência gramatical é dividida em métodos de enumeração e construção. Na primeira, todas as possíveis gramáticas de uma classe são listadas para escolher aquelas que podem ser utilizadas para gerar sentenças de amostras finitas, ou seja, se assume que é possível enumerar as hipóteses. Na segunda, as gramáticas são construídas diretamente pela base de treinamento (Sakakibara, 1995), em que são construídas novas hipóteses a cada novo exemplo. Neste trabalho é usado o método de construção, visto que não é possível enumerar todas as sentenças possíveis, pois cada imagem do problema em questão possuirá uma sentença diferente devido a variação de luz, ângulo da foro, etc.

A inferência gramatical é estudada por pesquisadores de diversas áreas em virtude de sua possibilidade de solução em tarefas de reconhecimento de padrões estruturais (Higuera, 2005), tais como: teoria da informação, linguagens formais, teoria dos autômatos, aquisição de linguagem, linguística computacional, aprendizagem de máquina, reconhecimento de padrão, redes neurais, dentre outras. Por essa razão, há aplicações direcionadas ao processamento de linguagem natural, análise de sequência biológica na bioinformática e, verificação e testes de software. Além disso, considerando que os padrões correspondem a símbolos não-terminais em uma gramática, existem vários tipos de problemas que podem ser resolvidos. Por exemplo, identificar regularidades em sequências biológicas, como no DNA e no RNA (Sakakibara, 2005; Searls, 2002; Durbin *et al.*, 1998), realçar padrões em contagens musicais (Cruz-Alcázar, 1998), descobrir propriedades de linguagem de textos, reconhecimento de voz (Miclet *et al.*, 2004), dentre outros.

Dentro da área de bioinformática, onde o objetivo é usar técnicas matemáticas e computacionais para a resolução de problemas na biologia molecular, vários modelos gramaticais foram aplicados no reconhecimento de sequências biológicas, por exemplo: gramáticas estocásticas regulares (Kashiwabara, *et al.*, 2007; Ross, 2002), gramáticas livres de contexto (Unold, 2007), gramáticas estocásticas livres de contexto (Sakakibara, 2005; Knudsen *et al.*, 2003), gramáticas de cláusulas definidas (*Definite Clause Grammars*) (Searls, 2002) e TAGs (*Tree Adjoining Grammars*) (Cai *et al.*, 2003). Também, um exemplo de processamento de linguagem natural por meio

da inferência gramatical pode ser encontrado em Geertzenl *et al.* (2004). Propostas híbridas também são encontradas na literatura, tal como a de Garcia-Gomez *et al.*, (2004), além de modelos com gramáticas *fuzzy* por Mordeson *et al.* (2002).

Por outro lado, trabalhos interessantes ainda podem ser desenvolvidos (Matsuno, 2006), tais como: dependência das características de conjunto de exemplos; quantidade mínima de exemplos para realizar o treinamento do modelo de dados; aumentar o processamento conforme a quantidade de exemplos; desenvolver algoritmos com complexidade polinomial; eliminação de não-determinismos; geração de gramáticas ambíguas; e, dificuldade em identificar e demonstrar problemas encontrados.

Para exemplificar o funcionamento de inferência gramatical e da produção de gramática, a seguir é apresentado brevemente o algoritmo sequitur, apenas para explanação, pois ele não é usado na aplicação desta tese.

Sequitur⁸ (Nevill-Manning e Witten, 1997) é um algoritmo que infere uma estrutura hierárquica de uma sequência de símbolos. Ele substitui frases repetidas baseando-se em uma regra gramatical que gera a frase, realizando esse processo de forma recursiva. O resultado é a representação hierárquica da sequência original, o que oferece uma visão em sua estrutura léxica. Inicialmente foi idealizado para trabalhar com compressão de textos, porém também é usado para reconhecer estruturas de documentos. Os autores defendem a ideia de que uma boa gramática é uma gramática compacta. Para isso, o algoritmo impõe duas restrições para reduzir o tamanho da gramática, produzindo uma estrutura como o seu subproduto:

- Nenhum par de símbolos adjacentes deve aparecer mais de uma vez na gramática; e,
- Toda regra deve ser usada mais de uma vez.

A primeira restrição garante que se um par de símbolos AB aparecer mais de uma vez na gramática, então deve-se criar uma regra $C \rightarrow AB$ e substituir todas as ocorrências de AB por C, tal que C implica em AB. Logo, sendo verdadeira a afirmação antes da implicação, então também será verdadeira a afirmação a sua direita. A

⁸ O projeto implementado em *ObjectPascal* pode ser encontrado em: <https://code.google.com/p/sequitur/>

segunda restrição garante que regras inúteis não sejam criadas, ou seja, se a partir de uma gramática é encontrada a regra $D \rightarrow DE$, mas essa regra aparece apenas uma vez na gramática, não é viável manter a regra, pois ela não é utilizada. Assim, o algoritmo aplica essas duas restrições em uma busca gulosa⁹ percorrendo todo o texto, construindo de forma incremental uma gramática.

O Sequitur pode processar 50.000 símbolos por segundo e já foi aplicada a uma vasta gama de sequências reais. Por outro lado, uma característica crítica nessa proposta é que ela gera apenas a sentença informada. Modificações em cima da abordagem apresentada foram propostas por Charikar *et al.* (2005) em que, a construção da gramática seja necessariamente a mínima. Além disso, há a possibilidade de estender a abordagem para construir uma gramática a partir de várias sentenças e não apenas de uma. A composição de duas ou mais gramáticas dependem de como as gramáticas são geradas (Wintner, 2003).

Um exemplo da geração do conjunto de regras pelo algoritmo Sequitur para a sentença “abcdbcabc” pode ser visualizada na Tabela 1. Cada símbolo é enviado como entrada de dado para composição (da esquerda para a direita) e análise da sentença. Até o passo 6, os símbolos são enviados e anexados na sentença, sem que haja repetição com mais de um símbolo na sentença. No passo 6, o Sequitur percebe que os símbolos “bc” se repetem e cria a regra $A \rightarrow bc$. Logo, é substituído as duas ocorrências na sentença por A, uma vez que A implica na igualdade de bc. Nos passos 7 e 8 envia e anexa dois símbolos sem constatar repetição. E, por fim, ao receber um novo símbolo no passo 9 é encontrada duas repetições, onde “bc” ocorre novamente e é substituído por “A” e, em seguida, o algoritmo encontra a repetição de “aA” duas vezes e substitui por uma nova regra B criada, tal que $B \rightarrow aA$. Como resultado final, foi criada duas regras para a sentença.

Passo	Entrada	Sentença	Gramática gerada	Observações
1	a	a	$S \rightarrow a$	
2	b	ab	$S \rightarrow ab$	
3	c	abc	$S \rightarrow abc$	

⁹ Algoritmo para resolver problemas de otimização, onde é realizada sempre a escolha que parece ser a melhor no momento, esperando que esta escolha conduza a uma solução ótima.

4	d	abcd	$S \rightarrow abcd$	
5	b	abcdb	$S \rightarrow abcdb$	
6	c	abcdbc	$S \rightarrow abcdbc$ $S \rightarrow aAdA; A \rightarrow bc$	bc ocorre duas vezes
7	a	abcdbca	$S \rightarrow aAdAa; A \rightarrow bc$	
8	b	abcdbcab	$S \rightarrow aAdAab; A \rightarrow bc$	
9	c	abcdbcabc	$S \rightarrow aAdAabc; A \rightarrow bc$ $S \rightarrow aAdAaA; A \rightarrow bc$ $S \rightarrow BdAB; A \rightarrow bc; B \rightarrow aA$	bc ocorre novamente aA duas vezes

Tabela 1. Exemplo passo a passo do funcionamento do algoritmo de inferência gramatical Sequitur.

Scale Invariant Feature Transform - SIFT

Para classificar imagens, primeiramente é preciso identificar suas características e descrevê-las, para posteriormente compará-las com uma nova imagem e verificar se os padrões de cada imagem se assemelham. Isso permite classificar uma imagem em um grupo de imagens já existente. Para tanto, é possível usar um extrator de características em imagens. O extrator pode determinar aspectos como, forma, dimensão, cor, textura, entre outros. O extrator descreve as características de um padrão ou objeto. Uma das maneiras de descrever e identificar características pertinentes em imagens é utilizando o algoritmo SIFT (Lowe, 2004).

A detecção e descrição de características de imagens tem sido extensivamente empregada na área de reconhecimento de imagens. Atualmente, o *Scale Invariant Feature Transform* – SIFT (Lowe, 2004) e o *Speeded Up Robust Features* - SURF (BAY *et al.*, 2006) são os dois detectores e descritores mais importantes. Eles 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, pois as configurações da imagem podem influenciar na quantidade de pontos extraídos. O descritor descreve o ponto de interesse por meio de um vetor numérico que possui

valores referentes a direção das variações ao redor do ponto de interesse. Esses extratores são extremamente relevantes para tarefas de reconhecimento de padrões em imagens, pois buscam ser invariantes a rotação, escala da imagem e mudança de iluminação. Os pontos de interesse tendem a ser representativos em uma imagem, de modo que possam ser diferenciados ou reconhecidos em uma outra imagem. A Figura 1 mostra uma imagem original de um peixe e a imagem com os vários pontos de interesse detectados pelo algoritmo SIFT. Conforme descrito na documentação da biblioteca OpenCV (2017)¹⁰, ao exibir os pontos de interesse, o algoritmo desenha um círculo com o tamanho de cada ponto, apontando também a sua orientação.



Figura 1. Exemplo de detecção de pontos de interesse utilizando o algoritmo SIFT.

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

¹⁰ É uma biblioteca multiplataforma para o desenvolvimento de aplicativos na área de visão computacional. Disponível em: <http://opencv.org/>

Histograma de palavras visuais (BOVW)

O histograma de palavras ou, simplesmente, *Bag of Words* – BOW (Sivic *et al.*, 2003; Csurka *et al.*, 2004), é um algoritmo muito utilizado na área de reconhecimento de textos. Essa técnica extrai do texto um conjunto de palavras desconsiderando sua estrutura gramatical e, até mesmo, sua ordenação no texto, apenas mantendo sua frequência de uso. Esse conjunto de palavras é utilizado para distinguir um texto e identificar seu gênero de modo que seja feita a classificação do texto.

O histograma de palavras visuais é uma técnica bastante utilizada em tarefas de classificação. Ao usá-la, basicamente é contada as ocorrências das palavras visuais de uma determinada imagem e feito a distribuição de frequência dessas palavras, gerando um histograma. Assim, para determinar a classe da imagem, compara-se o seu histograma com os 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 certo conjunto de palavras visuais.

Para se trabalhar com imagens, o algoritmo BOW passa a ser denominado de histograma de palavras visuais ou, *Bag of Visual Words* - BOVW¹¹. O conceito é o mesmo, porém agora o vetor de atributos é extraído de uma imagem. Para a extração desses atributos é necessário o uso de um algoritmo auxiliar, como o SIFT. Assim, o BOVW pode usar os descritores dos pontos de interesse extraídos pelo SIFT para determinar a palavra visual e associá-la a um ponto de interesse. Cada palavra visual é obtida pelo algoritmo k-médias, que representa um grupo de pontos de interesse semelhantes (Csurka *et al.*, 2004). Assim, na classificação de imagens, não é utilizado todo o processo do histograma de palavras visuais, e sim uma etapa desse processo, o k-médias, para determinar os centroides de cada grupo de pontos de interesse. Após a construção do alfabeto (palavras visuais) pelo k-médias, os pontos de interesse podem ser representados como palavras visuais, pois cada ponto de interesse terá uma palavra visual associada a ela.

Apesar dos avanços recentes e de resultados promissores, o poder descritivo dessa técnica é limitado, pois descarta informações espaciais das palavras visuais na imagem. Essas informações podem ser características importantes em tarefas de

¹¹ Alguns trabalhos utilizam a denominação de *bag-of-features* ou *bag-of-keypoints* referindo-se ao mesmo método.

classificação de imagens, pois quando são apenas 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 resultar na melhora do resultado da classificação.

Reconhecimento sintático de padrões

O reconhecimento sintático de padrões teve seu início nos anos 50 com técnicas probabilísticas. Posteriormente, surgiram técnicas baseadas na abordagem estrutural utilizando modelos sintáticos. O reconhecimento sintático de padrões destaca a descrição estrutural de padrões. Trata-se de uma abordagem em que os padrões de um objeto são tratados sob uma perspectiva hierárquica. Assim, o reconhecimento sintático de padrões pode ser usado no lugar do reconhecimento estatístico quando existir uma estrutura clara nos padrões. As informações estruturais descrevem cada padrão, por isso facilitam a descrição e a classificação de um objeto.

Essa é uma abordagem matemática que tem como fundamento a teoria das linguagens formais e autômatos. Sua importância deve-se às diversas ocorrências relacionadas a vida humana as quais possuem padrões, tornando-se uma alternativa importante para os sistemas inteligentes. Nesse tipo de sistema, muitas vezes as informações significativas em um padrão podem não consistirem em valores numéricos, mas em uma conexão de características que produzem informações estruturais relevantes para facilitar a descrição do objeto. Logo, é necessário extrair e quantificar as informações estruturais que descrevem cada padrão para avaliar a similaridade estrutural. Assim, o processo de reconhecimento sintático de padrões permite associar um padrão a uma classe específica, além de descrever características de um padrão que não o tornaram aceitável para classificá-lo em uma outra classe.

Normalmente, a descrição estrutural de padrões estabelece uma descrição hierárquica que é construída a partir de padrões mais simples, as quais são

construídas por sub-padrões ainda mais simples. Assim, os níveis mais baixos são os elementos mais elementares, extraídos a partir de dados de entrada, denominados de primitivas (Fu, 1982). Uma analogia formal pode ser traçada como um texto, em que é constituído por vários parágrafos; um parágrafo é constituído por frases; uma frase é constituída por palavras que, por sua vez, são constituídas por letras. A ideia principal é que um conjunto de padrões complexos possam ser descritos por meio de uma quantidade finita de primitivas e regras gramaticais. Contudo, as regras geradas por uma gramática, fornecem uma descrição da estrutura dos objetos e os padrões conforme suas primitivas (Oommen *et al.*, 1998). Nesse caso, diferenças nas estruturas entre as classes são codificadas como gramáticas diferentes. Logo, a gramática de cada classe de padrões deve ser inferida a partir dos casos de treinamento. Uma vantagem em trabalhar com modelos estruturais é em virtude da técnica fornecer uma descrição de como os padrões são gerados a partir de primitivas que podem ser repetidas ou que seguem determinadas ordens.

Na Figura 2 é possível visualizar as principais etapas de um sistema de reconhecimento sintático de padrões típico, onde o sistema é dividido em duas partes principais: o treinamento e o reconhecimento (Banks, 1990; Schalkoff, 1992). O primeiro passo é inerente ao treinamento e consiste em estimar um modelo para cada uma das classes a partir de exemplos. Nesse caso, é feita a seleção de primitivas, inferência gramatical e construção do autômato. O segundo passo é referente ao reconhecimento de uma instância e consiste em determinar uma classe a um padrão de entrada desconhecido. A classe que obtiver o maior valor de comparação dentre as outras classes é considerada como a mais provável para representar o novo padrão desconhecido. Nessa etapa, é feito o pré-processamento do padrão ou imagem, segmentação, reconhecimento de primitivas, construção da representação do padrão e a análise sintática do padrão recebido como dado de entrada.

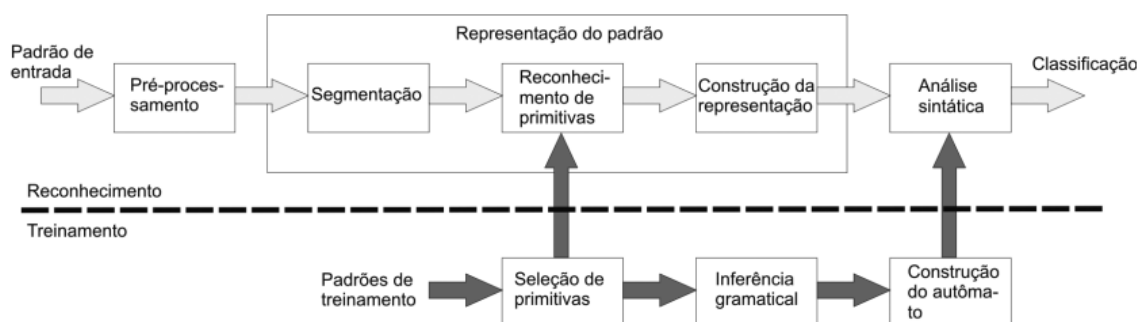


Figura 2. Etapas necessárias para um sistema de reconhecimento sintático de padrões típico (PIO, 2006).

Reconhecimento sintático de padrões aplicado em problemas de classificação de imagens

Um estudo recente realizado por Pistori *et al.* (2013) traz resultados interessantes ao utilizar a inferência gramatical no problema de classificação de imagens. Os autores apresentam uma nova estratégia para representar uma imagem por meio de uma sentença de modo que técnicas de inferência gramatical possam ser usadas em problemas de reconhecimento de imagens. Um ponto explorado nesse trabalho é sobre como representar as imagens baseando-se nas suas informações extraídas. A proposta explora o padrão textual da sentença até estratégias que tentam explicitamente preservar as informações espaciais. O trabalho de Pistori *et al.* (2013) é a base e o principal motivador deste trabalho.

