



FERNANDO RIBEIRO CASSIANO

**EXPOENTE DIRECIONAL DE HURST NA ANÁLISE DE
SIMILARIDADE DE IMAGENS DE SEMENTES**

LAVRAS – MG

2018

FERNANDO RIBEIRO CASSIANO

**EXPOENTE DIRECIONAL DE HURST NA ANÁLISE DE SIMILARIDADE DE
IMAGENS DE SEMENTES**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof. Dr^a. Thelma Sáfyadi
Orientadora

**LAVRAS – MG
2018**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Cassiano, Fernando Ribeiro.

Expoente direcional de Hurst na análise de similaridade de
imagens de sementes / Fernando Ribeiro Cassiano. - 2018.
50 p. : il.

Orientador(a): Thelma Sáfadi.

.

Tese (doutorado) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Autossimilaridade. 2. Classificação de imagens. 3.
Ondaletas. I. Sáfadi, Thelma. . II. Título.

FERNANDO RIBEIRO CASSIANO

**EXPOENTE DIRECIONAL DE HURST NA ANÁLISE DE SIMILARIDADE DE
IMAGENS DE SEMENTES**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 13 de junho de 2018.

Prof. Dr ^a . Angélica Sousa da Mata	UFLA
Prof. Dr ^a . Eniuce Menezes de Souza	UEM
Prof. Dr. Paulo Henrique Sales Guimarães	UFLA
Prof. Dr. Tales Jesus Fernandes	UFLA

Prof. Dr^a. Thelma Sáfadi
Orientadora

**LAVRAS – MG
2018**

AGRADECIMENTOS

É bem verdade que nunca pensei em escrever estes agradecimentos. Aquele garoto simples, de cidadezinha, nem nos seus melhores sonhos havia algo assim. Um doutorado. Então, já que isso me foi proporcionado...

Quero agradecer a Deus por eu pertencer à família que pertence, que desde sempre foi a minha referência. Pelo suporte recebido dos mais próximos. Meu avô José Ribeiro, minha avó Aurea, minha tia Nívia, meu primo Ricardo, minhas irmãs Raimara e Nayara foram e são os melhores!

Agradeço a Deus pela vida da tia Delci, que foi quem me deu a oportunidade de iniciar essa caminhada acadêmica.

Sou grato, Deus, pelos colegas que conheci; pelos amigos que encontrei, que estiveram e estarão presentes sempre. Pelo Diego e pelos amigos da república, pelo companheirismo, convivência e conversas viajadas. (risos)

Agradeço pela vida do Cristian, do Helvécio e da Adriana que, nas minhas idas a Lavras, me abrigaram em suas casas nesses últimos dois anos e meio. Foi muito bom estar com vocês.

Deus, muito obrigado pela vida da Thelma. Uma pessoa maravilhosa. Sua paciência, presteza, dedicação e tantas outras qualidades ficarão gravadas na minha memória. Sua orientação foi inspiradora. Muito obrigado.

Pelos demais professores, por ter aprendido bastante com eles.

Pelos funcionários da UFLA, o suporte deles foi fundamental. Em especial pela Josinha, pela Jose, pela Nádia, pelo Adoniram e o Ronaldo, muito obrigado.

Por fim, quero agradecer pelos idealizadores e mantenedores de agências como a CAPES e a FAPEMIG que, no tempo em que precisei, me apoiaram financeiramente.

*“Por que Dele, por Ele
e para Ele são todas as coisas.
A Ele seja a glória para sempre. Amém.”
(Romanos 11:36)*

