

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 AQUARIO28E40I¹, contendo 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.

¹ O banco de imagem pode ser obtido pelo link: <http://pistori.weebly.com/datasets.html>

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, 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.

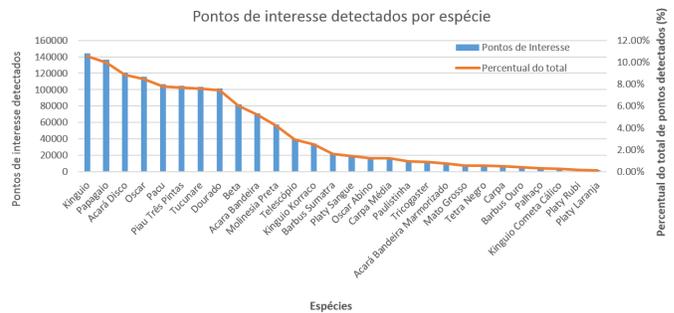


Figura 1: Total de pontos de interesse detectados por espécie considerando o banco de imagem utilizado neste trabalho.

Com a distribuição da quantidade de pontos de interesse por classe e, também, pela distribuição dos pontos de interesse baseado-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* significando 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_i do *k-testable* utilizado neste experimento foi $k_i = 2$, 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	
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.

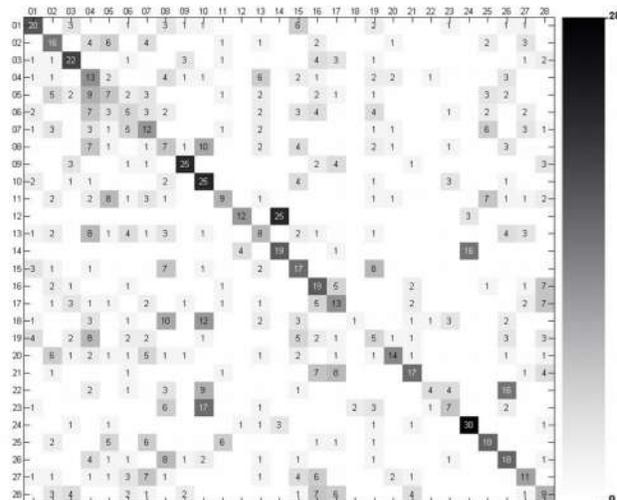


Figura 3: Matriz de confusão mostrando o número de imagens classificadas corretamente para cada espécie de peixe para o resultado de maior Medida-F, com resultado de 33,4%.

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.

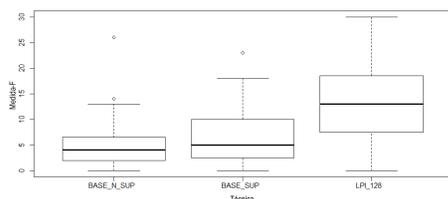


Figura 4: Diagrama de caixas mostrando o desempenho das abordagens baseado na Medida-F.

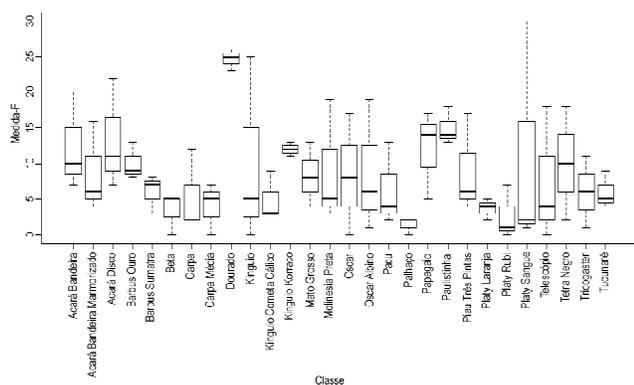


Figura 5: Diagrama de caixas mostrando o desempenho das classes individualmente baseado na Medida-F para a abordagem LPI_128.

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. Hripscak 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.