Pistori *et al.* (2013) exploram os pontos de interesse e palavras visuais como alternativa de facilitar o reuso de algoritmos que processam sentenças. A ideia principal da proposta é impor uma ordem para os n pontos de interesse detectados, produzindo uma sentença de tamanho n . Sobretudo, é analisado o impacto do tamanho do alfabeto, uma vez que também é modificada a ordenação dos pontos de interesse para gerar a gramática e realizar a classificação. No trabalho, dois grupos de imagens foram testados, um grupo com imagens criadas manualmente e outro grupo com um banco de imagens reais, a fim de explorar os parâmetros da nova estratégia e ter um primeiro resultado do seu desempenho em comparação com técnicas populares de aprendizagem de máquina. Essa abordagem utiliza uma estratégia de recuperação de erros simples, uma vez que alguns símbolos do alfabeto

podem não estar disponíveis em um determinado conjunto de imagens usadas para treinar uma classe, logo o AFD não teria uma transição relacionada a esse símbolo. Nesse caso, durante a execução do autômato, o símbolo é ignorado e o autômato segue lendo o próximo símbolo sem alterar o estado atual. A estratégia é utilizada para que o autômato sempre leia todos os símbolos da sentença de entrada e o número de erros encontrados na sentença. Após a sua leitura total é utilizado para classificar uma nova a imagem. Os resultados dos experimentos realizados mostraram ser encorajadores para que a área continue sendo explorada.

Posteriormente, Ribas *et al.* (2014) faz uma replicação da abordagem proposta por Pistori *et al.* (2013), porém buscando ampliar os experimentos com um tamanho de alfabeto maior e utilizando a abordagem supervisionada e não-supervisionada para a construção do alfabeto. Nesse trabalho, é usado no experimento um banco de imagens de cenas classificadas por seres humanos (Martinez-Gomez *et al.*, 2012). Além disso, os experimentos foram ainda mais motivadores a continuar a exploração dessa área.

Abordagens propostas para o reconhecimento de espécies de peixes

Existem alguns trabalhos correlatos à abordagem proposta nesta pesquisa, como o de Rova *et al.* (2007), em que os autores propõem uma solução baseando-se principalmente em estabelecer correlações entre modelos paramétricos, os quais combinam informações de textura e forma referente a duas espécies de peixes. O banco de imagens foi composto de 2 espécies de peixes com 320 imagens. Para o processo de classificação, foram utilizadas máquinas de vetores de suporte e obtida a exatidão de 90% no melhor caso.

Semelhantemente, Lee *et al.* (2003) propõem uma estratégia de reconhecimento e classificação direcionada na aquisição de informações das formas de nove espécies de peixes. Nessa abordagem, são propostos descritores compostos por alguns pontos específicos presentes no peixe que conjuntamente definem o seu contorno. Para a classificação foi utilizado o cálculo da distância Euclidiana entre os

pontos. Entretanto, não foi especificado nenhum experimento criterioso a respeito da exatidão dessa metodologia.

Outro trabalho é apresentado por Nery *et al.* (2005), onde os autores utilizam descritores mais robustos que os utilizados no trabalho de Lee *et al.* (2003), uma vez que é proposto o uso de diferentes características, sendo elas: forma, cor, textura e tamanho de espécies, além de utilizar um peso para cada uma dessas características, a fim de utilizar como informação *a priori* por um classificador bayesiano. O banco de imagens foi composto de 225 imagens divididas em 6 espécies de peixes e com variação no número de imagens por espécie. Nesse trabalho, foram apresentadas diferentes precisões de classificação baseando-se em combinações entre as características, porém o melhor resultado obtido baseando-se nas 6 espécies propostas foi de 81%.

METODOLOGIA

Esta seção apresenta os principais passos seguidos para a execução deste trabalho. Cada um dos objetivos específicos são detalhados a seguir.

Construção dos bancos de imagens de peixe

O objetivo dos bancos de imagens criados é conter imagens de espécies de peixe capturadas a partir de celulares, uma vez que um banco que contenha imagens que mostra apenas o peixe recortado não foi encontrado na literatura. Uma motivação para a criação de bancos de imagens a partir do uso de celulares é para ter futuramente uma aplicação em que o usuário possa enviar imagens para identificar a espécie (classificação) e, como consequência, essas imagens serão incrementadas no banco de imagem, tendo uma construção colaborativa do banco de imagens. É importante para a classificação, várias imagens de cada espécie em diferentes ângulos e distâncias, com variação de cor, textura, luminosidade, tamanho e forma. Algumas espécies se camuflam para melhor capturarem suas presas e outras se escondem na areia, o que pode mudar o tom da pele do peixe para que as presas não notem sua presença no ambiente (Hansen, 1998). Logo, foi relevante considerar essa informação no processo de obtenção de imagens, pois eles também se diferenciam por essas características.

Os bancos de imagens foram criados, pois não foram encontrados na literatura algum semelhante com a finalidade de mostrar o peixe na imagem. Então, um dos objetivos iniciais foi capturar imagens do Aquário Pantanal, de Campo Grande, Mato Grosso do Sul, porém as obras não foram finalizadas a tempo de capturar as imagens nesse local. Por isso, os bancos de imagens foram compostos de peixes coletados

em aquários. Algumas espécies foram obtidas no Aquário Municipal de Toledo¹², estado do Paraná, outras em uma casa de venda de peixes ornamentais, na loja Planeta Real¹³ em Campo Grande, Mato Grosso do Sul, e na loja Peixinho Dourado¹⁴ de Rondonópolis, Mato Grosso. As fotos foram tiradas de peixes nadando livremente no aquário em frente ao aquário. Não foram utilizadas câmeras submersas. A distância da câmera ao peixe foi variável, de 10 centímetros a 1 metro, com iluminação natural do ambiente. Por isso, interferências como alteração na cor, iluminação desigual, reflexo do vidro, fundo diferente, sedimentos na água, plantas subaquáticas, entre outras, podem ser encontradas em algumas imagens.

A Figura 3 ilustra o ambiente interno do Aquário Municipal de Toledo. Como muitas das imagens foram tiradas com vários peixes na mesma foto, elas foram recortadas para manter apenas um peixe centralizado em cada foto, conforme ilustra a Figura 4. Na Figura 4 (A) é apresentada uma imagem original e na Figura 4 (B) é apresentada a imagem com recortes realizados em volta do peixe. Todas as imagens do banco foram recortadas e estão desse modo.



Figura 3. Aquário Municipal de Toledo - Paraná.

¹² Maiores detalhes sobre o aquário podem ser encontrados em: <http://www.toledo.pr.gov.br/portal/meio-ambiente/aquario-municipal-romolo-martinelli/>

¹³ Avenida Afonso Pena, 1919 - Centro, Campo Grande - Mato Grosso Do Sul. Telefone: (67) 3025-4942.

¹⁴ Avenida Marechal Rondon ,1338, Centro, Rondonópolis, Mato Grosso.

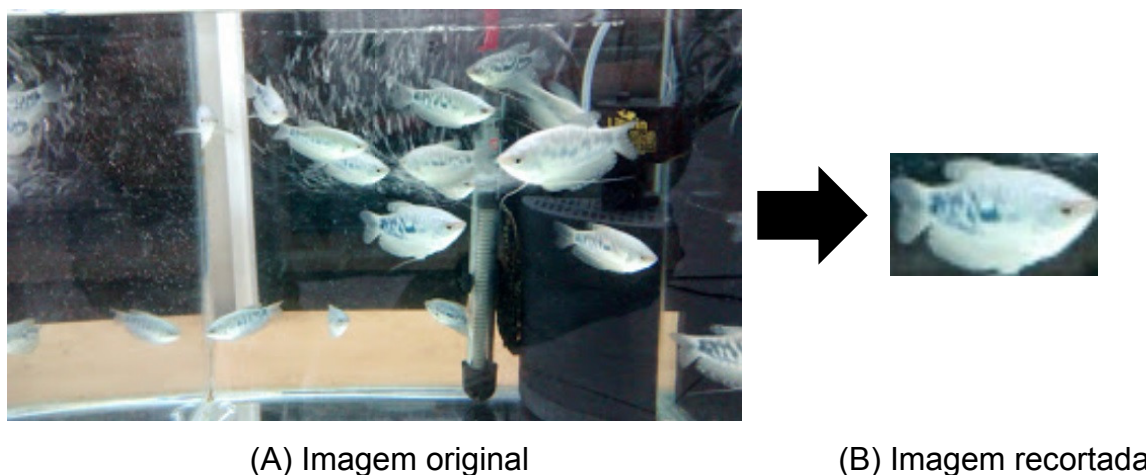


Figura 4. Exemplo de uma imagem original e uma imagem recortada para composição do banco de imagens.

A especificação de cada banco de imagens pode ser vista na Tabela 2. O banco de imagens “Aquario10e” foi o primeiro a ser criado e possui 596 imagens com 10 espécies de peixes diferentes (variando entre 41 e 101 imagens por espécie). Em seguida, foi criado o banco “AquarioSeg10e”, que diferencia-se por conter 10 imagens segmentadas de cada espécie do banco “Aquario10e”. As imagens foram selecionadas aleatoriamente e segmentadas uma a uma em um programa de edição de imagem. Por fim, foi criado o banco de imagens “Aquario28e40i”. Esse possui as espécies do “Aquario10e”, com adição de 18 novas espécies, totalizando 28 espécies. As imagens do banco “Aquario28e40i”, com os nomes populares e científicos, podem ser vistos na Tabela 3.

Nome do banco	Quantidade de espécies	Imagem Segmentada	Imagens por espécie	Total de imagens
Aquario10e ¹⁵	10	Não	Variado	596
AquarioSeg10e ¹⁶	10	Sim	10	100
Aquario28e40i ¹⁷	28	Não	40	1120

Tabela 2. Características dos bancos de imagens criados.

¹⁵ Disponível em: <https://goo.gl/GkCZRS>

¹⁶ Disponível em: <https://goo.gl/LWx6VS>

¹⁷ Disponível em: <https://goo.gl/A8mhBg>

Imagem	Espécie	Imagem	Espécie
	Acará bandeira (<i>Pterophyllum scalare</i>)		Oscar (<i>Astronotus ocellatus</i>)
	Acará Bandeira Marmorizado (<i>Pterophyllum scalare</i>)		Oscar Albino (<i>Astronotus ocellatus</i>)
	Acará Disco (<i>Symphysodon aequifasciatus</i>)		Pacu (<i>Piaractus mesopotamicus</i>)
	Barbus Ouro (<i>Puntius sachsii</i>)		Peixe Palhaço (<i>Amphiprion frenatus</i>)
	Barbus Sumatra (<i>Puntigrus tetrazona</i>)		Peixe Papagaio (<i>Sparisoma chrysopteron</i>)
	Beta (<i>Betta splendens</i>)		Paulistinha (<i>Danio rerio</i>)
	Carpa (<i>Cyprinus carpio</i>)		Piau Três Pintas (<i>Leporinus friderici</i>)
	Carpa Média (<i>Cyprinus carpio</i>)		Platy Laranja (<i>Xiphophorus maculatus</i>)
	Dourado (<i>Salminus brasiliensis</i>)		Platy Ruby (<i>Xiphophorus maculatus</i>)
	Kinguio (<i>Carassius auratus</i>)		Platy Sangue (<i>Xiphophorus maculatus</i>)
	Kinguio Cometa Calico (<i>Carassius auratus</i>)		Telescópio (<i>Carassius auratus</i>)
	Kinguio Karraco (<i>Carassius auratus</i>)		Tetra Negro (<i>Gymnocorymbus ternetzi</i>)
	Mato Grosso (<i>Hyphessobrycon eques</i>)		Tricogaster (<i>Trichogaster trichopterus</i>)
	Molinésia Preta (<i>Poecilia shenops</i>)		Tucunaré (<i>Cichla ocellaris</i>)

Tabela 3. Espécies de peixes do banco de imagens “Aquario28e40i” com o respectivo nome popular e científico de cada espécie.

Foram constituídos 3 diferentes bancos de imagens porque as fotos foram capturadas em datas diferentes. Inicialmente, foram gerados o “Aquario10e” e o

“AquarioSeg10e”, para iniciar os experimentos e, posteriormente, foi gerado o banco “Aquario28e40i” com 28 espécies. Por essa razão, o Capítulo 1 usa apenas os bancos “Aquario10e” e “AquarioSeg10e” e, o Capítulo 2 e Capítulo 3 usam do banco “Aquario28e40i”.

Proposta de um algoritmo para a classificação de espécies de peixes utilizando reconhecimento sintático de padrões

Esta tese se inspira e explora a abordagem que trata a representação de imagens como sentenças usada por Pistori *et al.* (2013). Nesta tese, as sentenças são geradas a partir de agrupamento baseado em estrutura hierárquica, o que diferencia do trabalho de Pistori *et al.* (2013).

Inicialmente, a abordagem proposta detecta pontos de interesse, utilizando o algoritmo *Scale Invariant Feature Transform* (SIFT). A partir dos pontos de interesse detectados é realizada uma limitação da quantidade de pontos baseado no conceito de cauda longa (*long tail*) e Lei de Zipf's. O efeito cauda longa, por sua vez, é previsto pela Lei de Zipf's (Zipf, 1949) (Newman, 2006). Essa lei está associada ao Princípio de Pareto (Newman, 2006), também conhecido pela regra dos 80-20. Essa regra determina que 80% das causas são responsáveis por 20% dos efeitos, enquanto que o restante dos 20% das causas são responsáveis por 80% dos efeitos (Brynjolfsson *et al.*, 2006) (Koch, 2000). À vista disso, podemos associar a regra dos 80-20 com os pontos de interesse e deduzir que muitos deles não são representativos, devido a estarem na cauda longa, enquanto que apenas poucos são representativos, o que justifica usar apenas a parte mais representativa deles na análise para a classificação. Nesse contexto, a filtragem e descarte de pontos de interesse pode tornar a descrição das classes mais relevante para a classificação.

Posteriormente, os pontos de interesse são descritos usando o descritor Opponent-SIFT. Em seguida, o alfabeto é construído pela aprendizagem supervisionada, usando o conjunto de descritores dos pontos de interesse. Na sequência, rotula-se cada um deles (associação do ponto de interesse à uma palavra visual). Como os resultados foram sempre superiores na aprendizagem

supervisionada, na versão final da proposta, Capítulo 3, foi utilizada apenas a aprendizagem supervisionada.

O processo de construção do alfabeto pode ser melhor visualizado na Figura 5, onde é exemplificado a construção pela aprendizagem supervisionada e não supervisionada (usada nos experimentos como forma de comparação). Na aprendizagem supervisionada, no passo(i) é realizada a detecção, a limitação e a descrição de pontos de interesse. No passo (ii) os centroides são gerados separadamente para cada classe (espécie) aplicando o algoritmo k-médias n vezes, tal que, n é o número de classes. Por exemplo, se existem 3 classes e o tamanho do alfabeto é 9 ($k = 9$), cada classe gerará 3 centroides. Caso o tamanho do alfabeto não seja múltiplo do número de classes, a quantidade remanescente do alfabeto é atribuída aleatoriamente para a(s) classe(s). No passo (iii) os centroides de cada classe são agrupados para compor um único conjunto com todos os centroides. Por fim, rotula-se cada ponto de interesse com a palavra visual do centroide mais próximo, conforme ilustrado no passo (iv). Se ocorrer a sobreposição de centroides é atribuído um centroide aleatório ao ponto de interesse. Diferentemente, na abordagem não supervisionada, após detectar e descrever os pontos de interesse (passo i), os centroides são gerados considerando todos os descritores de todas as classes, aplicando o algoritmo k-médias uma única vez, conforme ilustra o passo (ii). Em seguida, no passo (iii) realiza-se a rotulação dos pontos de interesse.

Com a aprendizagem supervisionada, cada classe pode gerar centroides mais característicos baseando-se em seus próprios descritores (apenas exemplos positivos da classe). Na abordagem não supervisionada, por sua vez, os centroides são gerados uma única vez com todos os descritores juntos, não considerando a separação de classe.

Na sequência, é realizada a geração de sentenças por agrupamento hierárquico. Para cada imagem é gerada uma sentença. Esse agrupamento une as palavras visuais baseando-se na posição espacial (x , y) entre cada par de pontos de interesse, até formar uma árvore binária completa. A distância entre os pontos é verificada entre cada par de pontos entre todos os existentes. Logo, a construção da árvore binária é feita baseando-se em cada de par de pontos mais próximos. Para tal

agrupamento é usado o algoritmo *linkage*¹⁸ implementado no Matlab. Esse algoritmo usa a informação de distância da posição espacial (x, y) do ponto de interesse a fim de determinar a proximidade de pontos entre si. A medida de distância usada é a distância Euclidiana.

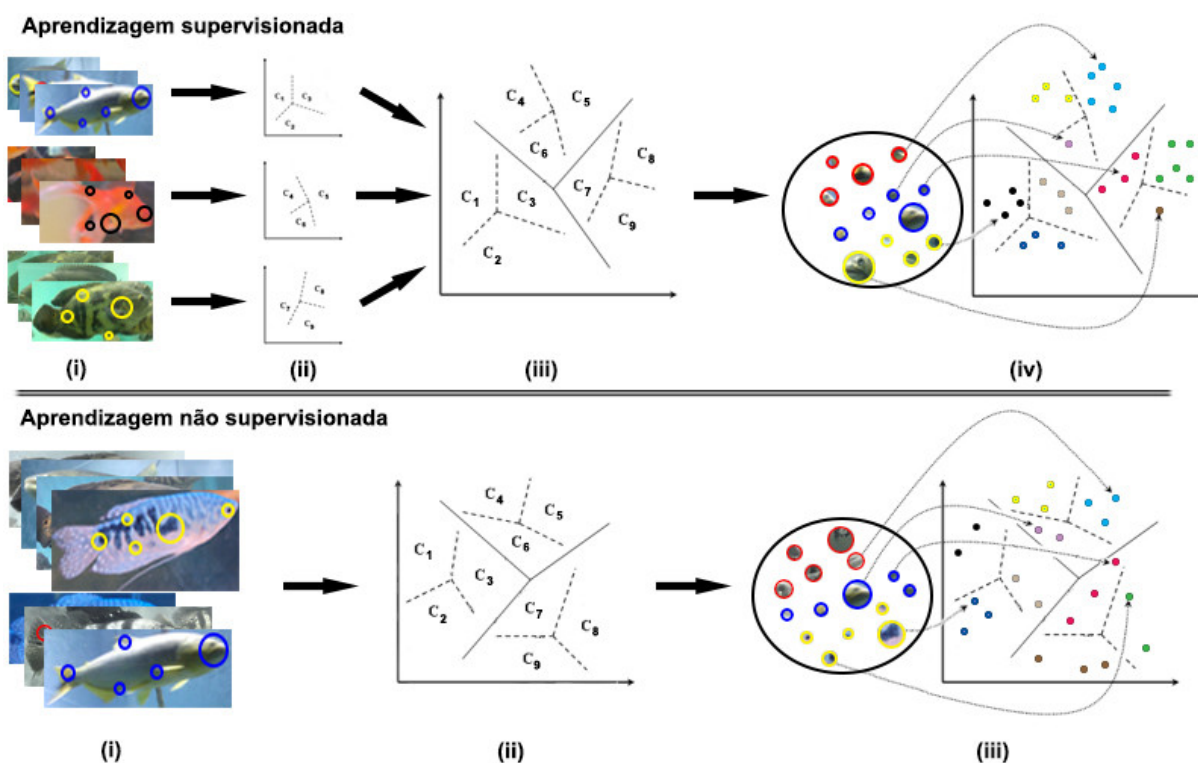


Figura 5. Ilustração da construção do alfabeto pela aprendizagem supervisionada e não supervisionada.