RESUMO

A modernização está presente em todos os campos do conhecimento. Técnicas cada vez mais sofisticadas e aparelhos mais modernos têm surgido com frequência. A decomposição em ondaletas é uma ferramenta que tem fundamental importância em muitos desses avanços. No que se refere à análise de imagens, essa ferramenta tem cooperado para criação de diversas novas técnicas, seja para reconstrução, compressão, eliminação de ruído, dentre outros. Outra ferramenta que auxilia na análise de imagens é o expoente de Hurst, que mede o quão autossimilar uma imagem é, de forma que se capte informação sobre características da imagem que a olho nu seria impossível. Com isso, o objetivo deste trabalho será combinar a técnica de decomposição em ondaletas com o cálculo do expoente de Hurst para analisar imagens de sementes e assim poder classificá-las em cheias, levemente danificadas ou danificadas. No cálculo do expoente de Hurst serão usadas como medida de localização a média e a mediana. Um support vector machine será usado para validação do método proposto. Para o grupo de todas as sementes a acurácia média do método, utilizando a média, foi de 74,5% e, com a mediana foi de 57,05%. Utilizando o grupo de sementes cheias e danificadas a taxa média de acurácia, com a média como medida de posição, foi de 99,76% e, com a mediana foi de 80,93%. No grupo contendo sementes levemente danificadas e danificadas a taxa média de acurácia, usando a média como medida de posição, foi de 99,26% e, com a mediana foi de 76,22%.

Palavras-chave: Autossimilaridade; Classificação de imagens; Ondaletas.

ABSTRACT

Modernization is present in all fields of knowledge. Increasingly sophisticated techniques and more modern devices have come up frequently. Wavelet decomposition is a tool that has fundamental importance in many of these advances. With regard to image analysis, this tool has cooperated to create several new techniques, like the ones for reconstruction, compression, noise elimination, among others. Another tool that assists in the analysis of images is the Hurst exponent, which measures how self-similar an image is, so that information about the characteristics of the image is captured, which would be impossible with the naked eye. Therefore, the objective of this work will be to combine the technique of wavelet decomposition with the calculation of the Hurst exponent to analyze seed images and thus be able to classify them as full, slightly damaged or damaged. In the calculation of the Hurst exponent it will be used as location measure the mean and the median. A support vector machine will be used for validation of the proposed method. For the group of all seeds the mean accuracy of the method, using the mean, was 74.5 % and with the median was 57.05 %. Using the group of full and damaged seeds the mean accuracy rate, with the mean as a measure of position, was 99.76 % and with the median was 80.93 %. In the group containing slightly damaged and damaged seeds the mean accuracy rate, using the mean as a measure of position, was 99.26 % and with the median was 76.22 %.

Keywords: Self-similarity; Image classification; wavelets.

LISTA DE FIGURAS

Figura 2.1 – Representação gráfica da ondaleta de Haar	16
Figura 2.2 – Representação das ondaletas de Daubechies, db2 e db3.	17
Figura 2.3 – Algoritmo multiresolução de Mallat, ou transformada decimada	20
Figura 2.4 – Diagrama de fluxo de uma transformada não decimada.	21
Figura 2.5 – Relação entre $\log(R/\sigma)$ e $\log(N)$ para sete grupos de fenômenos.	23
Figura 2.6 – Exemplos de imagens digitais. (a) paisagem natural, (b) esboço gerado sinteticamente, (c) pôster, (d) captura de tela, (e) ilustração em preto e branco, (f) código de barras, (g) impressão digital, (h) raio X, (i) imagem obtida por microscópio	25
Figura 2.7 – Processo de aquisição de uma imagem digital.	26
Figura 2.8 – Representação computacional de uma imagem.	27
Figura 2.9 – Exemplo de sementes de girassol: a) Elementos da semente de girassol, b) Raio X de uma semente de girassol.	28
Figura 2.10 – Decomposição de ondaletas da imagem e obtenção de diferentes inclinações.	30
Figura 2.11 – Mapeamento de um espaço de entrada para um espaço característico em um SVM.	31
Figura 3.1 – Exemplo de chapa de raio X utilizada para classificação das sementes como cheias, com danos leves e danificadas.	32
Figura 3.2 – Imagens de sementes classificadas como: a) semente cheia, b) semente com danos leves e c) semente danificada.	33
Figura 3.3 – Representação de como foram extraídas as partes internas de cada imagem para a decomposição.	33
Figura 3.4 – Representação da obtenção do expoente de Hurst por meio do coeficiente angular de uma regressão linear ajustada no log espectro.	34
Figura 4.1 – Distribuição empírica dos coeficientes de Hurst para as direções horizontal, vertical e diagonal, considerando a média como medida de localização.	37
Figura 4.2 – Boxplot dos coeficientes de Hurst para as direções horizontal, vertical e diagonal, considerando a média como medida de localização.	38
Figura 4.3 – Distribuição dos valores dos expoentes de Hurst na direção horizontal, vertical e diagonal considerando a mediana como medida de localização.	39

