Desenvolvimento de um sistema de classificação de grãos de arroz em Visão Computacional

Adriano Rodrigues de Sousa

Vandirleya Barbosa da Costa

**1. Introdução**

O arroz é um alimento fundamental e amplamente cultivado em todo o mundo, sendo um componente básico da dieta de muitas pessoas em diversos países. É uma fonte importante de energia e fornece carboidratos complexos, bem como vários nutrientes, como vitamina B, ferro e zinco. Além disso, o arroz é geralmente mais acessível e mais barato do que outros tipos de alimentos, o que o torna um alimento vital para muitas pessoas em todo o mundo, especialmente em países menos desenvolvidos. A demanda elevada por este grão também o torna um produto agrícola importante na economia global.

Para garantir o sucesso no cultivo e na comercialização do arroz, é fundamental primeiramente classificar os diferentes tipos de grãos. Sem essa classificação, os agricultores podem enfrentar dificuldades na produção do cereal, como a seleção inadequada de sementes, a preparação incorreta do solo e a exposição a doenças e pragas. Desse modo, este trabalho propõe mitigar essa dificuldade com intuito de ajudar a facilitar essa classificação, para isso foi selecionado um conjunto de dados com 75 mil imagens de grãos de arroz, divididos em cinco classes de tipos diferentes, cada uma contendo 15 mil imagens.

O referido conjunto de dados possui imagens de grãos de arroz com resolução de 250x250, juntamente com diversas características morfológicas, de forma e cor. A padronização no tamanho das imagens, bem como a disponibilidade de informações relacionadas e a qualidade do conjunto de dados, tornaram-no uma escolha ideal para o propósito deste estudo. Além disso, durante a realização deste trabalho, serão utilizados algoritmos para selecionar uma amostra de imagens de cada classe dentro deste conjunto de dados, com o objetivo de tornar o poder computacional disponível suficiente, acelerar o processamento das imagens e dividi-las em conjuntos de treinamento e teste.

Este trabalho tem como objetivo apresentar as etapas executadas durante o desenvolvimento de um sistema completo de visão computacional para a classificação de grãos de arroz. O sistema foi criado como parte de uma disciplina e utilizou uma metodologia pré definida em cada etapa do processo. O sistema de visão computacional desenvolvido compreende a concatenação de diversas técnicas e procedimentos, incluindo a extração de atributos de textura de primeira ordem, segmentação e filtragem espacial. As próximas seções detalham cada etapa das atividades que serviram de base para o referido sistema.

**2. Atributos de textura de primeira ordem**

Para a primeira atividade, foi definido utilizar alguns atributos de textura de primeira ordem no conjunto de dados escolhido. Esses atributos são calculados a partir dos histogramas das imagens para caracterizar as propriedades da textura dessas imagens. Desse modo, foi preciso separar um subconjunto menor do conjunto original para otimizar o processamento. Esse subconjunto sem processamento é usado como base para a criação de variações através da aplicação das funções de alargamento de contraste, log e gama sobre cada pixel das imagens. Depois de criar todas as variações de subconjuntos é realizada a extração de features, que é o processo onde cada imagem é carregada e os atributos são calculados. Uma breve descrição dos atributos usados nesta atividade está na Tabela 1.