Por fim, é realizada a aprendizagem da gramática para a classificação de imagens utilizando o algoritmo de inferência gramatical *k-testable*. A variável k do algoritmo *k-testable* significa uma quantidade k de caracteres anteriores ao caractere atual, analisado(s) na inferência gramatical. O valor de k utilizado foi 2, uma vez que número maiores não apresentaram melhores resultados devido à complexidade do problema.

Como resultado do processo de inferência gramatical, cada classe do problema em questão terá um autômato, que será a representação formal de suas regras (características), passível de ser validada ao receber uma sentença como entrada de dados. Assim, para classificar uma imagem, ela primeiramente é convertida em uma

¹⁸ Mais informações podem ser vistas em: <https://www.mathworks.com/help/stats/hierarchical-clustering.html>

sentença e, em seguida, submetida à validação em cada um dos autômatos. Cada autômato representará uma classe do banco de imagens. O autômato que gerar o menor número de erros será a espécie da sentença executada na validação. Os erros ocorrem quando há um caractere na sentença que não está previsto, não pertence a linguagem ou não existe a transição na linguagem do autômato.

A Figura 6 ilustra o processo de classificação de espécies de peixes desta tese. No passo 1, inicialmente é realizada a extração de pontos de interesse, a limitação do total de pontos por imagem e a criação do vetor de descrição de cada ponto. No passo 2, o alfabeto é construído pela abordagem supervisionada (ou não supervisionada) com base no conjunto de descritores dos pontos de interesse, rotulando cada um deles. No passo 3, é aplicado a técnica de agrupamento hierárquico, tendo como resultado as sentenças agrupadas do passo 4. No passo 5, a aprendizagem da gramática é realizada, tendo como produto final os autômatos para a validação de um algoritmo de inferência gramatical.

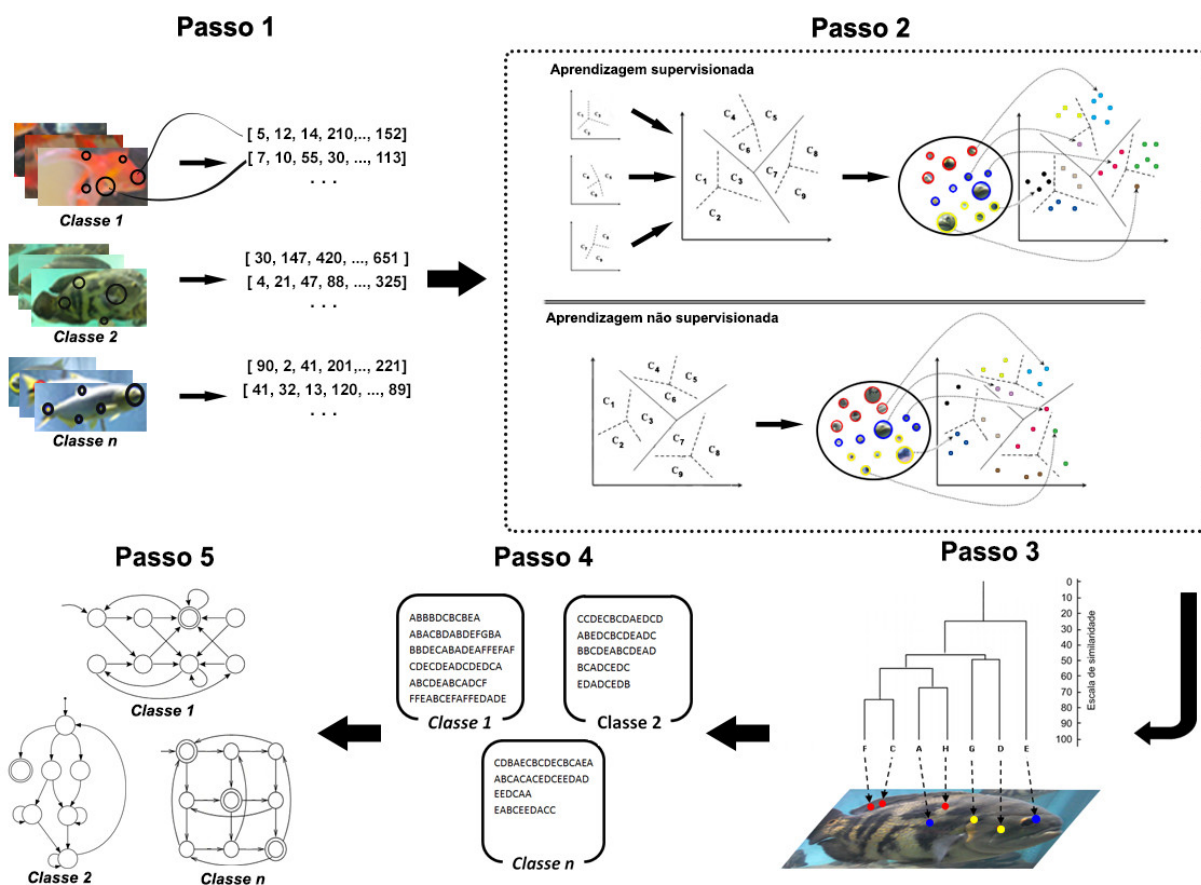


Figura 6. Passos da proposta de um algoritmo para a classificação de espécies de peixes utilizando reconhecimento sintático de padrões.

Implementação da abordagem proposta

O sistema computacional foi desenvolvido nas linguagens de programação Matlab¹⁹ e Java²⁰. No desenvolvimento com o Matlab foi usada a biblioteca *vl_feat*²¹ que implementa algoritmos populares de visão computacional especializados em compreensão de imagem, extração de características e casamento de dados. Essa biblioteca tem por objetivo extrair informações, de modo que o computador “aprenda” (no sentido de obter conhecimento), tentando de forma indutiva gerar hipóteses para obter soluções a partir de padrões encontrados.

Pela linguagem Matlab foi desenvolvida a geração do alfabeto, a rotulação das palavras visuais e geração de sentenças. Pela linguagem Java foi desenvolvida a detecção de pontos de interesse, a limitação e descrição de pontos de interesse e a classificação baseada na inferência gramatical usando o algoritmo *k-testable*. Para o uso se tornar viável em ambientes com execução em tempo real, torna-se necessário disponibilizar o processo de aprendizagem e classificação em um servidor dedicado para essa finalidade. A aplicação, por sua vez, enviará a imagem para processamento no servidor dedicado, e este retornará a resposta da classificação para a aplicação.

Foram seguidas regras definidas pelo grupo de pesquisa e desenvolvimento INOVISAO disponíveis no site do grupo²². O INOVISAO tem como base a metodologia de desenvolvimento de software SCRUM (Sims *et al.*, 2011). O código-fonte desenvolvido é mantido e monitorado pelo sistema de controle de versão distribuído Git²³ (Chacon, 2010). O padrão de documentação de código é baseado no *Javadoc*²⁴.

A Figura 7 ilustra a interface criada para realizar a classificação de imagens deste trabalho. Na parte superior é preciso informar as configurações gerais, que servirá para os 3 passos subsequentes da abordagem: extração de informação, geração da sentença e classificação. Esses passos são descritos abaixo:

- Passo 1: É realizado a detecção dos pontos de interesse, a limitação e a descrição;

¹⁹ Disponível em: <http://www.mathworks.com/products/matlab/>

²⁰ Disponível em: <http://www.java.com/>

²¹ Disponível em: <http://www.vlfeat.org>

²² Instruções para desenvolvedores estão disponíveis no link: <http://trac.gpec.ucdb.br>

²³ Disponível em: <http://git.inovisao.ucdb.br/>

²⁴ *Javadoc* é um padrão de documentação da Oracle Corporation. Tem disponível uma ferramenta para a geração de documentação em formato HTML a partir dos comentários presentes nos códigos-fonte.

- Passo 2: É realizado a geração do alfabeto, a rotulação das palavras visuais e geração de sentenças; e,
- Passo 3: É realizado o processo classificação usando o algoritmo de inferência gramatical.

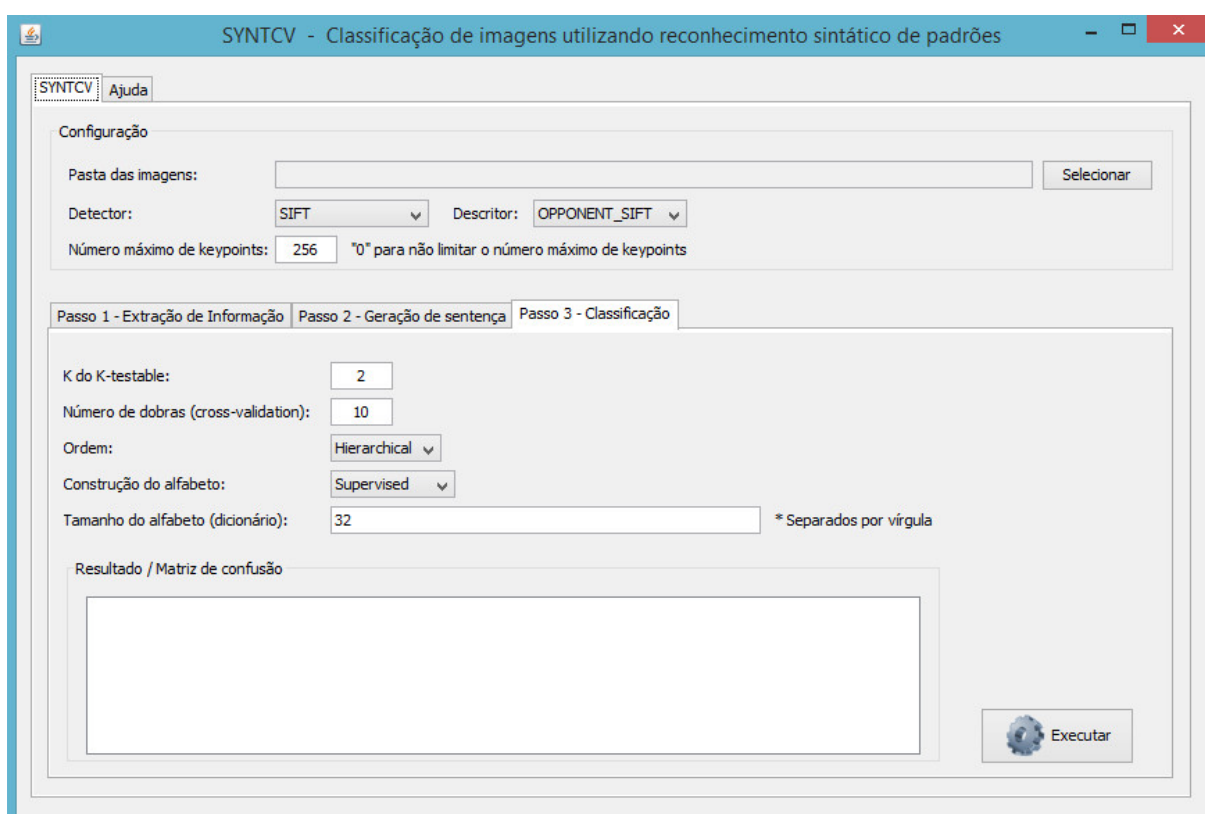


Figura 7. Interface proposta da aplicação para a classificação de espécies de peixes utilizando reconhecimento sintático de padrões.

Validação da abordagem proposta

A técnica de amostragem adotada foi a validação cruzada, com 10 dobras. Essa técnica avalia o potencial de generalização de uma abordagem a partir de um conjunto de dados. Ela divide o conjunto de dados em dois subconjuntos, um para determinar os parâmetros (treinamento) e outro para validar (teste). A partir disso, busca-se estimar o quão preciso é o modelo na prática, ou seja, prever qual seria o seu desempenho para um novo conjunto de dados.

Para as abordagens testadas, foi calculado o desempenho médio referente a métrica da Medida-F, também conhecida como *F-Score*. Essa métrica é usada para

discriminar os classificadores e extratores, obtida a partir da média harmônica da precisão e revocação. Ambas são calculadas utilizando os valores verdadeiros positivos (VP), falsos positivos (FP) e falsos negativos (FN), que podem ser obtidos por meio de uma matriz de confusão. Essa métrica é obtida a partir de problemas binários, entretanto é possível generalizar para problemas com múltiplas classes. Para isso, basta considerar que exemplos da classe são positivos e exemplos de qualquer outra classe são negativos, repetindo para as n classes do problema. A Medida-F é obtida a partir da equação:

$$\text{Medida } F = 2 \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}}$$

Outra análise utilizada neste trabalho é por meio da matriz de confusão ou de contingência. Ela mostra o número de classificações corretas e preditas para cada classe, a partir de um conjunto de exemplos. Nesse caso, o número de acertos para cada classe se localiza na diagonal principal da matriz. Os demais elementos representam erros de classificação. A matriz de confusão é uma estratégia relevante de ser adotada para avaliar os resultados, uma vez que facilita o entendimento da classificação de amostras positivas e negativas.

Para identificar se as abordagens diferem estatisticamente em relação ao desempenho, foi utilizado o teste não-paramétrico proposto por Friedman (1937). Esse modelo estatístico está disponível no software estatístico R²⁵. São reportados os valores-p e o nível de significância necessário para descartar a hipótese nula. Caso seja constatada diferença utilizando um nível de significância de 95% (valor $p < 0.05$), será realizado um pós-teste (*post-hoc*) e a geração de diagramas de caixa (*box-plots*) e analisados valores-p dois a dois. O pós-teste, também disponível no R, tem como base o teste de Wilcoxon com correção para FWER (*Family-wise Error Rate*), descrito por Hollander e Wolfe (1999).

Organização dos próximos capítulos

Na sequência, encontra-se uma divisão em 3 capítulos. O Capítulo 1 mostra os resultados dos primeiros experimentos realizados utilizando a abordagem base proposta. No Capítulo 2, são apresentados resultados intermediários obtidos a partir

²⁵ O software R está disponível em: <http://www.r-project.org/>

do uso da limitação de pontos de interesse, a partir de uma adaptação da abordagem base. No Capítulo 3, são apresentados os resultados da abordagem final desta Tese, usando a limitação de pontos de interesse, com o uso de cores para a descrição de cada ponto de interesse. Por isso, a descrição e validação de resultados varia de um capítulo para outro, dependendo do avanço da pesquisa, saindo da abordagem base (Capítulo 1) à abordagem completa (Capítulo 3). Por fim, são apresentadas as considerações finais deste trabalho.

CAPÍTULO I

O artigo a seguir foi publicado e apresentado no X Workshop de Tecnologia Adaptativa – WTA²⁶, São Paulo - SP. Neste artigo é descrita a abordagem base e experimentos iniciais desta tese.

Citação:

Marcelo Rafael Borth, Lucas Ribas, Hemerson Pistori, Wesley Gonçalves, Amaury Antônio de Castro Junior. Classificação de espécies de peixe usando inferência gramatical no reconhecimento de padrões em problemas de Visão Computacional. Em: Memórias do X Workshop de Tecnologia Adaptativa - WTA 2016. EPUSP, São Paulo. ISBN: 978-85-86686-86-3, pp. 60-69. 28 e 29 de Janeiro, 2016.

²⁶ Acesso ao website do evento: <http://lta.poli.usp.br/lta/wta/wta-2016>

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

M. R. Borth²⁷, L. C. Ribas²⁸, H. Pistori²⁹, W. N. Gonçalves³⁰, A. A. C. Junior³¹

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. Nesse contexto, a criação de softwares para monitoramento de peixes ou de seu ambiente possibilita adotar alternativas ou ações específicas para atuarem desde o controle preventivo de espécies de peixes até a criação de novas soluções. 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 que as permitem ser únicas, 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 ciência, 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 das 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 à apenas atributos relacionados a borda do objeto, diferentemente deste trabalho que pode utilizar informações de toda a parte da imagem.

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 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

²⁷ 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

²⁸ Mestrando em Ciência da Computação e Matemática Computacional pela Universidade de São Paulo. E-mail: lucascorreiaribas@gmail.com

²⁹ Doutor em Engenharia Elétrica pela Universidade de São Paulo. Professor da Universidade Católica Dom Bosco (UCDB). E-mail: pistori@ucdb.br

³⁰ 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

³¹ 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

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 [11]. 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*.

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* [12] 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 *aquario10e*³³ constituído de 596 imagens de peixes e o *aquarioSeg10e*³⁴ 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 [13] 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,
- $\varphi_i \in R^{128}$, 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 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

³³Disponível em: <https://www.dropbox.com/sh/crq14nug14juc4/AAAY-1ul8YBQ5AtXA7M1fz1wa>

³⁴Disponível em: <https://www.dropbox.com/sh/50u4z5o197bstqk/AACDf64uxVOFKTyti7smoIG-a>

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_t . 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_t-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].