Figura 4.4 – Boxplot dos coeficientes de Hurst para as direções horizontal, vertical e diagonal, considerando a mediana como medida de localização.	39
Figura 4.5 – Histograma da taxa de classificação correta para as sementes cheias (c) e danificadas (d), em 1000 simulações, considerando a média como medida de localização.	40
Figura 4.6 – Histograma da taxa de classificação correta para as sementes cheias (c) e danificadas (d), em 1000 simulações, considerando a mediana como medida de localização.	40
Figura 4.7 – Histograma da taxa de classificação correta para as sementes levemente danificadas (dl) e danificadas (d), em 1000 simulações, considerando a média como medida de localização.	42
Figura 4.8 – Histograma da taxa de classificação correta para as sementes levemente danificadas (dl) e danificadas (d), em 1000 simulações, considerando a mediana como medida de localização.	42
Figura 4.9 – Histograma da taxa de classificação correta para as sementes cheias (c) e levemente danificadas (dl), em 1000 simulações, considerando a média como medida de localização.	43
Figura 4.10 – Histograma da taxa de classificação correta para as sementes cheias (c) e levemente danificadas (dl), em 1000 simulações, considerando a mediana como medida de localização.	43
Figura 4.11 – Histograma da taxa de classificação correta para os três tipos de sementes, em 1000 simulações, considerando a média como medida de localização.	45
Figura 4.12 – Histograma da taxa de classificação correta para os três tipos de sementes, em 1000 simulações, considerando a mediana como medida de localização.	45

LISTA DE TABELAS

Tabela 2.1 – Possíveis valores que os pixels podem assumir.	27
Tabela 4.1 – Confundimento médio em 1000 simulações para sementes cheias (c) e danificadas (d), usando a média e a mediana como medidas de localização.	41
Tabela 4.2 – Confundimento médio em 1000 simulações para sementes danificadas e levemente danificadas, usando a média e a mediana como medidas de localização.	43
Tabela 4.3 – Confundimento médio em 1000 simulações para sementes cheias e levemente danificadas, usando a média e a mediana como medidas de localização.	44
Tabela 4.4 – Taxa média de acurácia para cada grupo de semente e metodologia	44
Tabela 4.5 – Confundimento médio para todas as sementes em 1000 simulações, considerando a média e a mediana como medidas de localização.	46
Tabela 4.6 – Taxa média de acurácia para cada grupo de semente e metodologia	46

SUMÁRIO

1	INTRODUÇÃO	10
2	REFERENCIAL TEÓRICO	13
2.1	Campos brownianos fractais	13
2.2	Fundamentos de Ondaletas	14
2.2.1	Análise multirresolução de Mallat	19
2.2.2	Transformada não decimada	20
2.3	Expoente de Hurst (H)	22
2.4	Análise de imagens	25
2.4.1	Análise de sementes	27
2.5	Support Vector Machines	30
3	MATERIAL E MÉTODOS	32
4	RESULTADOS E DISCUSSÕES	36
5	CONCLUSÕES	47
5.1	Trabalhos futuros	47
	REFERÊNCIAS	48

1 INTRODUÇÃO

A era tecnologia caminha a passos largos para uma modernização jamais vista. A cada dia a sofisticação e diversidade dos aparelhos eletrônicos nos levam a meditar sobre os limites da imaginação humana. A comunidade científica, como parte ativa nesse processo de modernização, tem buscado a criação e aprimoramento de técnicas que permitam o avanço do conhecimento humano nas mais diversas áreas do conhecimento.