| ATRIBUTOS | DESCRIÇÃO | FINALIDADE |
| --- | --- | --- |
| Média | A média é o valor médio dos pixels em uma imagem, é calculada somando-se todos os valores de pixel e dividindo pelo número total de pixels. | Pode ser útil para identificar o nível de brilho ou escuridão da imagem. |
| Variância | A variância mede a dispersão dos valores de pixel em torno da média, ela é calculada tomando a média dos quadrados dos desvios dos valores de pixel em relação à média. | Pode ser usada para avaliar a homogeneidade ou contraste da imagem. |
| Skewness | A skewness mede a assimetria da distribuição de pixel, é calculada tomando a média dos cubos dos desvios dos valores de pixel em relação à média, dividida pelo cubo da variância | Pode ser usada para identificar se a imagem tem uma tendência para ter valores de pixel mais altos ou mais baixos do que a média. |
| Kurtosis | A kurtosis mede o "achatamento" da distribuição de pixel. Ela é calculada tomando a média dos quartos dos desvios dos valores de pixel em relação à média, dividida pela quarta da variância. | Pode ser usada para identificar se a imagem tem uma distribuição de pixel mais concentrada ou mais dispersa em relação à média. |
| Energia | A energia é calculada somando os quadrados de todos os valores de pixel da imagem. | Pode ser usado para medir a intensidade da textura da imagem. |
| Entropia | A entropia é uma medida da aleatoriedade ou da variedade de valores de pixel na imagem, é calculada a partir da distribuição de probabilidade dos valores de pixel na imagem. | Pode ser útil na análise de imagens com padrões complexos ou com texturas variadas. |

Tabela 1: Atributos de textura de primeira ordem utilizados e suas finalidades.

Esses atributos de textura de primeira ordem são úteis na análise e comparação de imagens e podem ser usados em aplicações como classificação de imagens, detecção de padrões e segmentação de imagens.

**2.1 Resultados dos atributos**

Com todos os conjuntos preparados e os atributos escolhidos foi possível realizar um processo de extração de recursos, que consistem em extrair os atributos de todas as imagens de um conjunto. Os recursos extraídos são salvos em uma matriz, onde cada linha representa os atributos de uma única imagem. O processo de extração de recursos também cria um vetor com a mesma quantidade de linhas da matriz onde cada valor define a classe da imagem dos atributos extraídos em determinada linha. Depois que todos os atributos são calculados é possível treinar um modelo de classificação Perceptron Multi-Camadas (MLP). Um modelo MLP foi treinado para cada variação do subconjunto. Por fim, um gráfico de dispersão foi criado para exibir a relação entre dois atributos, como mostrado nos gráficos abaixo.

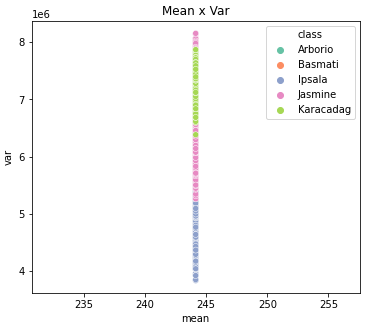


Gráfico 1: Relação entre os atributos de média e variância.

Ao observar o gráfico acima, notamos que as cinco classes de imagens formam uma linha reta, isso pode indicar que há uma forte relação linear entre a média e a variância dessas imagens, ou seja, imagens com médias mais altas tendem a ter uma variação de pixels mais alta, enquanto imagens com médias mais baixas tendem a ter uma variação de pixels mais baixa.

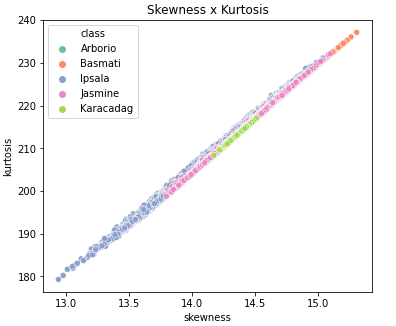


Gráfico 2: Relação entre os atributos de skewness e kurtosis.

Com base nos dados do gráfico 2, notamos que há uma relação não linear entre os atributos. Isso se deve ao fato de skewness ser uma medida de assimetria de uma distribuição de dados, enquanto kurtosis é uma medida da "achatamento" ou "curtição" da curva da distribuição de dados em comparação com uma distribuição normal. Portanto, para o conjunto de imagens em questão, a variação de pixels pode ser consistente em todas as classes de imagens.