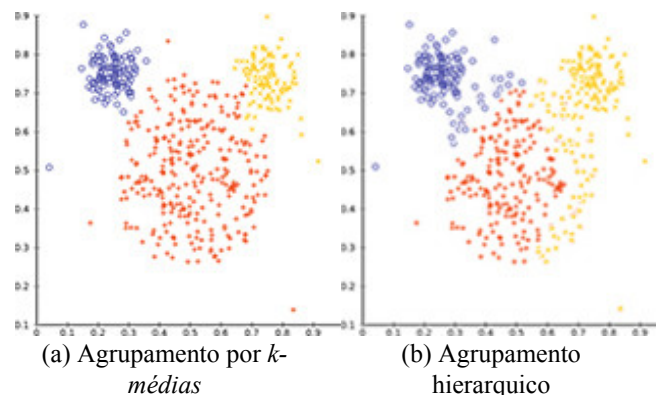


Figura 1. (a) Exemplo de agrupamento pelo algoritmo *k*-médias e (b) agrupamento hierárquico.

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.

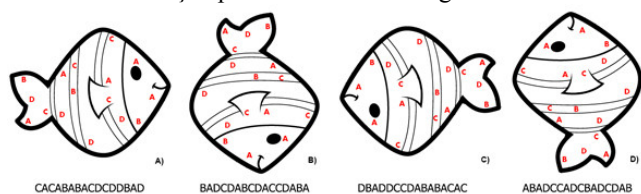


Figura 2. Quatro imagens de peixe com variação de 90° na rotação de uma para outra. Abaixo a sentença gerada de cada imagem a partir da ordenação dos pontos de interesse pela ordem de leitura ocidental (iniciando da parte superior e concatenando da esquerda para a direita).

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 centroe possui a mesma dimensão dos descritores dos pontos de interesse, em que $C \in R^{128}$. 2.

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 centroe 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:

Encontrar a similaridade entre cada par de pontos de interesse no conjunto de dados: nessa etapa, é calculada a distância entre cada ponto de interesse usando a distância Euclidiana; e,

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.

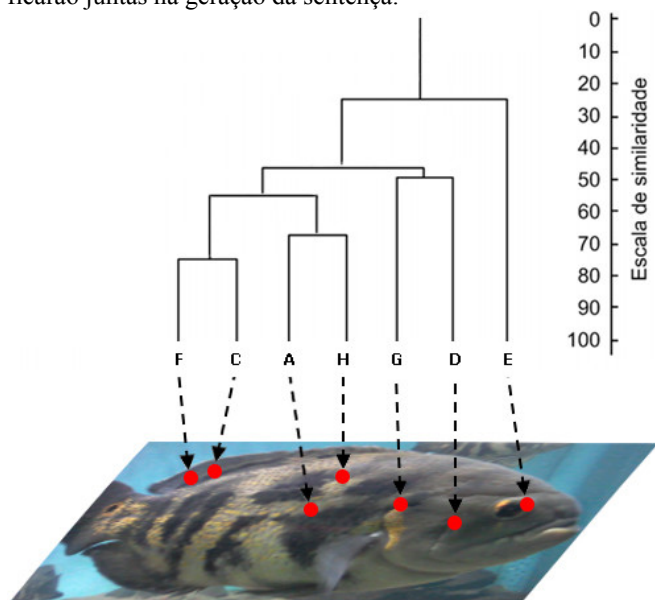


Figura 4. Ordenação dos pontos de interesse na imagem do peixe a partir do agrupamento hierárquico representado por um dendograma. A sentença gerada para essa imagem seria “F C A H G D E”.

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. 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.

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.

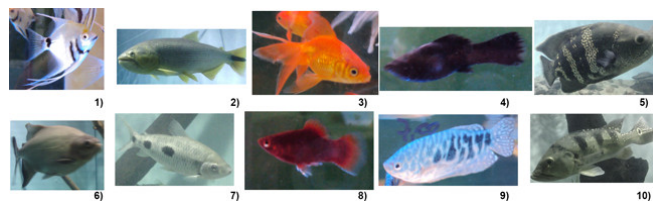


Figura 5. Imagens das espécies de peixes do banco de imagens *aquario10e*.

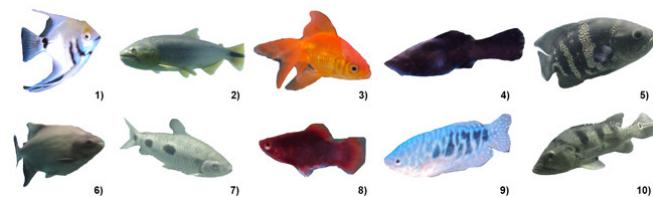


Figura 6. Imagens das espécies de peixes do banco de imagens *aquarioSeg10e*.

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.

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.

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.

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.

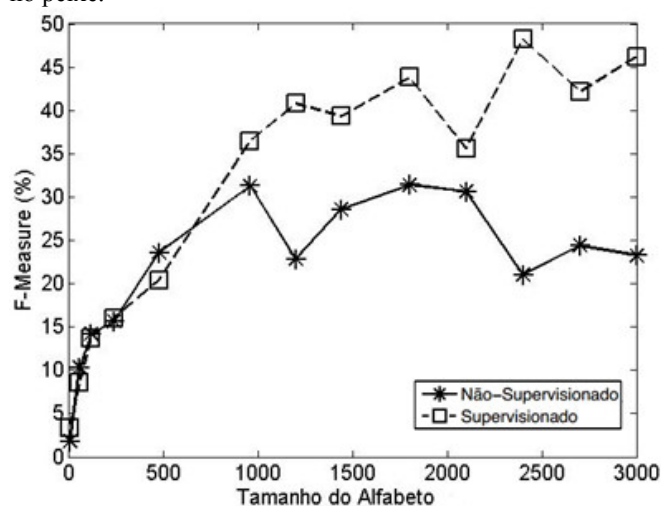


Figura 7. Gráfico comparativo da Medida-F para a abordagem supervisionada e não supervisionada observando o tamanho de cada alfabeto com o banco de imagens aquarioSeg10e e ordenação hierárquica.

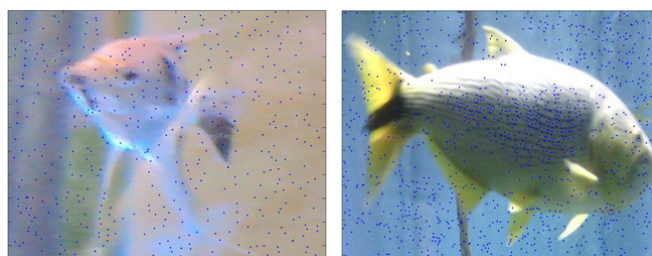


Figura 8. Exemplo de pontos de interesse detectados em imagens de duas espécies de peixe em que a maioria são ruídos

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 Molinésia 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

Kinguio 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.

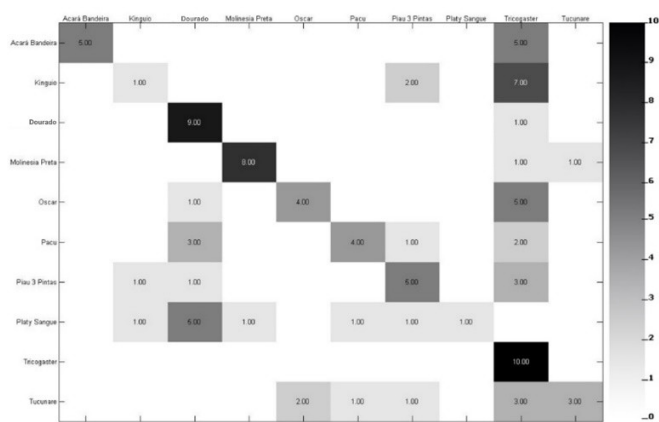


Figura 9. Matriz de confusão mostrando o número de imagens classificadas corretamente para cada espécie de peixe para o resultado de maior Medida-F (banco de imagens aquarioSeg10e, abordagem supervisionada, geração da sentença pela ordenação hierárquica e tamanho de alfabeto 2400).

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 [35] 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] 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.

CAPÍTULO II

O artigo a seguir foi publicado e apresentado no XII Workshop de Visão Computacional³⁵, na cidade Campo Grande - MS. Neste artigo é descrita a abordagem que usa a limitação de pontos de interesse desta tese.

Este artigo trata a limitação encontrada no Capítulo I, onde até o momento há a predominância de classes, em que elas possuem maior poder descritivo por ter mais pontos de interesse detectados que outras com menos. Como consequência, classificam bem mais imagens se comparadas a outras classes com poucos pontos de interesses. Até o momento, o desbalanceamento na classificação é evidente. Por conta disso, este artigo cria a limitação de pontos de interesse com o objetivo de equalizar a quantidade de informações interpretadas por imagem, para não ter prevalência de classes no processo de aprendizagem. Além disso, a proposta visa reduzir o tempo de execução do processo de aprendizagem (teste) e de classificação.

Citação:

Marcelo Rafael Borth, Wesley Nunes Gonçalves, Hemerson Pistori. Keypoints Reduction Applied to Image Classification Problems. Proceedings of XII Workshop de Visão Computacional (WVC 2016), pp. 382-387, November 09-11th, 2016.

³⁵ Acesso ao website do evento: <http://wvc2016.weebly.com/>

Keypoints Reduction Applied to Image Classification Problems

Marcelo Rafael Borth
IFPR – Instituto Federal do Paraná
Umuarama, Brazil
marcelo.borth@ifpr.edu.br

Hemerson Pistori
UCDB – Universidade Católica Dom Bosco
Campo Grande – Brazil
pistori@ucdb.br

Wesley Nunes Gonçalves
UFMS – Universidade Federal do Mato Grosso do Sul
Ponta Porã, Brazil
wesley.goncalves@ufms.br

Abstract — For the feature extraction in images there are algorithms such as SIFT which detect and describe keypoints. In the image classification process is common to have an image dataset for each set of images (class) and when using an algorithm to detect and describe local features in images, the number of keypoints detected by class can be a quite disproportionate, resulting a biased classification for classes with higher amount of keypoints. This paper presents a novel approach to reduce the amount of keypoints detected in images, so that computer vision techniques can be applied to image classification problems. An experiment was conducted using a new dataset, created by the authors of this work, in order to have the first results of the technique. As a result, the F-Measure was improved in 97% under a new dataset. In addition to that, the computer performance has increased in 300 times.

Keywords — *Keypoints Reduction; Imagem Classification; Pattern Recognition; Grammatical Inference.*

I. INTRODUÇÃO

Um dos desafios constantes quando se refere a classificação e análise de imagem é melhorar o processo de obtenção de características representativas. Para isso, os pontos de interesse são frequentemente usados na análise de características de imagens. Um algoritmo consolidado para a detecção e descrição de pontos de interesse é o *Scale Invariant Feature Transformation* (SIFT) [1].

Ao aplicar um algoritmo como o SIFT em um banco de imagens é comum que sejam detectadas quantidades diferentes de pontos para cada imagem. Quanto mais pontos identificados na imagem, maior o poder descritivo de características dessa imagem, o que pode ajudar posteriormente no processo de classificação. Entretanto, um fator importante é a forma do emprego desses algoritmos nas abordagens de classificação, pois considerando uma proposta que usa um banco de imagem dividido em classes de imagens, onde cada classe é um grupo semelhante de imagens, é possível que uma classe possua mais pontos de interesse detectados que uma outra. Isso é comum, porque os algoritmos de detecção não limitam a quantidade de pontos detectados por imagem, uma vez que eles analisam a

importância do ponto detectado baseado em um limiar (*threshold*). Assim, os algoritmos detectam os pontos de interesse, sem se preocupar com a quantidade detectada. Diante do exposto, existe o problema do desbalanceamento da quantidade de pontos de interesse detectados por classe. Dessa forma, é possível que algumas classes possuam muitos pontos de interesse detectados enquanto que outras possuam poucos, podendo acarretar em uma classificação tendenciosa para as classes de maior poder descritivo.

Este trabalho apresenta uma nova forma para limitar e reduzir a quantidade de pontos de interesse extraídos por algoritmos de detecção e descrição como o SIFT. A redução de pontos de interesse tem como objetivo eliminar informações redundantes e que tendem a não ser tão informativas para a descrição do conjunto de imagens, bem como equalizar a quantidade de pontos para cada classe do banco de imagem e, como consequência, reduzir o tempo de processamento do processo de classificação. A partir da redução de informação é possível melhorar a qualidade do modelo construído durante um processo de classificação de imagens, tornando-se mais imparcial o processamento direcionado aos problemas de classificação de imagens. Logo, a redução de pontos de interesse prevê diminuir o número total de pontos selecionados para compor o novo subconjunto de informações, aumentar o percentual de acertos do classificador e a minimizar a complexidade da geração de conhecimento como, modelos conceituais, autômatos, etc. Este trabalho faz a limitação e redução da quantidade de pontos de interesse para as imagens e não a redução de dimensionalidade do vetor de descrição criado pelo algoritmo de detecção.

Para o processo de classificação de imagens, este trabalho ordena os pontos de interesse detectados em cada imagem, produzindo uma sentença. A sentença, também conhecida por palavra ou cadeia de caracteres, é uma sequência finita de palavras visuais ordenadas, por exemplo, “ABACDECBCADA”. Para gerar a sentença, a ordenação é baseada no agrupamento hierárquico dos pontos de interesse

que são rotulados e identificados como palavras visuais, unindo pontos de interesse similares entre si a partir de uma posição espacial (x, y) que possuem maior similaridade, até formar a estrutura de uma árvore binária completa. A palavra visual é apenas uma representação simbólica, em forma de letra, do descritor do ponto de interesse extraído pelo SIFT. A ideia principal deste trabalho é extrair pontos de interesse de uma imagem, criar palavras visuais para rotular os pontos de interesse e ordená-los produzindo uma sentença de tamanho n , tal que n é a quantidade de pontos de interesse detectados na imagem. O tamanho do alfabeto é definido usando o algoritmo k -médias (ou k -means), como no histograma de palavras visuais [2] [3]. O alfabeto, também conhecido como dicionário, é um conjunto finito de símbolos, números ou caracteres. Logo, a sentença é construída com base no tamanho ou distinção de um alfabeto, uma vez que o seu tamanho determina o quão uma imagem é bem descrita computacionalmente. Embora seja utilizada neste trabalho uma abordagem de reconhecimento sintático de padrões, a redução de pontos de interesse proposta aqui pode ser aplicada à qualquer abordagem de classificação de imagens.

O experimento realizado foi utilizando imagens de peixes, a fim de analisar a estratégia em um possível problema real. Para o experimento, o algoritmo de detecção de pontos de interesse SIFT e o algoritmo de inferência gramatical k -testable [4] foram utilizados. Entretanto, a técnica pode ser aplicada usando qualquer algoritmo de detecção de pontos de interesse que seja possível quantificar a sua importância e de inferência gramatical disponíveis.

Outra contribuição deste trabalho é a publicação de um novo banco de imagem chamado AQUARIO28E40I³⁶, contudo a abordagem funciona para qualquer conjunto de imagens. O novo banco possui 40 imagens de peixes para cada uma das 28 espécies, totalizando 1120 imagens. Ele diferencia-se por conter apenas imagens capturadas por câmeras de celulares, visto que grande maioria dos bancos de imagens atuais fornecem apenas imagens de câmeras fotográficas.

II. O PROCESSO DE CLASSIFICAÇÃO DE IMAGENS

A classificação de imagens deste trabalho é baseada na proposta desenvolvida por [5] e pode ser dividida em 5 etapas:

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. Esse algoritmo extrai da imagem uma coleção de vetores de características locais, contendo 128 valores que descrevem a região ao redor do ponto 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. Essa técnica busca ser invariante a rotação, escala da imagem e mudança de iluminação e é extremamente relevante para a tarefa de reconhecimento de padrões.

B. Construção do Alfabeto

A construção do alfabeto é feita pela abordagem supervisionada, conforme trabalho de Ribas *et al.* (2014). Ela se baseia no conjunto de descritores D dos pontos de interesse detectados, 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 .

Na abordagem supervisionada, 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 :

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

$$C_p = k\text{-médias}(D_p)$$

Cada classe terá um conjunto de centroides, definidos a priori, chamado de alfabeto. Por fim, os conjuntos de centroides são concatenados para obter o alfabeto completo.

C. Rotulação de Pontos de Interesse

Dados os pontos de interesse p_i^j para a imagem j , cada ponto de interesse 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 , $|\varphi_i^j, C_l|$ é a distância Euclidiana e k é o tamanho do alfabeto. 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.

D. Geração da Sentença

Nessa etapa, os pontos de interesse da imagem são acessados e gerado uma sentença com as palavras visuais correspondentes de cada ponto de interesse de cada imagem. Para gerar a sentença a partir das palavras visuais da imagem, a ordenação baseada em agrupamento hierárquico foi usada. 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

³⁶ O banco de imagem pode ser obtido pelo link: <http://pistori.weebly.com/datasets.html>

são agrupadas, em uma árvore hierárquica binária, para formar a sentença.

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. A variável *k* do algoritmo *k-testable* significa uma quantidade *k* de caracteres anteriores ao caractere atual analisado(s) na inferência gramatical.

Nesse momento, as sentenças que representam as imagens de teste são verificadas nos autômatos gerados. Para verificar 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.

III. REDUÇÃO DE PONTOS DE INTERESSE

No processo de classificação de imagens é possível ter uma grande quantidade de informações para representar imagens, de modo que torna-se necessário separar atributos relevantes dos não relevantes, a fim de otimizar o processamento da classificação.

Na abordagem proposta neste trabalho, a redução de pontos de interesse é realizada após a extração e descrição de pontos de interesse realizada pelo algoritmo SIFT. Teoricamente, é possível imaginar que quanto maior a quantidade de pontos de interesse detectados na imagem, maior é a quantidade de informações disponíveis para processamento do algoritmo de classificação. Logo, o processo de redução de pontos de interesse permite filtrar de acordo com um critério de importância, reduzindo a quantidade de possíveis pontos ruidosos ou não tão relevantes.