No Brasil, uma das principais atividades econômicas é a agropecuária e, como tal, é uma área em constante pesquisa e avanço tecnológico. Técnicas apuradas são buscadas para que se potencialize a capacidade de produção nas diversas atividades deste ramo de atuação. Uma atividade bastante importante dentro da agropecuária é a produção de grãos. Dentro dessa produção destaca-se aqui o cultivo do girassol (*Helianthus annuus* L.). Por causa da sua maior resistência à seca, à pragas e doenças, além de outros fatores, o girassol tem se expandido e consolidado como cultura técnica e economicamente viável. Dentre as culturas anuais o girassol é responsável por 16% da produção de óleo comestível do mundo, está na terceira posição na produção mundial de óleo comestível (EMBRAPA, 2018). Dessa forma, se faz necessária a criação e desenvolvimento de técnicas que auxiliem na melhoria da qualidade de tais grãos.

Uma forma de melhorar a qualidade dos grãos a serem produzidos é através da seleção de sementes saudáveis para o plantio. Assim, aumenta-se a chance de ter uma plantação saudável e produtiva. A melhoria no processo de seleção de tais sementes reduzirá o tempo gasto entre a colheita e a seleção de novos grãos para uma nova plantação, a disponibilização das sementes para comercialização e conseqüentemente a qualidade do produto final para o consumidor. Provoca-se, assim, celeridade no processo de produção. Dentre as diversas técnicas utilizadas para esse fim, a decomposição em ondaletas e os algoritmos de aprendizagem de máquina têm ganhado destaque.

A utilização da análise de ondaletas tem provocado uma enorme evolução do conhecimento científico nos últimos anos. Essa técnica permite decompor uma função, $f(t)$, em uma combinação linear de outras funções, chamadas ondaletas e função escala. Essa decomposição permite analisar $f(t)$ em vários níveis de detalhes.

A transformada discreta de ondaletas é uma técnica que proporcionou um grande avanço em análise de imagens e processamento de sinais. Tem sido uma técnica alternativa à análise de Fourier por seu caráter flexível, preservando os fenômenos locais, não periódicos e de multiescala (RODRIGUES, 2015).

Na literatura há vários exemplos da utilização da transformada discreta de ondaletas para a análise de imagens. Em muitos desses trabalhos o expoente de Hurst é utilizado como ferramenta chave para a tomada de decisão. Há várias maneiras de se calcular tal expoente, uma delas é usando o espectro de energia da transformada. Nesse trabalho o espectro foi calculado de duas formas. Na primeira delas utilizou-se o logaritmo da média dos quadrados dos coeficientes de detalhes, obtidos por meio da transformada discreta de ondaletas (NICOLIS; RAMIREZ-COBO; VIDA KOVIC, 2011). O outro método foi usar a mediana do logaritmo do quadrado dos coeficientes de detalhes (KANG; VIDA KOVIC, 2017). Em ambas maneiras usou-se medidas centrais na distribuição dos logaritmos dos coeficientes. A mediana foi uma alternativa mais robusta a possíveis *outliers* que poderiam surgir após a transformação logarítmica dos quadrados dos coeficientes. O expoente pode ser usado como uma medida de autossimilaridade ao longo das escalas da decomposição da imagem. Essa decomposição foi sucedida por métodos que envolvem algoritmos de aprendizagem de máquina, que auxiliaram na validação dos métodos propostos.

São várias as famílias de algoritmos de aprendizagem de máquina. Fernández-Delgado et al. (2014), por exemplo, estudaram 179 classificadores de 17 famílias com o intuito de compará-los. Esses mesmos autores usaram 121 bancos de dados para testar os classificadores. Os classificadores com melhor desempenho foram da família Random Forest (RF) seguidos pela família Support Vector Machine (SVM) com núcleo gaussiano. Importante ressaltar que a diferença entre a acurácia dos classificadores não foi estatisticamente significativa (FERNÁNDEZ-DELGADO et al., 2014). Nesse trabalho o SVM foi a família utilizada para a validação do método proposto. Na literatura essas técnicas aparecem nas mais diversas áreas do conhecimento. Dentre muitas outras aplicações, Leite, Sáfadi e Carvalho (2013) e Sáfadi et al. (2016) usaram algoritmos de aprendizagem para classificar imagens de sementes de girassol, Nicolis, Ramirez-Cobo e Vidakovic (2011) utilizaram alguns desses algoritmos na classificação de tumores cancerígenos, etc.