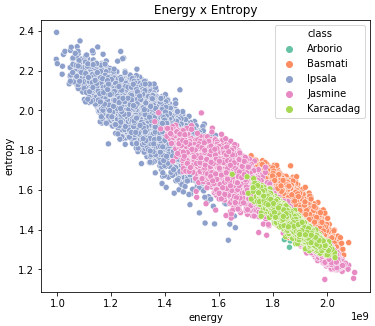


Gráfico 3: Relação entre os atributos de energia e entropia.

Ao analisar o gráfico 3, percebemos que a maior variabilidade na dispersão pode ser um indicativo de que há uma ampla variedade de imagens no conjunto de dados, com diferentes características de pixel. Desse modo, imagens com áreas muito escuras ou muito claras podem ter energias mais altas, mas entropias mais baixas, enquanto imagens com áreas mais uniformes de luz e escuridão podem ter energias mais baixas e entropias mais altas. A Tabela 1 contém os resultados dos treinamentos de todas as variações definidas na atividade considerando os seis atributos.

Tabela 1: Resultados dos treinamentos

| **Dataset** | **Acurácia** |
| --- | --- |
| original | 0.2 |
| withoutProcessing | 0.2 |
| negative | 0.18 |
| alargamentoDeContraste-v1 | 0.19 |
| alargamentoDeContraste-v2 | 0.2 |
| alargamentoDeContraste-v3 | 0.22 |
| alargamentoDeContraste-v4 | 0.19 |
| log-v1 | 0.19 |
| log-v2 | 0.16 |
| log-v3 | 0.18 |
| log-v4 | 0.2 |
| gama-v1 | 0.24 |
| gama-v2 | 0.23 |
| gama-v3 | 0.21 |
| gama-v4 | 0.19 |

A Tabela 2 contém os primeiros resultados significativos dos treinamentos. É válido ressaltar que somente algumas das variações foram usadas por causa da limitação de poder computacional. O conjunto de imagens original obteve a maior acurácia, entretanto, considerando que a variação negativa tem apenas mil imagens é possível afirmar que a acurácia pode vir a crescer aumentando sua quantidade.

Tabela 2: Resultados dos treinamentos considerando skewness e entropy

| **Dataset** | **Acurácia** |
| --- | --- |
| original | 0.74 |
| withoutProcessing | 0.5 |
| negative | 0.61 |
| alargamentoDeContraste-v1 | 0.56 |
| log-v2 | 0.54 |
| log-v4 | 0.48 |
| gama-v2 | 0.49 |

A Tabela 3 é o resultado de mais alguns testes com a intenção de encontrar uma combinação de atributos que tornasse a acurácia do modelo de classificação mais atrativa para uso. Considerando apenas os atributos de kurtosis e entropy, não houve resultados chamativos como no teste da Tabela 2, mas esclareceu que o conjunto de imagens em sua variação negativa ainda retorna um resultado maior que as demais variações.

Tabela 3: Resultados dos treinamentos considerando kurtosis e entropy

| **Dataset** | **Acurácia** |
| --- | --- |
| withoutProcessing | 0.18 |
| negative | 0.45 |
| alargamentoDeContraste-v2 | 0.39 |
| log-v4 | 0.2 |
| gama-v3 | 0.26 |

O modelo Perceptron Multi-Camadas (MLP) foi utilizado para a classificação de imagens de grãos de arroz com base em atributos de primeira ordem. No entanto, os resultados obtidos apresentaram uma acurácia relativamente baixa, variando entre 15% e 24%. Isso sugere que o modelo teve dificuldades em realizar a classificação dos grãos de arroz de forma precisa. Uma possível explicação para esse resultado pode ser a incompatibilidade com o modelo de classificação. Uma vez que, em trabalhos relacionados, modelos como a Rede Neural Convolucional (CNN) têm apresentado uma precisão significativamente maior.