A discrepância da quantidade de informação para cada classe pode resultar em uma grande diferença na classificação de uma para outra. O algoritmo de aprendizagem pode ser prejudicado tanto em seu desempenho, por causa da quantidade de dados, quanto no percentual de classificação correta, por conta de informações redundantes que podem confundir o algoritmo, prejudicando-o na construção de um modelo adequado para o conhecimento exposto [6]. Ao visualizar a matriz de confusão da classificação apresentada por [5] é possível notar a predominância de classe na classificação. Por exemplo, no experimento, uma única classe (espécie de peixe) classificou 37% de todas as imagens contidas no banco de imagens. Em geral, as classes com maior poder descritivo obtiveram mais imagens classificadas pelo algoritmo, pois descreveram mais características quando comparadas às classes com menos pontos de interesse detectados. Conforme afirma [7], 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 desbalanceamento pode prejudicar o classificador quando novos exemplos precisam ser classificados, fornecendo uma

classificação tendenciosa. Além disso, essa situação pode piorar conforme a prevalência de classe torna-se mais evidente, podendo aumentar a precisão de classes majoritárias e reduzir de classes minoritárias.

Por essa razão, propomos o uso de uma estratégia para limitar a quantidade de pontos de interesse na fase de extração de características, deixando a quantidade de pontos semelhantes (iguais ou bem próximas) para cada classe. O objetivo de equalizar o total de pontos de interesse entre as classes pode ser relevante a fim de determinar proporcionalmente o poder descritivo entre cada uma delas, diminuindo a prevalência de classe.

Na Figura 1 é apresentado o total de pontos de interesse detectados das imagens utilizadas neste trabalho, separados por espécie. Pode ser observado que existe uma tendência de algumas classes terem uma quantidade significativa de informação (pontos de interesse detectados), enquanto que outras tem poucas informações para representar a classe de imagens da espécie de peixe. Se compararmos duas espécies, a de maior quantidade de pontos identificados (Kinguio) com a de menor quantidade (Platy Laranja), a primeira tem 87 vezes mais pontos de interesse detectados que a segunda. Kingiuo obteve 10.5% do total de pontos detectados de todo o banco de imagem, enquanto que Platy Laranja obteve apenas 0.1% do total. Logo, o desbalanceamento e a prevalência de informação são evidentes.

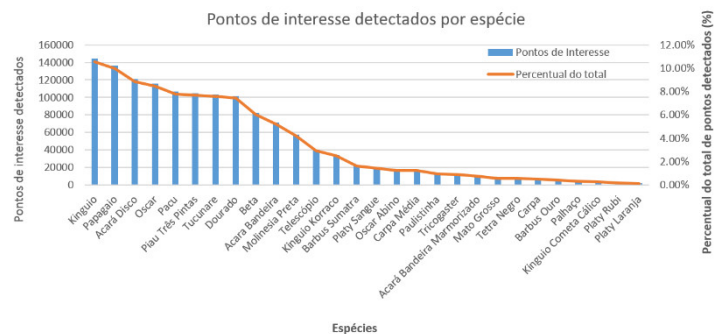


Figura 1: Total de pontos de interesse detectados por espécie considerando o banco de imagem utilizado neste trabalho.

Com a distribuição da quantidade de pontos de interesse por classe e, também, pela distribuição dos pontos de interesse baseando-se na sua importância, podemos notar um efeito conhecido como “cauda longa” (*long tail*) [8], onde poucos pontos de interesse tem um alto valor de importância e muitos têm um baixo valor de importância para a classificação. Os de baixa importância não tornam-se relevantes manter para o aprendizado da gramática e classificação. O efeito cauda longa, por sua vez, é previsto pela Lei de Zipf’s [9]. Essa lei está associada ao Princípio de Pareto [9], também conhecido pela regra dos 80-20. Essa regra determina que 80% das causas são responsáveis por 20% dos efeitos, enquanto que o restante dos 20% das causas são responsáveis por 80% dos efeitos [10] [11]. À vista disso, podemos associar a regra dos 80-20 com os pontos de interesse e deduzir que muitos deles não são

representativos, devido a estarem na cauda longa, enquanto que apenas poucos são representativos, o que justifica usar apenas a parte mais representativa de cada imagem para a classificação. Logo, na tentativa de equalizar a quantidade de informação entre as classes e deixar a classificação mais equitativa, realizaremos a redução de pontos de interesse.

Com base nos valores de *val* e *score* da implementação da linguagem de programação C do algoritmo SIFT é possível rejeitar pontos identificados e que não são tão “fortes” ou representativos comparados a outros detectados. Na implementação, *val* é a “força” da derivada e *score* significa se o ponto de interesse possui resposta de borda forte em uma única direção. Em nossa implementação, obtivemos um valor de importância para o ponto de interesse, sendo:

$$\text{Importância} = \text{val} * 0,5 + \text{score} * 0,5$$

Após obter o valor de importância de cada ponto de interesse, elimina-se os menores baseado no número máximo determinado para cada imagem.

A Figura 2 ilustra a limitação progressiva de pontos de interesse em seis imagens e em uma delas é ilustrado todos os pontos de interesse detectados. Percebe-se que, quanto mais pontos de interesse detectados, maior a quantidade de pontos no fundo da imagem. Os pontos do fundo da imagem podem ser aqueles de menor importância na imagem. A limitação por imagem foi de 32, 64, 128, 256, 512 e 1024 pontos, respectivamente, 2^5 , 2^6 , 2^7 , 2^8 , 2^9 e 2^{10} . Também, podemos observar a imagem com todos os pontos de interesse. Nela, existem centenas de pontos identificados no fundo da imagem que não são relevantes para a descrição do peixe. Embora poderiam ser relevantes para identificar o contexto de onde o peixe está inserido, não são relevantes para a descrição de características do peixe, considerado neste trabalho.



Figura 2. Seis imagens de peixes exibindo os pontos de interesse detectados com suas respectivas limitações e uma imagem exibindo todos os pontos de interesse detectados.

IV. EXPERIMENTO

O experimento realizado utilizou imagens de peixes de 28 espécies, contidas no banco de imagens AQUARIO28E40I. Ao total são 1120 imagens coloridas divididas em 28 classes, onde cada classe possui 40 imagens cada. As imagens variam umas das outras em tamanho e resolução, pois elas foram capturadas de diferentes celulares e posteriormente recortadas manualmente.

Em [5], a construção do alfabeto foi realizada de duas formas, supervisionada e não-supervisionada. Entretanto, para a redução de pontos de interesse deste trabalho, utilizamos apenas a abordagem supervisionada, por ela implicar em melhores resultados. A abordagem não supervisionada será utilizada apenas para comparação de resultados. A ordenação usada foi baseada no agrupamento hierárquico. O valor do parâmetro k_t do *k-testable* utilizado neste experimento foi $k_t = 2$, uma vez que usando um valor maior, a geração da gramática não correspondeu a uma melhora ao 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*) [12], também conhecida como *F-Score*, foi adotada como parâmetro de comparação entre as abordagens testadas. Essa métrica é a média harmônica da precisão e revocação.

A quantidade máxima de pontos de interesse por imagem é múltipla de 2, sendo elas: 2^5 , 2^6 , 2^7 , 2^8 , 2^9 e 2^{10} . Essa é a quantidade máxima analisada por imagem, pois depende da quantidade de pontos de interesse que o algoritmo de extração identifica, uma vez que é possível identificar uma quantidade menor da máxima estipulada.

No experimento cada técnica será representada por uma sigla: **LPI_128** – Limitação de Pontos de Interesse com os melhores 128 pontos de interesse extraídos para cada imagem (128, pois foi a quantidade do alfabeto de melhor resultado); **BASE_SUP** – construção do vocabulário pela forma supervisionada; e **BASE_NSUP** – construção do vocabulário pela forma não supervisionada.

V. RESULTADOS E DISCUSSÃO

Na Tabela 1 são apresentados os resultados do experimento realizado com o banco de imagem AQUARIO28E40I. A Medida-F mais alta obtida está marcada com asteriscos. O melhor resultado obtido usando a limitação de pontos de interesse (LPI_128) foi de 33,4% usando ao máximo 128 pontos de interesse por imagem. O maior valor alcançado pela abordagem que utiliza todos os pontos de interesse foi de 16,9%, construindo o alfabeto pela forma supervisionada (BASE_SUP) e 11% para a não supervisionada (BASE_NSUP). Portanto, usando a abordagem proposta por este trabalho podemos dizer que houve um aumento de 97% da Medida-F utilizando o novo banco de imagem, sendo que o percentual foi obtido pela razão do maior valor alcançado da abordagem LPI_128 pelo maior valor da abordagem BASE_SUP.

Com a redução do total de pontos de interesse, menos pontos são utilizados no processo de construção do alfabeto, geração de sentença e geração do autômato para o reconhecimento de padrões que, consequentemente, reflete na redução do tempo de processamento do algoritmo como um todo. A configuração do experimento mudou apenas na redução de pontos de interesse atribuídos por imagem para compor o conjunto final. Logo, se compararmos o desempenho computacional da classificação que resultou nas melhores Medidas-F para as duas abordagens supervisionadas, LPI_128 e BASE_SUP, a abordagem proposta

neste trabalho (LPI_128) executou 300 vezes mais rápido que a abordagem de base (BASE_SUP).

Tamanho do alfabeto (k)	Supervisionada						Não Superv.	
	Limitação de pontos de interesse (LPI) - Número máximo para cada imagem						Todos os pontos de interesse	
	32	64	128	256	512	1024	BASE	BASE
32	14.4	10.2	11.9	10.0	7.6	5.0	2.2	0.6
64	19.1	23.0	22.4	18.9	13.6	11.1	1.4	2.1
128	17.5	24.6	*33.4*	25.7	23.1	17.7	2.8	2.0
256	16.0	20.7	20.1	26.6	30.1	24.0	2.8	1.9
512	23.5	27.9	26.0	28.8	28.2	27.3	3.4	5.5
1024	25.4	28.1	23.1	27.0	26.1	24.9	6.0	6.9
2048	21.9	25.7	24.0	21.1	22.4	21.8	9.2	9.1
3072	22.6	20.1	25.5	21.4	20.5	21.6	11.8	8.3
4096	21.6	12.7	28.2	22.2	22.2	19.6	11.5	9.8
5120	26.1	25.3	25.6	19.3	18.7	22.2	11.9	8.9
6144	24.4	24.4	22.6	23.9	20.0	20.4	16.2	11.0
7168	23.4	28.0	28.4	22.2	20.3	20.5	14.6	10.4
8192	30.3	25.6	18.9	20.9	21.6	21.0	16.9	9.3

Tabela 1: Medidas-F para a abordagem proposta executada no banco de imagens aquario28e40i. Os valores mais altos para cada tamanho do alfabeto para ambos os bancos de imagens estão marcados em negrito. A Medida-F mais alta da tabela está exibida entre dois asteriscos.

Ao distribuir os valores obtidos para cada alfabeto, obteve-se outra informação relevante, em que pode ser percebido que os resultados de todos os tamanhos de alfabetos para a abordagem de limitação de pontos de interesse (LPI_128) foram superiores à abordagem base, que não limita a quantidade de pontos de interesse. Além disso, se comparar cada tamanho de alfabeto da Tabela 1, a limitação de pontos de interesse teve resultado superior para todas as configurações (BASE_SUP e BASE_NSUP). Acreditamos que esse resultado advém pela relevância dos pontos de interesse detectados e escolhidos para representar a imagem, uma vez que ao limitar a quantidade de informação para análise, teremos apenas os melhores n pontos de interesse de cada imagem. Por outro lado, a quantidade excessiva de pontos detectados e usados nas abordagens que não usam a limitação, possui mais ruídos, reduzindo o desempenho do classificador, pois em várias imagens a maioria dos pontos de interesse são detectados no fundo da imagem e não no peixe.

A Figura 3 apresenta a matriz de confusão para o melhor resultado obtido no experimento com a Medida-F de 33,4%. Os peixes ilustrados por números nessa figura correspondem aos mesmos ilustrados na Figura 5. Nela é possível visualizar as classificações corretas e incorretas para cada espécie, considerando que na diagonal principal está representada a quantidade de imagens classificadas corretamente. Logo, o valor fora dessa diagonal representam erros da classificação. Percebe-se que a taxa de classificação correta se sobressai em alguns elementos da matriz de confusão, como na diagonal principal e para algumas espécies em que a classificação pode ser vista nas colunas verticais da matriz de confusão. 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, o que torna a tarefa com um alto grau de complexidade.

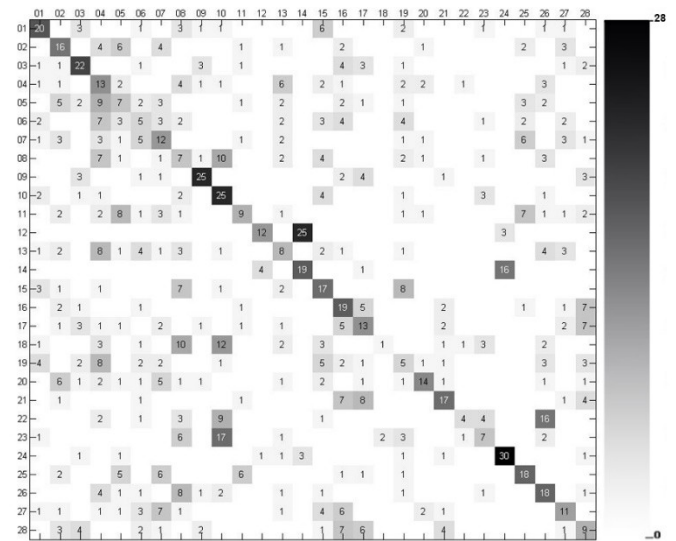


Figura 3: Matriz de confusão mostrando o número de imagens classificadas corretamente para cada espécie de peixe para o resultado de maior Medida-F, com resultado de 33,4%.

Nesse tipo de experimento, questões como a mudança de cor, iluminação desigual, fundo diferente, sedimentos na água e plantas subaquáticas podem gerar ruídos na classificação. Além disso, 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. Todos esses desafios dificultam o processo de classificação de imagens.

Para identificar se o desempenho dos classificadores difere estatisticamente, foi utilizado o teste proposto por Tukey, da diferença honestamente significativa (*Honestly Significant Difference* – HSD), disponível no software estatístico R Studio, “versão 0.99.896”. A estratégia desse método consiste em definir a menor diferença significativa. Para isso, foi empregado o teste chamado ANOVA com o intervalo de confiança de 95% ($p < 0,05$). Sumarizando, o valor de p foi de 3.74×10^{-6} , com isso podemos rejeitar a hipótese nula e afirmar que as abordagens são estatisticamente diferentes. Entretanto, ao comparar a abordagem proposta neste trabalho (LPI_128) com as outras duas, ela se mostra estatisticamente superior que ambas. O teste de Tukey mostrou que existe diferença entre as abordagens, pois o valor- p em comparação da abordagem proposta (LPI_128) com a BASE_SUP foi de $0,3 \times 10^{-3}$ e para a BASE_NSUP foi de $0,61 \times 10^{-5}$, conforme pode ser visto na Tabela 3.

A Figura 4 ilustra o diagrama de caixas para cada abordagem com base na Medida-F. É possível visualizar que a abordagem LPI_128 possui uma maior volatilidade. Além disso, a abordagem LPI_128 possui a mediana mais alta. Também, todo seu corpo é classificado acima da mediana da abordagem BASE_SUP. Na Figura 5 pode ser observada a classificação de cada espécie separadamente. É possível perceber algumas espécies com resultados melhores que outras

como, por exemplo, a espécie Dourado, pois ela tem pouca variação das três abordagens usadas nesse experimento, sugerindo que ela pode ser mais fácil de realizar a classificação.

Técnica	BASE_SUP	BASE_NSUP
LPI_128	$0,3 \times 10^{-3}$	$0,61 \times 10^{-5}$
BASE_SUP	—	0,558

Tabela 3: Valor de p obtido a partir da execução do teste Tukey, usando a Medida-F aplicada para cada técnica.

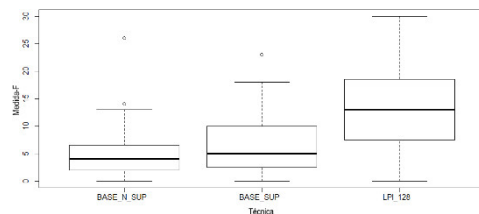


Figura 4: Diagrama de caixas mostrando o desempenho das abordagens baseado na Medida-F.

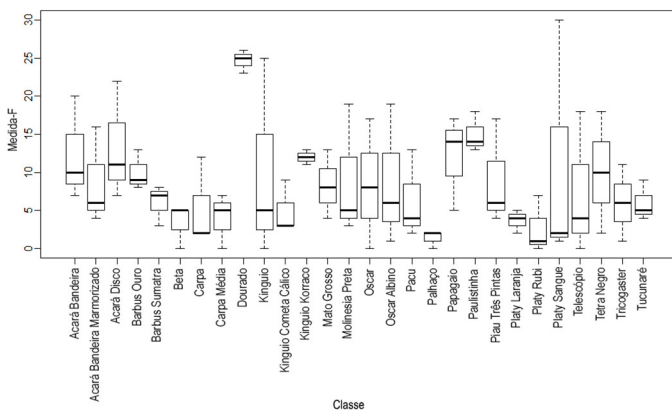


Figura 5: Diagrama de caixas mostrando o desempenho das classes individualmente baseado na Medida-F para a abordagem LPI_128.

VI. CONCLUSÕES

Esta nova abordagem mostrou-se vantajosa para seguir com novos estudos e aprimoramentos, uma vez que houve um bom desempenho considerando a complexidade do problema envolvido e ao aumento de desempenho. Como pôde ser visto, a abordagem proposta por este trabalho aumentou em 97% a Medida-F usando um novo banco de imagem de imagens de peixes. Além disso, ela se mostrou estatisticamente distinta das abordagens comparadas, uma vez que rejeitou a hipótese nula ao aplicar o teste estatístico ANOVA com o intervalo de confiança de 95% ($p < 0,05$). Assim, utilizando a abordagem deste trabalho, obteve-se um aumento do número de pontos que são corretamente detectados e relevantes para se ter uma melhor performance do classificador, aprimorando a identificação entre os pontos. A análise apresentada traz evidências de que a abordagem para redução do número de pontos de interesse pode

melhorar o espaço de características construído por meio das características visuais. Além disso, forneceu uma melhora significativa avaliando o desempenho computacional, em que o tempo de processamento foi 300 vezes mais rápido. Como trabalhos futuros, será verificado se a abordagem proposta irá melhorar a Medida-F em outras bases de imagens.