O objetivo, neste trabalho, foi utilizar a transformada discreta não decimada de ondaletas juntamente com o expoente direcional de Hurst para extrair características de imagens de sementes, feitas com raio X, e classificá-las, com o auxílio de um algoritmo de aprendizagem de máquina, como cheias, levemente danificadas ou danificadas. Para o cálculo do expoente de Hurst foram usadas duas estatísticas como medidas de localização: a média e a mediana. A comparação da acurácia proporcionada pelo emprego dessas duas medidas foi feita neste

trabalho. Por ser uma metodologia não invasiva, os procedimentos que são propostos evitarão a destruição de amostras de sementes para a quantificação da qualidade de semeadura dos lotes além de aumento no leque de opções de análise. Espera-se, com isso, melhorar a forma e diminuir o tempo necessário de classificação e seleção das sementes.

O trabalho está organizado como segue. No Capítulo 2 apresenta-se o conteúdo teórico da tese, conceitua-se os principais assuntos tratados neste trabalho. No Capítulo 3 apresenta-se o material e os procedimentos usados para realizar este trabalho. No Capítulo 4 discute-se os resultados encontrados. Por fim, no Capítulo 5 apresenta-se as conclusões obtidas com o presente trabalho e ideias para trabalhos futuros.

2 REFERENCIAL TEÓRICO

Ao longo deste capítulo são apresentados conceitos necessários para o desenvolvimento deste trabalho. A definição do conceito de movimento browniano fractal é apresentada na Seção 2.1. Na Seção 2.2 estão os conceitos fundamentais sobre a teoria de ondaletas. Na Seção 2.3 tem-se os fundamentos da teoria sobre o expoente de Hurst. A Seção 2.4 traz a apresentação dos elementos que constituem uma imagem, do ponto de vista computacional, e alguns métodos de classificação já utilizados. Por fim, na Seção 2.5, são apresentadas as principais ideias sobre a família Suport Vector Machine.

2.1 Campos brownianos fractais

O movimento browniano fractal (mbf) é um processo estocástico gaussiano não estacionário com média zero. É um dos mais populares modelos para fenômenos autossimilares. Representado por $\{B_H(t), t \in \mathbb{R}\}$, em que H é um parâmetro de autossimilaridade, conhecido também como expoente de Hurst, no caso unidimensional o processo browniano fractal é caracterizado pela seguinte função de autocorrelação

$$R_{B_H} = E [B_H(t) B_H(s)] = \frac{\sigma_H^2}{2} [|t|^{2H} + |s|^{2H} - |t-s|^{2H}],$$

em que $t, s \in \mathbb{R}$, $\sigma_H^2 = \Gamma(1-2H) \frac{\cos(\pi H)}{\pi H}$, Γ representa a distribuição Gama e $0 < H < 1$. Além disso, esse processo estocástico é dito autossimilar, ou seja, para todo $a > 0$ ele satisfaz $B_H(at) \stackrel{d}{=} a^H B_H(t)$ onde $\stackrel{d}{=}$ indica igualdade em distribuição (NICOLIS; RAMIREZ-COBO; VIDAKOVIC, 2011).

A generalização do mbf para dimensões maiores não é única. Uma possível generalização (NICOLIS; RAMIREZ-COBO; VIDAKOVIC, 2011) para \mathbb{R}^2 é o campo browniano fractal (cbf). Tal campo é um campo aleatório gaussiano, $B_H(\mathbf{u})$, onde \mathbf{u} é a posição em algum domínio específico, com média zero. A função de autocorrelação do campo $B_H(\mathbf{u})$ é dada por

$$R_{B_H}(\mathbf{u}, \mathbf{v}) = E [B_H(\mathbf{u}) B_H(\mathbf{v})] = \frac{\sigma_H^2}{2} [\|\mathbf{u}\|^{2H} + \|\mathbf{v}\|^{2H} - \|\mathbf{u} - \mathbf{v}\|^{2H}],$$

em que $\mathbf{u}, \mathbf{v} \in \mathbb{R}^2$ e $\|\cdot\|$ é a norma euclidiana usual e

$$\sigma_H^2 = \frac{2^{-(1+2H)} \Gamma(1-H)}{\pi H \Gamma(1+H)}.$$

Nesse cenário descrito acima serão utilizadas as técnicas de ondaletas.