A relação entre os atributos pode ter afetado os resultados obtidos e novos métodos de pré-processamento podem ser considerados para melhorar a precisão. Testes foram realizados para utilizar os atributos que têm uma melhor relação entre si para analisar os dados. A relação entre energia e entropia mostrou resultados satisfatórios com uma precisão de 74% com as imagens sem processamento. Em contraste, a relação entre média e qualquer outro atributo não foi eficaz. Isso pode ser visto nos gráficos 1 e 3.

Desse modo, a classificação deste conjunto de imagens com aprendizado de máquina pode ser um desafio, é importante considerar diversos fatores que podem influenciar nos resultados obtidos, como a relação dos atributos. No entanto, mesmo diante das dificuldades encontradas neste primeiro trabalho, ainda há espaço para otimizar os resultados e obter uma maior precisão na classificação dos grãos de arroz.

**3. Filtragem espacial**

Nesta fase do estudo, foram conduzidas implementações de filtragem espacial, tanto linear quanto não-linear. Na primeira parte, que se refere à filtragem espacial linear, foi aplicada uma função de convolução. Essa função tem como finalidade aplicar um filtro de convolução em uma imagem, recebendo como parâmetros a imagem a ser processada, o filtro a ser aplicado e a técnica de tratamento a ser utilizada na borda da imagem. O filtro é uma matriz com dimensões MxN, onde M e N são ímpares, maiores ou iguais a 3 e menores que o tamanho da imagem em cada dimensão. O parâmetro da borda pode assumir 4 valores possíveis: "ignore", "espelho", "zero" ou "replicar".

O processo tem início com a verificação da validade do tamanho do filtro, ou seja, se este é ímpar e menor que a imagem. Em caso negativo, é gerado um erro. Em seguida, é feita uma cópia da imagem para que a original não seja alterada. Então, é definida a técnica de tratamento da borda, que consiste em adicionar colunas e linhas na borda da imagem conforme o parâmetro escolhido. Por fim, o filtro é aplicado na imagem, percorrendo-a linha por linha e coluna por coluna. Para cada posição, é obtida uma submatriz da imagem que será processada pelo filtro, e é aplicado o filtro nessa submatriz. O resultado é atribuído à posição correspondente na imagem processada.

Na filtragem espacial não-linear, foram utilizados quatro filtros: mediana, média, máximo e mínimo. Posteriormente, foi implementada uma máscara de nitidez. Para tal, é necessário criar uma matriz do kernel de máscara com valores -1, -1, -1, -1, 9 e -1, -1, -1, -1. Em seguida, essa máscara é aplicada na imagem utilizando o algoritmo de convolução mencionado anteriormente. Isso realça as bordas e detalhes na imagem, dando a impressão de nitidez. Para a implementação da filtragem high-boost, é necessário escolher um valor para k (0 < k < 2). Em seguida, cria-se uma matriz do kernel de máscara com valores -1, -1, -1, -1, 1+k e -1, -1, -1, -1. Essa máscara é então aplicada na imagem utilizando o algoritmo de convolução mencionado anteriormente. Isso realça as bordas e detalhes na imagem, com maior intensidade do que a máscara de nitidez, devido ao valor de k.

**4. Segmentação**

Nesta atividade foram explorados métodos de segmentação de imagem, onde o primeiro método implementado foi o algoritmo K-means, que foi aplicado em uma base de dados fornecida na disciplina para fins de teste. A função K-means utilizada neste trabalho retornou a imagem segmentada juntamente com os centróides de cada cluster. É importante mencionar que os centróides podem variar em cada execução da função, uma vez que é utilizado um valor aleatório como semente. Por essa razão, a lógica do algoritmo retorna os índices dos pixels mais próximos do centróide de menor e maior distância, permitindo assim que as labels intermediárias sejam removidas.

Além do método K-means, foram aplicadas duas técnicas de segmentação por binarização, global (Otsu) e local (Niblack), bem como o algoritmo de agrupamento de máxima verossimilhança Expectation-Maximization (EM). Uma vez obtidos os resultados das segmentações nas imagens de ultrassom, foram buscadas métricas em artigos com foco em segmentação similar ao deste estudo. Dessa forma, as métricas utilizadas foram: Índice de Jaccard (IoU), Coeficiente de Similaridade de Dice (DSC), Acurácia, Recall, Precisão e F1.