AGRADECIMENTOS

Este trabalho recebeu apoio financeiro da Universidade Católica Dom Bosco, UCDB e da Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia do Estado de Mato Grosso do Sul, FUNDECT. Um dos autores recebeu bolsa da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES.

REFERÊNCIAS

- [1] D. G. LOWE, Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [2] J. SIVIC J and A. ZISSERMAN, “Video Google: A text retrieval approach to object matching in videos”. In *Proceedings of ICCV, Nice, France*, volume 2, oct, pp. 1470–1477, 2003.
- [3] G. CSURKA, C. R. DANCE, L. FAN, J. WILLAMOWSKI and C. BRAY, Visual categorization with bags of keypoints. In *Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.
- [4] H. I. AKRAM, C. DE LA HIGUERA, H. XIAO and C. ECKERT, “Grammatical inference algorithms in matlab”. In *ICGI 2010: Proceedings of the 10th International Colloquium on Grammatical Inference*. Valencia, Spain: Springer-Verlag, 2010.
- [5] M. R. BORTH, L. C. RIBAS L. C, H. PISTORI, H. W. N. GONÇALVES and A. A. C. JUNIOR, “Classificação de espécies de peixe usando inferência gramatical no reconhecimento de padrões em problemas de Visão Computacional”. Em: *Memórias do X Workshop de Tecnologia Adaptativa - WTA 2016*. EPUSP, São Paulo. pp. 60-69, 2016.
- [6] K. KIRA and L. A. RENDELL, “The feature selection problem: traditional methods and a new algorithm”. In: *Conference on Artificial Intelligence*, vol. 2, pp. 129-134, 1992.
- [7] M. PEURA and J. LIVARINEN, “Efficiency of Simple Shape Descriptors”. In: *Aspects of Visual Form*. World Scientific, Singapore, pp. 443–451, 1997.
- [8] J. R. ANDERSON, “Cognitive Psychology and its Implications”. New York: W. H. Freeman and Company, 4 ed, 1995.
- [9] M. E. J. NEWMAN, “Power Laws, Pareto Distributions and Zipf's Law”. *Statistical Mechanics. Contemporary Physics*, 46, pp. 323-351, 2006.
- [10] E. BRYNJOLFSSON, Y. HU and D. SIMESTE, “Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales”. SSRN, 2006 Disponível em <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=953587>. Acesso em 09 sep. 2016.
- [11] N. P. KOCH, “Software Engineering for Adaptive Hypermedia Systems Reference Model, Modeling Techniques and Development Process”. Tese de Doutorado em Engenharia de Software. Ludwig – Maximilian – Universität Munchen. Munique, 2000.
- [12] G. Hripcsak and A. S. Rothschild, "Agreement, the F-Measure, and Reliability in Information Retrieval". *Journal of the American Medical Informatics Association*, 12.3, pp. 296-298, 2005

CAPÍTULO III

O artigo a seguir foi submetido para a revista *Pattern Analysis and Applications*³⁷ e estamos aguardando o parecer dos avaliadores. Este artigo apresenta a abordagem completa desta Tese. A abordagem usa a limitação de pontos de interesse juntamente com o uso de cores no vetor de descrição para descrever os pontos de interesse.

Devido ao SIFT utilizar apenas informações em tons de cinza para descrever os pontos de interesse e considerando a complexidade do banco de imagens deste trabalho, foi utilizado informações de cores nos descritores utilizando o algoritmo Opponent-SIFT. Neste artigo, é utilizada a limitação da quantidade de pontos de interesse para a classificação de imagens juntamente com o uso de informações de cores nos descritores de cada ponto de interesse. Por conta dos peixes terem uma grande variedade de cores, inclusive para uma mesma espécie, utilizamos um descritor que usa cores para descrever melhor os pontos de interesse e ajudar na classificação de imagens.

³⁷ Acesso ao website da revista: <http://paaa.edmgr.com/>

A New Approach for Image Classification Applying Reduction of Colored Keypoints

Marcelo Rafael Borth
IFPR – Federal Institute of Paraná
Umuarama, Brazil
marcelo.borth@ifpr.edu.br

Wesley Nunes Gonçalves
UFMS – Federal University of Mato Grosso do Sul
Ponta Porã, Brazil
wesley.goncalves@ufms.br

Hemerson Pistori
UCDB – Dom Bosco Catholic University
Campo Grande – Brazil
pistori@ucdb.br

Abstract — There are algorithms for feature extraction such as SIFT and Opponent-SIFT which detect and describe keypoints. In image classification, it is common to have an image dataset. So when using an algorithm to detect and describe local features in a set of images, the number of keypoints detected by class can be disproportionate, resulting in a biased classification for classes with higher amount of keypoints. This paper presents a novel approach to reduce the number of keypoints (and colored keypoints) in images after the feature extraction process, so computer vision techniques can be applied to image classification problems. This approach uses Zipf's Law and Pareto Principle to conduct the new strategy to reduce keypoints. An experiment was conducted comparing 4 different strategies. Results are encouraging and the proposal opens new paths of exploration for keypoints reduction and syntactical pattern recognition. The classification reached a F-Measure of 76,8% and the computer performance (execution time) has increased from 9 to 1900 times.

Keywords — *Keypoints Reduction; Image Classification; Image Processing; Pattern Recognition; Grammatical Inference.*

1. INTRODUCTION

One of the most challenging problems in computer vision is the image classification. The analysis to improve the process of obtaining representative features is not trivial, once the feature extraction process is one of the most important steps for object/scene classification. For this, keypoints are used often in analysis of images. A consolidated and widely used algorithm for detecting and describing keypoints is the Scale-Invariant Feature Transformation (SIFT) proposed by Lowe [1], and its variation with color information descriptors is Opponent-SIFT [2]. This algorithm extracts from the image a collection of local features, describing the region around the keypoint.

When applying an algorithm like SIFT in an image dataset it is common to have different numbers of keypoints for each image. The more keypoints identified in the image, the greater will be its own descriptive power, which may support in the classification process. Considering a dataset divided into classes, it is possible some classes has much more keypoints detected than others do. This is common because the detection algorithms do not limit the number of keypoints per image, since they analyze the importance of the keypoint based on a threshold. Thus, algorithms detect keypoints without care

about the amount detected. In front of this, there is the problem of the unbalanced numbers of keypoints per class. In this way, it is possible these classes have many keypoints while others have few, and may lead to a biased classification for those classes with greatest descriptive power. Therefore, balancing the number of keypoints is essential to improve discrimination power.

This work presents a new approach to reduce keypoints extracted by local descriptors algorithms such as SIFT and Opponent-SIFT. Reduction of keypoints aims to eliminate noises and redundant information, as well as equalize the number of keypoints for each class of the image dataset. As consequence, it reduces the execution time of the classification process. From the information reduction, it is possible to improve the quality of the model built during the image classification process, leaving it unbiased to image classification. Therefore, keypoints reduction decrease the total number of keypoints to compose the new subset of information, increasing the percentage of correctness of the classifier and minimizing the complexity of the knowledge generation, such as conceptual models, automata, etc. Instead, this work does the reduction of keypoints for images and not the dimensionality reduction of the descriptor vector created by the detection algorithm.

The main idea of this classification process is to extract and limit the number of colored keypoints. From keypoints, we can create visual words to label and order them, producing a sentence of size n , such that n is the number of keypoints. The alphabet size is defined using the k-means algorithm, as in the bag of visual words [3-4]. The alphabet, also known as dictionary, is a finite set of symbols, numbers, or characters. Thus, the sentence is constructed based on the size or distinction of an alphabet, since its size determines how well an image is described computationally. Before the image classification process, this work order keypoints in each image, producing a sentence. A sentence, also known as word or string, is a finite sequence of ordered visual words, for example, "ABACDECBCADA." In order to generate the sentence, the ordering is based on the hierarchical clustering of keypoints, joining keypoints based on the spatial position (x, y) until it forms a complete binary tree. A visual word is only a symbolic representation, by letters, of the keypoint descriptor extracted by Opponent-SIFT. Experiments were

carried out using fish images in order to analyze the approach in a possible real problem. The grammatical inference algorithm k-testable [5] was used.

2. WHY REDUCE KEYPOINTS?

In the image classification process, it is possible to have a large amount of information to represent images. In order to improve and optimize the classification, it becomes necessary to separate relevant features from those are not relevant. In this work, keypoints reduction is performed before keypoint description. In practice, it is possible to imagine the greater the number of keypoints, the greater the amount of information available for classification, increasing as consequence the processing time. Therefore, keypoints reduction allows filtering according to a criterion of importance, reducing the number of possible noises or irrelevant keypoints.

Huge difference in the amount of information of each class can reduce correctness in classification from one to another. Learning algorithm can be impaired in its performance, because of the amount of data, and in the percentage of correct classification, because of redundant information can confuse the algorithm, harming it in the construction of a suitable model for the exposed knowledge [6]. When viewing the confusion matrix presented by [7], it is possible to note the class predominance in the classification. For example, in the experiment, a single class ranked 37% of total images contained in the dataset. This is not so bad until consider the dataset has 28 classes and each one should have only ~3.5% of classified images. In general, classes with greater descriptive power got more images classified, since they described more characteristics when compared to classes with less keypoints detected. According to Peura and Livarinen [8], the ability to recognize patterns of an image over a set of images depends on the amount of information is known a priori of the respective object. Therefore, Peura and Livarinen [8] point out one of the main problems faced in the imbalance of features in the classification.

Unbalance can harm the classifier when new examples need to be classified, providing a biased classification. Moreover, this situation may be worse as the class prevalence becomes more evident, being able to increase the precision of those classes with more keypoints and reduce from those classes with less keypoints.

For this reason, we propose an approach to reduce the number of keypoints after the feature extraction phase, leaving the number of keypoints similar (or very close to) for each class. The aim of equalizing the number of keypoints between classes is relevant in order to determine proportionally the descriptive power between each of them, decreasing the prevalence of classes.

Figure 1 shows the number of keypoints detected from the images used in this work, separated by classes. It can be observed there is a tendency for some classes to have a significant amount of information (keypoints), while others have little information to represent the class of images. If we compare two classes, with the highest number of keypoints (Kingiuo) with the lowest one (Platy Laranja), the first has 87

times more keypoints than the second does. Kingiuo obtained 10.5% of the total keypoints from the entire image dataset, while Platy Orange obtained only 0.1%. Therefore, the difference between each other and the prevalence of information is evident.

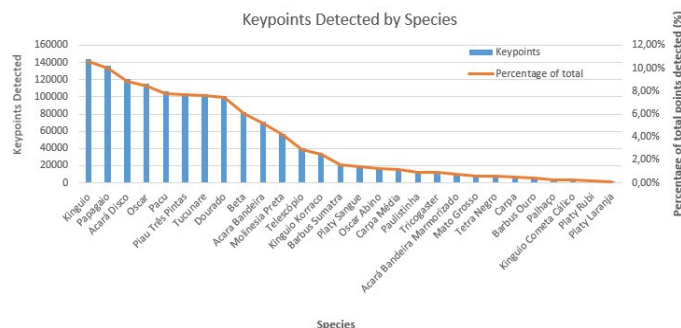


Fig. 1 Number of keypoints detected by species and percentage of total keypoints detected for whole dataset, considering the image dataset used in this work

From the distribution of the number of keypoints per class and the distribution of keypoints based on their strength (importance), we can notice in both an effect known as "long tail" [9]. Bringing it to our context, there are few keypoints with high value of importance and many keypoints with low value of importance. Those with low importance cannot become relevant to maintain for learning process and classification.

Zipf's Law [10] provides the long tail effect, where is associated with the Pareto Principle [10], also known as 80-20 rule. This rule states that 80% of causes account for 20% of effects, while the remaining 20% of causes account for 80% of effects [11-12]. Because of this, we can associate the 80-20 rule with the keypoints and deduce many of them are not representative because they are in the long tail, while only a few are representative, which justifies using only the most representative part of each set of keypoints for classification. Therefore, to equalize the amount of information between classes and leave the classification more equitable and fair, we suggest the keypoints limitation.

Figure 2 illustrates six images with the progressive keypoints limitation and another with all keypoints. The limitation per image was 32, 64, 128, 256, 512 and 1024 keypoints, respectively, 25, 26, 27, 28, 29 and 210. We can note the greater the number of keypoints, the greater the number of keypoints in the background of the image. Keypoints in background may be those less important in the image. In this case, there are hundreds of keypoints identified on background that are not relevant to fish description. Although those keypoints could be relevant to identify the context of the fish, they are not relevant for the description of fish characteristics.

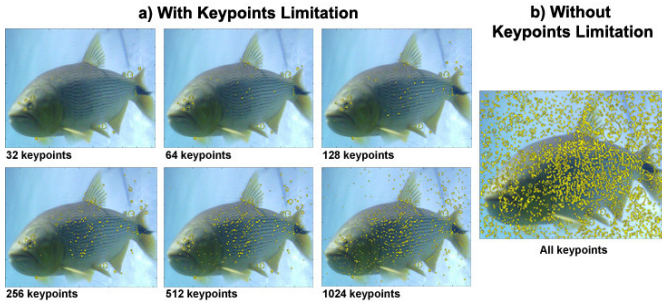


Fig. 2 (a) Six fish images presenting the keypoints with their respective limitations (maximum number of keypoints per image are indicated below of the image) and (b) one image presenting all keypoints

With keypoints reduction, fewer information are used in the process of the alphabet construction, sentence generation and automaton generation for pattern recognition, which consequently reflects directly in reducing the processing time of the whole approach.

3. THE IMAGE CLASSIFICATION PROCESS

The approach proposed in this paper consists of a supervised learning strategy that combines reduction keypoints with colored descriptors. The image classification of this work is based on the proposal developed by [13], with some improvements. The process can be divided into 7 steps:

Keypoints Extraction

First, keypoints are detected in images using SIFT. Each keypoint describes a small part of the image and the number of keypoints detected can be different from one image to another. This technique seeks to be rotational and scale invariant that is relevant for pattern recognition tasks.

Keypoints Reduction

In this step, the maximum number of keypoints N is defined to limit the best keypoints for each image. Therefore, only the N strongest keypoints are maintained based on the maximum number chosen a priori. In case of identifying less points of the maximum assigned by image, the quantity analyzed will be the total of identified points.

For this, an attribute in SIFT algorithm called *response* is responsible by the keypoint “power”, where it can be used for sorting or subsampling. From the response value, it is possible to remove keypoints that are not so strong or representative compared to others. After obtaining the importance value of each keypoint, the lowest are eliminated for each image based on the maximum number assigned.

Keypoints Description

A drawback of the SIFT is because the information colors are ignored [1], since only gray-scale are used for the descriptor computation. Because of this, we are using in this work Opponent-SIFT instead of the standard SIFT. In Opponent-SIFT, a color image is converted from RGB color space to opponent color space. Descriptor is calculated over each three channel of the opponent color space and it is a

combination of 3×128 dimension vector. Consequently, the dimension of the Opponent-SIFT descriptor is 384.

Alphabet Construction

Alphabet construction is done by supervised approach as used in [14], which implies better results as showed in previous works [7, 15]. It is based on the set of descriptors and the alphabet is constructed separately for each class. Therefore, the number of visual words P_c of the alphabet for each class c is given by:

$$P_c = k / nc$$

where, k is the alphabet size and nc is the number of classes. If k is not divisible by nc , the remainder is randomly distributed, one per class. Centroids C_p , are obtained using the k-means algorithm, applied in a subset D_p containing only the descriptors belonging for class p :

$$C_p = \text{k-means}(D_p)$$

$$D_p = [\varphi_i^j], \text{ if } j \text{ belongs to class } p$$

Each class will have a set of centroids, defined a priori, called the alphabet. Finally, sets of centroids are concatenated in order to obtain the complete alphabet.

Labeling Keypoints

Given the keypoints p_i^j for image j , each keypoint is labeled with the nearest, which is the nearest centroid, as shown below:

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

where r_i^j corresponds to the label of the keypoint i of the image j , $|\varphi_i^j, C_l|$ is the Euclidian Distance and k is the alphabet size. After this step, each keypoint will have a visual word, or label, where similar keypoints will have the same visual words.

Sentence Generation

In this step, a sentence is generated with the corresponding visual words of each keypoint. To generate the sentence from the visual words in the image, the hierarchical clustering ordering was used. In this case, visual words are selected from the image according to Euclidean Distance of the spatial position from one keypoint to others, where the closest visual words are grouped in a binary hierarchical tree to generate the sentence.

Grammar Learning

The last step is the grammar inference (generation rules) for each class from the sentences, resulting in an automaton for each class. For the inference, the k-testable algorithm [5] was used and it is able to infer k-testable languages in polynomial time. In essence, this algorithm finds prefixes, substrings and suffixes that occur in the training data [16]. Variable k of the k-testable algorithm means a quantity of k characters prior to the current character analyzed in grammatical inference. Implementation of k-testable available with GI Toolbox for Matlab has been used in this work [5].

We have used $kt = 2$, since using a greater value do not improved the results.

At this point, the sentences representing the test images are verified in the automata. In order to check a sentence, the number of errors is counted for each automaton. Errors occur when there is a character in the sentence that does not belongs to the language or there is no transition in the language of an automaton. The sentence tested will belong to the automaton that returns the fewest errors.

4. EXPERIMENT

In this section, we describe the experiments properties in a real application of fish species recognition. The AQUARIO28E40I dataset contains 40 images for each of 28 fish species, where all classes can be seen in Figure 3. In total, the dataset contains 1120 colored images divided into 28 classes. Images has variations from one to another in size and resolution, as they were captured from different cell phones and then manually cropped.