2.2 Fundamentos de Ondaletas

As funções ondaletas têm se mostrado uma ferramenta muito versátil para os diversos campos do conhecimento. O primeiro trabalho utilizando tal ferramenta foi o feito por Alfréd Haar em 1910. Porém, o desenvolvimento e disseminação aconteceram a partir dos trabalhos de Meyer, em 1989, e Daubechies (1992). Lima (2002) afirma que os trabalhos de Daubechies formam o alicerce para as aplicações atuais de ondaletas. Essa técnica tem auxiliado diversos pesquisadores a avançarem o conhecimento científico em todo o mundo. Seu caráter interdisciplinar faz com que as ondaletas ganhem cada vez mais visibilidade. Vários trabalhos podem ser encontrados na literatura com uso dessa ferramenta.

Delouille, Jansen e Sachs (2006) propuseram uma técnica utilizando ondaletas para eliminação de ruído para dados bidimensionais. Pimentel e Silva (2011) utilizaram técnicas de ondaletas para fazer uma análise de volatilidade e correlação de séries de índices financeiros. Ramirez-Cobo e Vidakovic (2013) utilizam técnicas de ondaletas para mapear e extrair características de imagens de mamogramas, permitindo a identificação, ou não, da presença de câncer de mama. Remenyi et al. (2014) propuseram uma técnica para eliminação de ruído em imagens bidimensionais utilizando a teoria de ondaletas combinada com métodos bayesianos. Ramirez-Cobo et al. (2011) extraíram medidas de autossimilaridade em mapas de precipitação ao longo do tempo. Antonini et al. (1992), Shapiro (1993), Chang, Yu e Vetterli (2000) exemplificaram como a teoria de ondaletas se mostra como uma potente ferramenta na compressão de imagens. Dentre outras aplicações.

Por causa de suas propriedades únicas, ondaletas foram usadas em análise funcional matemática, em estudos de propriedades (multi)fractais, singularidades e oscilações locais de funções, em soluções de equações diferenciais, em reconhecimento de padrões, em compressão de imagens e de sons, em processamento de geometria digital, na solução de vários problemas de física, biologia, medicina, astronomia, acústica, engenharia nuclear, grupo de renormalização em mecânica estatística, problemas de computação gráfica, neurofisiologia, música, ressonância magnética, identificação de vozes, ótica, fractais, turbulência, previsão de terremotos, radar, visão humana. (LIMA, 2002)

Em um sentido amplo, um espaço funcional é simplesmente uma coleção de funções que satisfazem certas propriedades matemáticas. Por exemplo, o espaço de energia finita $L^2(-\infty, \infty)$

é um conjunto de funções que são quadrado integrável (GOSWAMI; CHAN, 2011), que é

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < \infty.$$

As ondaletas são definidas neste espaço. Outra propriedade que as ondaletas compartilham é

$$\int_{-\infty}^{\infty} \psi(t) dt = 0.$$

Uma função que obedece estas características é dita ondaleta mãe. A partir desta ondaleta pode-se gerar ondaletas por operações de dilatação e translação. Então, para $a, b \in \mathbb{R}$, $a \neq 0$,

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right). \quad (2.1)$$

Para o caso discreto tem-se $j, k \in \mathbb{Z}$ e pode-se formar

$$\psi_{j,k}(t) = 2^{j/2} \psi(2^j t - k). \quad (2.2)$$

Nas Equações 2.1 e 2.2 representam-se as ondaletas que são geradas pela ondaleta mãe no caso contínuo e discreto, respectivamente. Na equação 2.1, b é um parâmetro de localização e a um parâmetro de escala. Da mesma forma, na equação 2.2, k é um parâmetro de localização e j um parâmetro de escala. Esta parametrização de tempo ou localização no espaço, k , e a frequência ou escala, j , é muito prática no sentido de que é fácil identificar o local e em que escala há maiores alterações nos coeficientes (BURRUS; GOPINATH; GUO, 1998).