Vale ressaltar que a maioria das métricas utilizadas considera a posição dos pixels, entretanto, a maioria das imagens da base de dados possui fundo preto, gerando um desbalanceamento que pode influenciar nos resultados, levando a conclusões tendenciosas. Por essa razão, as métricas que não levam em consideração a posição dos pixels, como o IoU, apresentam resultados mais realistas e confiáveis. Deve-se enfatizar que a análise cuidadosa das métricas é essencial para avaliar a eficácia da segmentação em imagens médicas, como as de ultrassom.

**5. Descrição de imagens**

Para esta etapa foram utilizados alguns dos métodos apresentados anteriormente nas seções. O foco deste trabalho foi desenvolver um sistema completo para a classificação dos grãos de arroz, com ênfase na utilização dos descritores de imagens. A descrição de imagens é um processo essencial na área de processamento de imagens e visão computacional, onde é possível representar uma região de duas formas: por suas características externas, ou seja, sua fronteira, e por suas características internas, que são os pixels que constituem a região. As propriedades de objetos em imagens podem ser medidas, tais como forma, cor e textura, e estas medidas são conhecidas como características da imagem.

As características da imagem são normalmente agrupadas em um vetor de escalares, denominado descritor de imagem, onde cada objeto é representado por um ponto em um espaço RN para n características. É importante que um descritor seja invariante a transformações afins, ou seja, que ele seja capaz de identificar a mesma imagem independentemente de sua posição, escala ou orientação. Essa invariância é fundamental para a robustez e eficiência dos algoritmos de descrição de imagens. Entre os descritores apresentados na disciplina estão: Gray-Level Co-Occurrence Matrix (GLCM), Speeded Up Robust Features (SURF), Gray-Level Run-Length Matrix (GLRLM), Histograms of Oriented Gradients (HOG) e Local Binary Pattern (LBP).

**5.1 Resultados dos descritores**

Para a elaboração do sistema de visão computacional, definiu-se a necessidade de empregar diversos métodos, incluindo um método de pré-processamento, um método de segmentação e dois métodos de descrição, com exceção do descritor CNN, além de dois classificadores (utilizando o método K-fold) e a apresentação de pelo menos três métricas, que devem englobar o uso de todos os métodos citados anteriormente. A escolha de cada método é determinada pelos autores do trabalho.

Para dar início ao desenvolvimento do sistema de visão computacional, optou-se por empregar o método de pré-processamento negativo, uma vez que este mostrou resultados expressivos na análise da base de grãos de arroz utilizada. Posteriormente, foi empregado o método de segmentação Otsu. A escolha desse método decorreu de sua utilização na disciplina, o que favoreceu sua aplicação nesta atividade. Por fim, foram extraídas as características (features) utilizando os classificadores Random Forest (RF) e MLP, empregando-se o método K-Fold para a validação dos resultados. As métricas utilizadas para avaliar o desempenho do sistema foram a acurácia, kappa e F1, considerando os descritores GLCM e HOG. A média e o desvio padrão dos classificadores com os descritores também foram obtidas para uma melhor compreensão. A Tabela 4 e 5 apresenta os resultados das métricas utilizando os classificadores MLP e RF com o descritor GLCM.