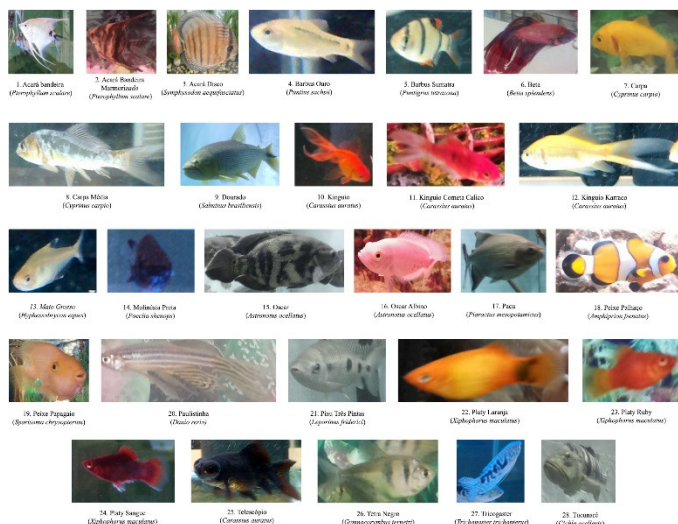


Fig. 3 All 28 fish species contained in the image dataset AQUARIO28E40I.

To divide the training and test dataset, the 10-fold cross-validation model was used. The F-Measure [17], also known as F-Score, was adopted for the comparison between the strategies once is a measure of a test's accuracy. This metric is the harmonic mean of precision and recall.

5. RESULTS AND DISCUSSION

Table 1 presents the results of the experiments. The best accuracy obtained using the proposed approach was 76.8%, using up to 256 keypoints per image.

Alphabet Size	Maximum number of keypoints per image					
	32	64	128	256	512	1024
32	39.6	32.0	33.8	35.2	26.6	16.7
64	37.3	42.5	41.7	43.1	42.3	27.6
128	29.5	39.3	44.4	49.4	49.2	40.0
256	39.5	41.1	49.2	50.2	53.0	57.9
512	74.4	71.9	72.5	72.1	73.2	64.5
1024	59.1	67.7	75.4	*76.8*	71.9	72.5
2048	54.6	69.7	66.7	67.2	71.7	73.4
3072	57.4	49.8	49.4	60.3	48.9	59.6
4096	57.1	43.6	66.5	54.0	69.4	69.2
5120	53.9	64.1	60.7	69.4	74.5	71.0
6144	58.5	46.2	62.7	51.3	52.3	63.8
7168	50.8	60.1	60.6	50.9	74.2	66.7
8192	44.2	59.8	51.9	63.1	74.8	55.9

Table 1 F-Measure for each configuration of the proposed approach. Highest values for each maximum number of keypoints by image are marked in bold. Highest F-Measure is highlighted and displayed between two asterisks

Figure 4 shows the confusion matrix for the best result obtained in this work. Vertical and horizontal numbers represented in Figure 4 correspond to each fish numbered in Figure 3. Confusion matrix shows the correct and incorrect classifications for each species. All correct guesses are located in the diagonal, so it is easy to visually inspect the misclassification, as they are represented by values outside the diagonal. Therefore, it is noticed the correct classification values stands out in the matrix's diagonal. In front of this, we believe the misclassifications is due to the complexity of the dataset, since the classes have very similar characteristics among them, which takes this task a high degree of complexity. The values in the vertical increase and then decrease, this is due to finding an approximate point of the ideal for that description. In this case, the alphabet finds an alphabet size good enough to describe this type of problem, and decreasing or greatly increasing that value will confuse the classification algorithm by not describing it so well with a smaller or larger alphabet.

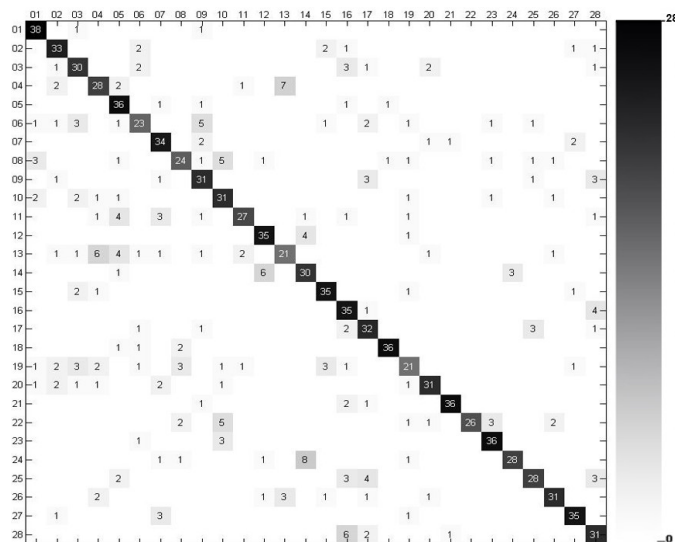


Fig. 4 Confusion matrix for the best result of the experiment, with the F-Measure of 76,8%

To show the performance of this approach, we compare it with 3 other similar approaches in the classification process, once each one is based on the main strategy proposed by [13]. The approaches compared are indicated by the names:

1. **BASE**: Base inference grammatical approach for image classification without keypoint limitation and colored keypoints description (base approach based on [13]);
2. **BASE + LPI**: Base inference grammatical approach using keypoints limitation [15];
3. **BASE + OppC**: Base inference grammatical approach using colors to describe keypoints (proposed in this work and used only for comparing results);
4. **BASE + LPI + OppC**: Base inference grammatical approach using keypoints limitation technique and colors to generate the keypoint descriptors (**proposed in this work**).

Comparing the results of these approaches, the study proposed in this work (BASE + LPI + OppC) obtained the best result, as we can see in Figure 5. The F-Measure of each approach was 16.9% (± 5.4), 33.4% (± 3.4), 71.4% (± 4.7) and 76.8% (± 4.2), respectively, from the lowest to highest. We can note the use of color in keypoint descriptors had relevant results with the approach BASE + OppC, since by itself, there is already an improvement in the result. Nevertheless, its combination with the limitation of keypoints (BASE + LPI + OppC) had the best result overall.

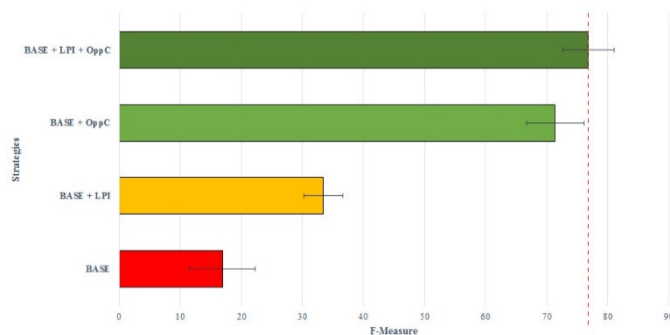


Fig. 5 F-measure of the 4 strategies. Error bars indicate the standard deviation of the F-Measure

We believe the improvement of this result comes from the relevance of the keypoints, once the algorithm analyze only the best N points for each image. As shown in [15], the excessive amount of information and its use without keypoints limitation has more noises in the data, reducing the performance of the classifier, since most of keypoints are detected in the image background and not in the fish.

In real applications, issues such as color change, irregular lighting, different background, sediments in the water and underwater plants can generate noises in the extraction process, influencing the classification. In addition, recognition process is not a trivial task, since the fish may appear in images with different sizes, shapes, scales, orientations, distortions,

colors (even for the same species) and background context. All of these challenges harms the image classification process.

When comparing the four strategies, the approach proposed in this work stands out in relation to the other strategies. We performed ANOVA test with 95% confidence interval ($p < 0.05$), followed by Tukey's HSD (honest significant difference) post-hoc test. Summarizing, p-value was $< 2 \times 10^{-16}$. Therefore, we can reject the null hypothesis and conclude the approaches are statistically different. However, comparing pair-wise the approach proposed in this work (BASE+LPI+OppC) exceeds two of them. In comparison, BASE+LPI+OppC is far superior than BASE and BASE+LPI, but when compared to BASE+OppC, the F-Measure is greater but it does not proved statistically difference, since the p-value was 0.905.

Figure 6 presents the box plot diagram for each approach based on F-Measure. It can be seen that BASE+OppC has the best median value. However, BASE+LPI+OppC is less volatile and has more consistent results, which makes it more attractive to use in applications. Box plot shows the use of the keypoints limitation in the approach BASE+LPI+OppC can remove noises in the information extraction, since the minimum whisker is smaller than the approach BASE+OppC, besides to removing the two outliers. We consider the limitation of keypoints so important that using only colors characteristics would not have all the positive effects, in both quality and performance aspects.

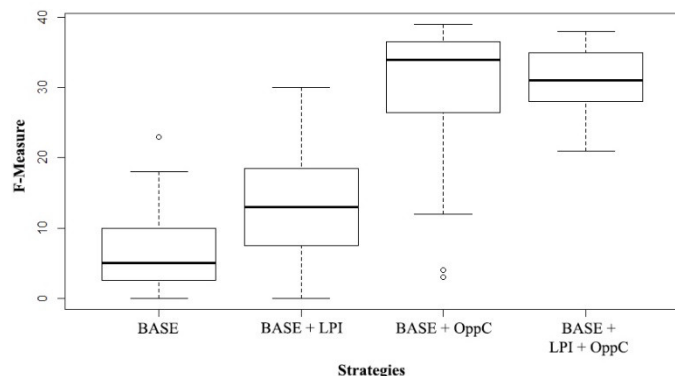


Fig. 6 Box plot diagram showing the performance of each strategies based on the F-Measure measurement

Considering the processing time in the image classification process for the best result of each one, the approach proposed in this work (BASE+LPI+OppC) has an execution time 9x better than BASE+OppC and 1900x than BASE. So bearing in mind all results and performances, we consider this work is feasible to continuous studying and use ahead in an application.

6. FINAL CONSIDERATIONS

This new approach showed good results considering the complexity of the problem involved and the increased performance compared to other approaches. The main contribution of this work is the proposal of a novel approach

to reduce keypoints, which can head a way for new improvements in this area when applied to image processing.

The approach proposed in this work has the best F-Measure compared to other approaches, reaching 76.8%. Furthermore, applying the keypoints reduction approach produced a decrease from 9 to 1900x in the computational execution time, which makes the approach even more interesting. Thus, it was obtained an increase in the number of keypoints that are correctly detected and relevant for a better performance of the classifier, improving the knowledge generation for the classification. The analysis presents evidence the keypoints reduction improves the information space constructed through the visual characteristics. In addition, the proposed BASE+LPI+OppC can extract and maintain more effective and discriminative local features for object/scene classification.

Suggestion for future work includes the study of new forms to sentence generation; expand to other grammatical inference algorithms and others types of error recovery strategy; use adaptive automata to perform classification; and verifying the feasibility of using regular grammars as well as positive and negative examples may be an alternative to improve classification performance.

ACKNOWLEDGEMENTS

This work has received financial support from the Dom Bosco Catholic University, UCDB, the Foundation for the Support and Development of Education, Science and Technology from the State of Mato Grosso do Sul, FUNDECT, and the National Council for Scientific and Technological Development (CNPq). Some of the authors have been awarded with Scholarships from the Coordination for the Improvement of Higher Education Personnel, CAPES.

REFERENCES

- [1] Lowe DG (2004) Distinctive image features from scale-invariant keypoints. *International journal of computer vision*. Nov 1;60(2):91-110.
- [2] Van De Sande K, Gevers T, Snoek C (2010) Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*. Sep;32(9):1582-1596.
- [3] Sivic J, Zisserman A (2003) Video google: A text retrieval approach to object matching in videos. In *ICCV*, 2(1470), pp 1470-1477.
- [4] Csurka G, Dance C, Fan L, Willamowski J, Bray C (2004) Visual categorization with bags of keypoints. In *Workshop on statistical learning in computer vision, ECCV*, vol 1:1-22.
- [5] Akram HI, De La Higuera C, Xiao H, Eckert C (2010) Grammatical inference algorithms in MATLAB. In *International Colloquium on Grammatical Inference*, Springer Berlin Heidelberg, pp. 262-266, 2010.
- [6] Kira K, Rendell LA (1992) The feature selection problem: Traditional methods and a new algorithm. In *AAAI*, 2:129-134.
- [7] Borth MR, Ribas LC, Pistori H, Gonçalves WN, Junior AA (2016) Classificação de Espécies de Peixe usando Inferência Gramatical no Reconhecimento de Padrões em Problemas de Visão Computacional. Em: *Memórias do X WTA 2016*. EPUSP, São Paulo. pp. 60-69.
- [8] Peura M, Iivarinen J (1997) Efficiency of simple shape descriptors. In *Proceedings of the third international workshop on visual*, 443: 1-9.
- [9] Anderson JR (1990) *Cognitive psychology and its implications*. 4ed, WH Freeman/Times Books/Henry Holt & Co.
- [10] Newman ME (2005) Power laws, Pareto distributions and Zipf's law. *Contemporary physics*. 46(5):323-351.
- [11] Brynjolfsson E, Hu Y, Simester D (2011) Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*. 57(8):1373-1386.
- [12] de Koch, NP (2001) *Software Engineering for Adaptive Hypermedia Systems-Reference Model, Modeling Techniques and Development Process*. Thesis, Universitat Munchen.
- [13] Pistori H, Calway A, Flach P (2013) A new strategy for applying grammatical inference to image classification problems. In *ICIT*, pp 1032-1037.
- [14] Ribas LC, Borth MR, Castro Jr AA, Gonçalves WN, Pistori H (2014) Grammatical Inference and SIFT for Scene Recognition. In *X Workshop de Visão Computacional, Uberlândia-MG*, pp 293-298.
- [15] Borth MR, Gonçalves WN, Pistori H (2016) Keypoints Reduction Applied to Image Classification Problems. *XII Workshop de Visão Computacional, Campo Grande-MS-Brazil*, pp. 382-387.
- [16] De la Higuera C (2010) *Grammatical inference: learning automata and grammars*. Cambridge University Press.
- [17] Hripcsak G, Rothschild AS (2005) Agreement, the f-measure, and reliability in information retrieval. *Journal of the American Medical Informatics Association*. 12(3):296-298.

CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi criar uma abordagem baseada no reconhecimento sintático de padrões que seja capaz de classificar espécies de peixes, a partir de imagens capturadas por celulares. Como não foi encontrado nenhum banco de imagens de peixes nesse estilo na literatura, foram criados 3 bancos de imagens distintos. Além disso, a proposta consiste no reconhecimento de espécies de peixe a partir de algoritmos de inferência gramatical, na implementação da abordagem proposta e, na validação da proposta com experimentos baseados nos bancos de imagens criados.

Foram realizados três trabalhos com resultados graduais e com melhoras em cada uma das adaptações, até chegar na abordagem final deste trabalho (Capítulo 3). Ao analisar os resultados da abordagem proposta neste trabalho, conforme apresentada nos Capítulos 1, percebe-se que para o banco de imagem Aquario10e, a maioria dos pontos de interesse são detectados no fundo da imagem e não no peixe. Por outro lado, para o banco de imagens AquarioSeg10e esse problema não ocorre, pois as imagens estão segmentadas. Todavia, é inviável adotar estratégias de classificação com milhares de imagens realizando a segmentação manual. Assim, foi desenvolvida a estratégia apresentada no Capítulo 2 para limitar os pontos de interesse e ter uma redução daqueles pontos não tão significantes para a classificação. A proposta de limitação mostrou-se viável e que é possível reduzir pontos de interesse ruidosos identificados em imagens. Não obstante, mesmo aplicando a limitação de pontos de interesse, percebeu-se que eles perdiam informações no seu descritor, por conta do algoritmo SIFT utilizar apenas informações em tons de cinza para descrever o ponto, desconsiderando informações de cores. Diante disso, foi desenvolvida a estratégia da limitação de pontos de interesse juntamente com o uso da informação de cor na descrição do ponto, apresentada no Capítulo 3. Os resultados apresentados da Medida-F são crescentes conforme os

capítulos são apresentados, e de bons resultados, considerando a complexidade do problema enfrentado para classificar espécies de peixes tão semelhantes e com características tão próximas umas das outras. O melhor resultado obtido da Medida-F foi de 76.8%. Além disso, o tempo de processamento de um classificador pode ser dividido entre aprendizado e classificação. E, considerando apenas o tempo de classificação, esta abordagem melhora o tempo de processamento em até 1900 vezes se comparada a uma utilização normal sem a limitação de pontos de interesse.

Uma das principais dificuldades enfrentadas para o desenvolvimento desta tese foi o alto tempo de execução para cada rodada de execução da configuração do experimento. Para a classificação deste trabalho, são analisados muitas informações extraídas de imagens que, por sua vez, possuem o processo de geração do alfabeto, ordenação das palavras visuais, construção dos autômatos e classificação. Embora tenha sido reduzido consideravelmente o tempo de execução da aprendizagem (teste) com o uso da limitação de pontos de interesse, o tempo de execução continuou sendo uma dificuldade enfrentada. Além disso, o tempo era maior na proporção em que o tamanho do alfabeto aumentava. Por conta disso, alguns experimentos demoram dias a serem executados para então ter os resultados e verificar se a alteração em algum passo da abordagem tenha provocado efeito positivo. Nesse sentido, cada adaptação da proposta que devia ser testada perdia-se um bom tempo para se ter os resultados.

Como trabalho futuros podem ser considerados alguns estudos como:

- Estudar outras formas de geração de sentença que não haja perda da informação espacial entre os pontos de interesse;
- Testar outros algoritmos de inferência gramatical, por exemplo, algoritmos que analisam exemplos positivos e negativos;
- Replicar os experimentos usando autômatos adaptativos para realizar a classificação;
- Estudar locais de corte na árvore binária gerada a partir da ordenação hierárquica para a geração da sentença. Com esses corte, a sentença poderá ser gerada com pontos de interesse mais próximo entre si, eliminando ainda mais os pontos mais distantes que podem não fazer parte do objeto a ser analisado, mas sim do fundo da imagem; e,
- Usar outras estratégias de recuperação de erros.

REFERÊNCIAS BIBLIOGRÁFICAS

ABDULRAHMAN, W. **Overview on Industrial Vision Systems**. In International Journal of Engineering Research & Technology (IJERT), Vol. 3, n. 5, 2014.