Segundo Oliveira (2007) e Daubechies (1992) as ondaletas são funções matemáticas capazes de separar os sinais em suas diferentes componentes e extrair cada componente com uma resolução correspondente à sua escala. A transformada de ondaletas analisa o sinal em diferentes escalas e se desloca analisando cada ponto do sinal.

Lima (2002) afirma que as ondaletas podem ser vistas como ferramentas para a decomposição ou quebra de sinais nas suas partes constituintes, permitindo fazer uma análise dos dados em diferentes domínios de frequências com a resolução de cada componente atrelada à sua escala. Para o autor, as ondaletas fornecem uma técnica elegante na representação dos níveis de detalhes presentes, de modo que com essa técnica é possível contemplar “a floresta e as árvores”. Ela permite dar um “zoom” nos dados.

Para Morettin (2014) as ondaletas podem ser funções simétricas ou não, suaves ou não, e podem ser representadas por meio de expressões matemáticas simples ou não. Para o mesmo autor, as ondaletas mais utilizadas são:

a) Haar:

Primeira ondaleta mencionada na literatura, introduzida pelo matemático Haar em 1910 e forneceu a primeira representação em uma série ortogonal de ondaletas. A ondaleta de Haar possui suporte compacto e forma analítica, mas não é contínua. A seguir são apresentadas as expressões da ondaleta de Haar para os casos contínuo e discreto, respectivamente:

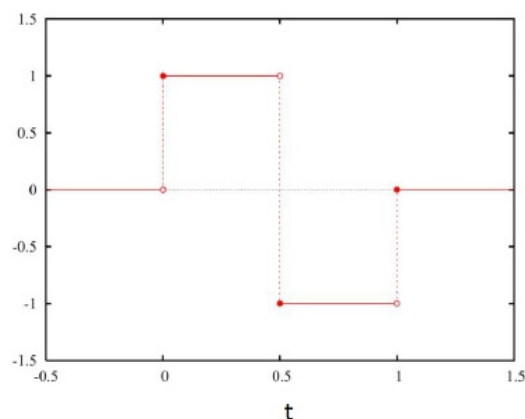
$$\psi(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{caso contrário} \end{cases}$$

e

$$\psi_{j,k}(t) = \begin{cases} 2^{j/2}, & 2^{-j}k \leq t < 2^{-j}(k+1/2) \\ -2^{j/2}, & 2^{-j}(k+1/2) \leq t < 2^{-j}(k+1) \\ 0, & \text{caso contrário.} \end{cases}$$

Na Figura 2.1 apresenta-se o gráfico da ondaleta de Haar.

Figura 2.1 – Representação gráfica da ondaleta de Haar

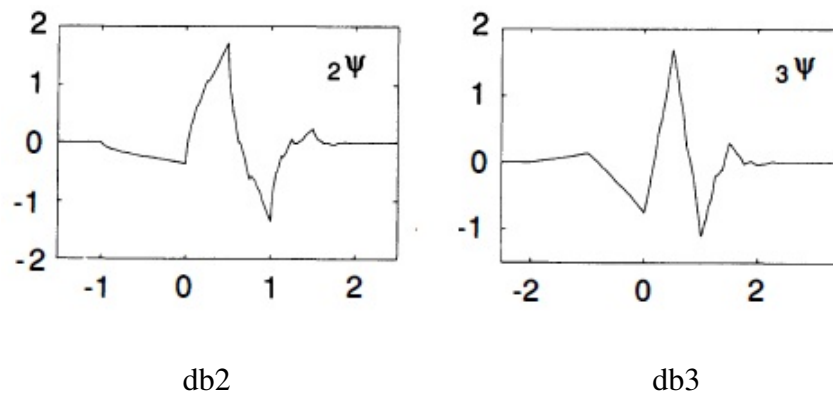


Fonte: Daubechies (1992)

b) Daubechies:

As ondaletas de Daubechies são calculadas de forma iterativa e não possuem forma explícita. Tais ondaletas geram bases que são ortogonais e têm suporte compacto. Elas são classificadas de acordo com o número de momentos nulos (db1, db2, db3, etc). A db1 corresponde à ondaleta de Haar. Na Figura 2.2 apresenta-se a db2 e db3.