Tabela 4: Métricas com RF e GLCM

|  | Accuracy | Kappa | f1 |
| --- | --- | --- | --- |
| 0 | 0.680 | 0.602139 | 0.610667 |
| 1 | 0.705 | 0.628335 | 0.645028 |
| 2 | 0.775 | 0.717496 | 0.752109 |
| 3 | 0.695 | 0.618082 | 0.629920 |
| 4 | 0.695 | 0.620399 | 0.612370 |

Os resultados apresentados na tabela indicam que a média das métricas obtidas com o uso do classificador RF em conjunto com o descritor GLCM foi de 0.71 para a acurácia, 0.63 para o kappa e 0.65 para o f1. Em relação ao desvio padrão, observou-se um valor de 0.037 para a acurácia, 0.045 para o kappa e 0.058 para o f1. Os resultados referente a Tabela 5 para a média foram de 0.70 para a média, 0.62 para o Kappa e 0.66 para o f1. Se tratando do desvio padrão, para a acurácia o resultado foi de 0.067, para o Kappa foi de 0.083 e de 0.088 para o f1.

Tabela 5: Métricas com MLP e GLCM

|  | Accuracy | Kappa | f1 |
| --- | --- | --- | --- |
| 0 | 0.705 | 0.630742 | 0.623125 |
| 1 | 0.625 | 0.528909 | 0.588864 |
| 2 | 0.810 | 0.761329 | 0.802756 |
| 3 | 0.690 | 0.612015 | 0.688825 |
| 4 | 0.685 | 0.609955 | 0.600222 |

A Tabela 6 e 7 mostra os resultados das métricas obtidas através da avaliação do classificador MLP com o descritor HOG. Na Tabela 6 a média alcançada para o classificador MLP com o descritor HOG foi de 0.72 para acurácia, 0.64 para Kappa e de 0.68 para o f1. Já para o desvio padrão os resultados foram de 0.03 para a acurácia, 0.04 para Kappa e de 0.03 para f1.

Tabela 6: Métricas com MLP e HOG.

|  | Accuracy | Kappa | f1 |
| --- | --- | --- | --- |
| 0 | 0.74 | 0.672565 | 0.673289 |
| 1 | 0.74 | 0.670459 | 0.687375 |
| 2 | 0.69 | 0.608968 | 0.675536 |
| 3 | 0.75 | 0.687314 | 0.749247 |
| 4 | 0.68 | 0.597800 | 0.643892 |

A Tabela 7 mostra os resultados utilizando o classificador RF juntamente com o descritor HOG foi alcançada a média de 0.71 para a acurácia, 0.636 para o Kappa e de 0.638. O desvio padrão para essas métricas foi de 0.029 para a acurácia, de 0.034 para o Kappa e de 0.03 para o f1.

Tabela 7: Métricas com RF e HOG.

|  | Accuracy | Kappa | f1 |
| --- | --- | --- | --- |
| 0 | 0.715 | 0.640662 | 0.646035 |
| 1 | 0.755 | 0.690295 | 0.687649 |
| 2 | 0.675 | 0.596674 | 0.595431 |
| 3 | 0.705 | 0.631595 | 0.632255 |
| 4 | 0.700 | 0.624248 | 0.631209 |

Ao analisarmos os valores obtidos por meio das tabelas acima, bem como a média e o desvio padrão, podemos inferir que o sistema desenvolvido apresentou bons resultados na classificação dos grãos de arroz. Os valores médios das métricas obtidas foram consideravelmente elevados, enquanto os valores do desvio padrão foram baixos. Esse conjunto de evidências sugere que os classificadores foram capazes de identificar de forma eficiente os grãos de arroz, e que o conjunto de imagens utilizado para treinamento e teste foi homogêneo e representativo.

**6. Conclusão**

Neste estudo, um sistema de visão computacional foi desenvolvido para classificar grãos de arroz usando diferentes métodos, incluindo pré-processamento, segmentação, extração de características e classificação. Os resultados mostraram que o sistema teve um bom desempenho na identificação de grãos de arroz, evidenciado pelos altos valores médios e baixo desvio padrão das métricas de avaliação. O uso dos classificadores Random Forest e MLP, juntamente com os descritores GLCM e HOG, mostrou-se eficaz na obtenção de resultados precisos. No geral, este estudo destaca o potencial para explorar a utilização de outros descritores e métodos de classificação, almejando resultados ainda melhores.