AKRAM, H. I.; 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.

ALMEIDA, F. **O mundo dos negócios e o meio ambiente no século 21**. O meio ambiente no século 21. André Trigueiro (cord). Rio de Janeiro: Sextante, p. 123-141, 2003.

ANDERSON, J. R. **Cognitive Psychology and its Implications**. New York: W. H. Freeman and Company, 4 ed., 1995.

ASSAD, L.T.; BURSZTYN, M. **Aquicultura Sustentável**. In: Aqui cultura no Brasil: bases para um desenvolvimento sustentável. Brasília: CNPq/Ministério da Ciência e Tecnologia, p. 303-323, 2000.

BANKS, S. **Signal Processing, Image Processing and Pattern Recognition**. Prentice Hall, 1990.

BAY, H.; TUYTELAARS, T.; VAN GOOL, L. J. **SURF: Speeded-Up Robust Features**. In ECCV, p. 404–417, 2006.

BEZERRA, C. A.; KURSTEN, A. F.; SCHIAVON S. **Sistema informatizado de acompanhamento de aquicultura**. ENEGEP 2003: XXIII Encontro Nac. de Eng. de Produção – Ouro Preto, MG, Brasil, 21-24 de out, 2003.

BORGHETTI, J. R. **Estimativa da pesca e aquicultura de água doce e marinha**. Série Relatório Técnico do Instituto de Pesca (ISSN: 1678-2283) 3: p. 8-14, 2000.

BRABO, M. F.; PEREIRA, L. F. S.; SANTANA, J. V. M.; CAMPELO, D. A. V.; VERAS, G. C. **Cenário atual da produção de pescado no mundo, no Brasil e no estado do Pará: ênfase na aquicultura**. Acta of Fisheries and Aquatic Resources 4, no. 2, pp. 50-58, 2016. Disponível em:

<www.seer.ufs.br/index.php/ActaFish/article/download/5457/4607>. Acesso em 09/02/2017.

BRAMLEY, R. G. V. **Lessons from nearly 20 years of Precision agriculture research, development, and adoption as a guide to its appropriate application.** *Crop and Pasture Science*, 60 (3), pp. 197-217, 2009.

BRYNJOLFSSON, E.; HU, Y.; SIMESTE, D. **Goodbye Pareto Principle, Hello Long Tail: The Effect of Search Costs on the Concentration of Product Sales.** SSRN. 2006. Disponível em <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=953587>. Acesso em 08 ago. 2015.

BUCKUP, P. A.; MENEZES, N. A.; GHAZZI, M. S. **Catálogo das espécies de peixes de água doce do Brasil.** Rio de Janeiro: Museu Nacional. 195p., 2007.

CADIEUX, S.; MICHAUD, F.; LALONDE, F. **Intelligent system for automated fish sorting and counting.** *IEEE/RSJ International Conference on Intelligent Robots and Systems*, v.2, pp. 1279-1284, 2000.

CAI, L.; MALMBERG, R. L.; WU, Y. **Stochastic modeling of RNA pseudoknotted structures: a grammatical approach.** *Bioinformatics*, v. 19, p. 166-173. Oxford University Press, 2003.

CHACON, S. **Git official site.** 2010. Disponível em: <<http://git-scm.com/>>. Acesso em: 04 Set, 2015.

CHARIKAR, M.; LEHMAN, E.; LIU, D.; PANIGRAHY, R.; PRABHAKARAN, M.; SAHAI, A.; SHELAT, A. **The smallest grammar problem.** *IEEE Transactions on Information Theory*, v. 51, n.7, p. 2554-2576. IEEE Press, 2005.

CRUZ-ALCÁZAR, P.P.; VIDAL-RUIZ, E. **Learning Regular Grammars to Model Style: Comparing Different Coding Schemes.** In: *Grammatical Inference, ICGI'98*, number 1433. In: *Lecture Notes in Artificial Intelligence*, pages 211-222. Springer Verlag, 1998.

CSURKA, G.; DANCE, C.; FAN, L.; WILLIAMOWSKI, J.; BRAY, C. **Visual categorization with bags of keypoints.** In workshop on Statistical Learning in Computer Vision, *ECCV*, pp. 59–74, 2004.

DURBIN, R.; EDDY, S. R.; KROGH, A. L.; MITCHISON, G. **Biological Sequence Analysis - Probabilistic Models of Proteins and Nucleic Acids.** New York: Cambridge University Press, 356 p. 1998.

FAO. **The State of world fisheries and aquaculture.** Rome, 2016. Disponível em: <<http://www.fao.org/3/a-i5555e.pdf>>. Acesso em: 09/02/2017.

FAUGERAS, O. **Three-dimensional computer vision: a geometric viewpoint**. MIT Press, 1993.

FEITOSA, F. D.; da SILVA, C. A. **Avaliação do Teste de Desempenho Geral do Aquisys Pelos Produtores de Tilápia de Sergipe Acessado Via Internet**. Embrapa Tabuleiros Costeiros-Artigo em anais de congresso (ALICE). In: SEMINÁRIO DE INICIAÇÃO CIENTÍFICA E PÓS-GRADUAÇÃO DA EMBRAPA TABULEIROS COSTEIROS, 4, Brasília, DF: Embrapa, 2014.

FORSYTH, D. A.; PONCE, J. **Computer Vision: A Modern Approach**. 2nd Edition. Prentice Hall, 2012.

FOUNTAS, S.; BLACKMORE, S.; ESS, D.; HAWKINS, S.; BLUMHOFF, G.; LOWENBERG-DEBOER, J.; SORENSEN, C. G. **Farmer experience with precision agriculture in Denmark and the US Eastern Corn Belt**. Precision Agriculture 6 (2), 2005.

FRIEDMAN, M. **The use of ranks to avoid the assumption of normality implicit in the analysis of variance**. Journal of the American Statistical Association (American Statistical Association), 32, 675–701, 1937.

FU, L. M. **Rule Generation from Neural Networks**. IEEE Transactions on Systems: Man and Cybernetics, v. 28, n. 8, p. 1114-1124, 1994.

GARCIA-GOMEZ, J. M.; BENEDI, J. M. **Corpus based learning of stochastic context-free grammar combined with Hidden Markov models for tRNA modelling**. In: Proceedings of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS'04), p. 2785-2788. IEEE Press, 2004.

GEERTZENL, J.; ZANEN, M. V. **Grammatical Inference Using Suffix Trees**. G. Paliouras and Y. Sakakibara (Eds.). In: Grammatical Inference and Applications, ICGI 2004, in: Lecture Notes in Artificial Intelligence, LNAI 3264, p. 163-174, 2004.

GODINHO, H.; GODINHO, A. **Fish Communities in Southeastern Brazilian River Basins Submitted to Hydroelectric Impoundments**. Acta Lim. Brasiliensia, v.5, p.187–197, 1994.

GONZALEZ, R. C.; WOODS, E. R. **Digital Image Processing**. 3rd Edition. Prentice Hall, 2007.

GRIFFIN, T. W.; LOWENBERG-DEBOER, J.; LAMBERT, D. M.; PEONE, J.; PAYNE, T.; DABERKOW, S. G. **Adoption, profitability, and making better use of precision farming data**. Staff Paper. In: Department of Agricultural Economics, Purdue University, 2004.

HANSEN, J.D.; ZAPATA, A.G. **Lymphocyte development in fish and amphibians**. Immunological Reviews, Volume 166, pp. 199–220, 1998.

HARTLEY, R.; ZISSERMAN, A. **Multiple View Geometry in Computer Vision**. Cambridge University Press, 2004.

HIGUERA, C. A. **Bibliographical study of grammatical inference**. Pattern Recognition, v. 38, n. 9, p. 1332-1348. Elsevier, 2005.

HOLLANDER, M.; WOLFE, D. **Nonparametric statistical methods**. Wiley Series in Probability and Statistics – Applied Probability and Statistics Section. Wiley, 1999.

IBGE. **Produção da Pecuária Municipal**. Vol. 42, pp. 1-39, 2014. Disponível em: <http://biblioteca.ibge.gov.br/visualizacao/periodicos/84/ppm_2014_v42_br.pdf>. Acesso em: 30/03/2017.

KNUDSEN, B.; HEIN, J. **Pfold: RNA secondary structure prediction using stochastic context-free grammars**. Nucleic Acids Research, v. 31, n. 13, p. 3423-3428. Oxford University Press, 2003.

KOCH, N. P. **Software Engineering for Adaptive Hypermedia Systems Reference Model, Modeling Techniques and Development Process**. Tese de Doutorado em Engenharia de Software. Ludwig – Maximilian – Universität München. Munique, 2000.

LIMA, M. F. S. **Inovapeixe: a inovação da piscicultura através do software livre**. Anais do Congresso Nacional Universidade, EAD e Software Livre. v.2, n.2, 2012.

MARTINEZ-GOMEZ, J.; GARCIA-VAREA, I.; CAPUTO, B. **Overview of the ImageCLEF 2012 Robot Vision Task**. CLEF (Online Working Notes/Labs/Workshop). 2012.

MATSUNO, I. P. **Um Estudo dos Processos de Inferência de Gramáticas Regulares e Livres de Contexto Baseados em Modelos Adaptativos**. Dissertação (Mestrado em Engenharia Elétrica) - Universidade de São Paulo, São Paulo-SP, 121 p., 2006.

MENEZES, P. B. **Linguagens Formais e Autômatos**. 6ed. Porto Alegre: Bookman, Série livros didáticos informática UFRGS, 256 p., 2011.

MICLET, L.; ONCINA, J.; CARRASCO, R.; CASACUBERTA, P.; EYRAUD, R.; EZEQUEL, P.; FERNAU, H.; MURGUE, T.; THOLLARD, F.; VIDAL, E. **Applications of Grammatical Inference**. Grammatical Inference, 2004.

MORDESON, J. N.; MALIK, D. S. **Fuzzy Automata and Languages: Theory and Applications**. Boca Raton: Chapman & Hall/CRC, 576 p., 2002.

MPA, Ministério da Pesca e Aquicultura. **Boletim Estatístico da Pesca e Aquicultura**. Brasília – DF, 60p., 2011. Disponível em: <<http://bibspi.planejamento.gov.br/handle/iditem/191>>. Acessado em: 30 de março de 2017.

NELSON, J. S. **Fishes of the World**. Nova Jersey, EUA: John Wiley e Sons, Inc., n.4, 600 p., 2006.

NERY, M. S. **Determining the Appropriate Feature Set for Fish Classification Tasks**. Graphics, Patterns and Images, SIBGRAPI Conference on, pp. 173-180, 2005.

NEVILL-MANNING, C. G.; WITTEN, I. H. **Compression and explanation using hierarchical grammars**. The Computer Journal, v. 40, p. 103-116. Oxford Journals, 1997.

NEWMAN, M. E. J. **Power Laws, Pareto Distributions and Zipf's Law**. Statistical Mechanics. Contemporary Physics, 46, pp. 323-351, 2006.

OOMMEN, B. J., KASHYAP, R. L. **A Formal Theory for Optimal and Information Theoretic Syntactic Pattern Recognition**. Pattern Recognition, vol. 31, n. 8, pp. 1159-1177, 1998.

OPENCV. **OpenCV 3.2.0 - Open Source Computer Vision**, 2017. Disponível em: <http://docs.opencv.org/3.2.0/da/df5/tutorial_py_sift_intro.html>. Acesso em: 08/02/2017.

OSTRENSKY, A.; BOEGER, W. A. **Principais problemas enfrentados atualmente pela aqüicultura brasileira**. In: OSTRENSKY, A.; BORGHETTI, J.R.; SOTO, D. Aqüicultura no Brasil: o desafio é crescer. Brasília: Secretaria Especial de Aqüicultura e Pesca/FAO, p. 279, 2007.

PERELMUTER, G.; CARRERA, E. V.; VELLASCO, M.; PACHECO, A. **Reconhecimento de Imagens Bidimensionais Utilizando Redes Neurais Artificiais**. Anais do VII SIBGRAPI, p. 197-203, 1995.

PEURA, M.; LIVARINEN, J. **Efficiency of Simple Shape Descriptors**. In: Aspects of Visual Form, pp. 443–451. World Scientific, Singapore, 1997.

PIO, J. L. S.; OLIVEIRA, C. J. S.; ARAÚJO, A. A.; OLIVEIRA, N. J. **Reconhecimento Sintático e Estrutural de Padrões**. Atualizações em Informática. Rio de Janeiro. Editora PUC-Rio: SBC, cap. 6, 2006.

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.

PORTAL DA SAÚDE. **Ministério da Saúde incentiva consumo regular de peixe**, 2012. Disponível em: <<http://portalsaude.saude.gov.br/portalsaude/noticia/6934/162/ministerio-da-saude-incentiva-consumo-regular-de-peixe.html>>. Acessado em: 14/06/2013.

PROCHMANN, Â. M.; TREDEZINI, C. A. O. **A piscicultura em Mato Grosso do Sul, como instrumento de geração de emprego e renda na pequena propriedade**. Congresso da sociedade brasileira de economia e sociologia rural, v.42, 2004.

REIS, R.E., KULLANDER S.O.; FERRARIS JUNIOR, C. J. **Check list of the freshwater fishes of South and Central America**. Editora da Pontifícia Universidade Católica do Rio Grande do Sul, Porto Alegre, 729 p., 2003.

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.

RODRIGUES, P. O. A.; [et al.]. **Piscicultura de água doce: multiplicando conhecimentos**. Brasília, DF: Embrapa, 440 p., 2013.

ROSS, B. J. **The evolution of stochastic regular motifs for protein sequences**. New Generation Computing, v. 20, n. 2, p. 187-213. Ohmsha, 2002.

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.

RUSSEL, S.; NORVIG, P. **Artificial Intelligence: A Modern Approach**. 2ed. New York: Prentice Hall, 1132 p., 2003.

SABINO, J.; ANDRADE, L. P. **Uso e conservação da ictiofauna no ecoturismo da região de Bonito, Mato Grosso do Sul: o mito da sustentabilidade ecológica no rio Baía Bonita (Aquário Natural de Bonito)**. Biota Neotrópica, v3 (n2), pp.1-9., 2003.

SAKAKIBARA, Y. **Grammatical inference: An old and new paradigm**. Lecture Notes in Artificial Intelligence, 997:1-24, 1995.

SAKAKIBARA, Y. **Grammatical inference in bioinformatics**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 27, n.7, p. 1051-1062. IEEE Press, 2005.

SANTOS, G. M.; SANTOS, A. C. M. **Sustentabilidade da pesca na Amazônia**. Estudos avançados, São Paulo, v. 19, n. 54, Aug. 2005. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0103-40142005000200010&lng=en&nrm=iso>. Acesso em: 12 Set, 2015.

SATOLANI, M. F.; CORREA, C. C.; FAGUNDES, M. B. B. **Análise do ambiente institucional e organizacional da piscicultura no estado de Mato Grosso do Sul.** Revista de Economia e Agronegócio, v.6, n.2, 2008.

SCHALKOFF, R. J. **Pattern Recognition: Statistical, Structural, and Neural Approaches.** New York: Wiley, 1992.

SEARLS, D. B. **The language of genes.** Nature, v. 420, p. 211-217, 2002.

SHAPIRO, L.; STOCKMAN, G. **Computer vision.** New Jersey: Prentice Hall, 2001.

SIDONIO, L. *et al.* **Panorama da aquicultura no Brasil: desafios e oportunidades.** BNDES Setorial–Agroindústria 35: p. 421-463, 2012.

SIMS, C.; JOHNSON, H. L. **The Elements of Scrum.** Dymaxicon, 2011.

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.

TRUCCO, E.; VERRI, A. **Introductory Techniques for 3-D Computer Vision.** Prentice Hall, 1998.

TURECK, C. R.; OLIVEIRA, T. N. **Sustentabilidade ambiental e maricultura.** Revista Saúde e Ambiente 4.2: 22-26, 2009.

UNOLD, O. **Grammar-based classifier system for recognition of promoter regions.** In: Proceedings of the International Conference on Adaptive and Natural Computing Algorithms (ICANNGA). Lecture Notes in Computer Science, v. 4432, p. 798-805. Springer-Verlag, 2007.

VAL, A. L.; ROLIM, P. R.; RABELO, H. **Situação atual da aquicultura na região norte.** Em: Valenti, W. C., Poli, C. R., Pereira, J. A., Borghetti, J. R. Aquicultura no Brasil: bases para um desenvolvimento sustentável. Brasília; CNPQ/ Ministério da Ciência e Tecnologia, p. 247-266, 2000.

VALENTI, W.; POLI, C. R.; PEREIRA, J. A.; BORGHETT, J. R. **Aquicultura no Brasil: bases para um desenvolvimento sustentável.** Ed. CNPq/Ministério da Ciência e Tecnologia, Brasília, p. 399, 2000.

VAN DE SANDE, K. E. A.; GEVERS, T.; SNOEK, C. G. M. **Color descriptors for object category recognition.** In Conference on Colour in Graphics, Imaging, and Vision. Society for Imaging Science and Technology, v. 2008, pp. 378–381, 2008.

VAN DE SANDE, K. E. A.; GEVERS, T.; SNOEK, C. G. M. **Evaluation of color descriptors for object and scene recognition.** In IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1-8, 2010.

PATEL, V. M.; GOPALAN, R.; LI, R.; CHELLAPPA, R. **Visual domain adaptation: A survey of recent advances**. IEEE signal processing magazine, vol. 32, n. 3, p. 53-69, 2015

WINTNER, S. **On the semantics of unification grammars**. Grammars, v.6, n.2, p. 145-153, Springer, 2003.

WOYNAROVICH, E. **The Hydroelectric Power Plants and the Fish Fauna**. Verh. Int. Verein. Limnol, v.24, p.2531–2536, 1991.

ZHANG, N.; WANG, M.; WANG, N. **Precision agriculture-a worldwide overview**. Computers and Electronics in Agriculture 36, pp.113-132, 2000.

ZIPF, G.K. **Human Behavior and the Principle of Least Effort**. Addison-Wesley, Cambridge, Massachusetts, 1949.