Figura 2.2 – Representação das ondaletas de Daubechies, db2 e db3.



Fonte: Daubechies (1992)

Como na análise de Fourier, a ideia é escrever uma função $f(t)$ como combinação linear de outras funções, ondaletas, que é

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \varphi_k(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t),$$

em que $c_k = \int f(t) \varphi_k(t) dt$ e $d_{j,k} = \int f(t) \psi_{j,k}(t) dt$ são chamados coeficientes de ondaletas. Mais especificamente, c_k é conhecido como coeficiente de escala e $d_{j,k}$ são os coeficientes de detalhes (NASON, 2008). Nessas equações, $\varphi(t)$ é dita função escala e está associada à função ondaleta $\psi(t)$.

A expansão anterior, em termos discretos é dada por

$$f(t) = \sum_{k=-\infty}^{\infty} c_{j_0,k} \varphi_{j_0,k}(t) + \sum_{j=0}^{\infty} \sum_k d_{j,k} \psi_{j,k}(t). \quad (2.3)$$

Esta expansão, 2.3, em ondaletas é conhecida como transformada discreta de ondaletas da função $f(t)$ (BURRUS; GOPINATH; GUO, 1998). Na Equação 2.3, j_0 é o nível de resolução mais baixo ou a escala mais grosseira.

Nason (2008) descreve algumas propriedades importantes das ondaletas, sendo elas:

- esparsidade na representação de um grande número de funções, incluindo aquelas com descontinuidades;
- a habilidade de dar “zoom” para analisar funções em um número de escalas e também manipular informações nessas escalas;
- habilidade de detectar e representar características locais e também criar características localizadas em síntese;
- eficiência em termos de rapidez computacional e armazenamento.

“As propriedades individuais, e a combinação delas, são razões porque as ondaletas são úteis para um grande número de problemas estatísticos” (NASON, 2008).

As funções ondaletas definidas em \mathbb{R} podem ser generalizadas para espaços multidimensionais. No presente trabalho o foco será sobre imagens. Portanto, o espaço bidimensional será o espaço em foco. Ramirez-Cobo e Vidakovic (2013) apresentaram uma transformada de ondaletas para o caso bidimensional. Segundo os autores as bases ondaletas bidimensionais são construídas por translações e dilatações de um produto de ondaleta univariada e função escala. São definidas por

$$\varphi(t_1, t_2) = \varphi(t_1) \varphi(t_2),$$

$$\psi^h(t_1, t_2) = \varphi(t_1) \psi(t_2),$$

$$\psi^v(t_1, t_2) = \psi(t_1) \varphi(t_2),$$

$$\psi^d(t_1, t_2) = \psi(t_1) \psi(t_2),$$

em que h, v e d simbolizam as direções horizontal, vertical e diagonal, respectivamente. Então, considerando os elementos das bases de ondaletas, tem-se

$$\varphi_{j,\mathbf{k}}(\mathbf{t}) = 2^{2j} \varphi(2^j t_1 - k_1, 2^j t_2 - k_2),$$

$$\psi_{j,\mathbf{k}}^i(\mathbf{t}) = 2^{2j} \psi^i(2^j t_1 - k_1, 2^j t_2 - k_2),$$

para $i = h, v, d$, em que $\mathbf{t} = (t_1, t_2) \in \mathbb{R}^2$ e $\mathbf{k} = (k_1, k_2) \in \mathbb{Z}^2$. Então uma imagem X (pensada como uma função em $L^2(\mathbb{R}^2)$) pode ser representada como

$$X(\mathbf{t}) = \sum_{\mathbf{k}} c_{j_0, \mathbf{k}} \varphi_{j_0, \mathbf{k}}(\mathbf{t}) + \sum_{j \geq j_0} \sum_{\mathbf{k}} \sum_i d_{j, \mathbf{k}} \psi_{j, \mathbf{k}}^i(\mathbf{t}). \quad (2.4)$$

Os coeficientes dessa decomposição serão analisados com o objetivo de se encontrar características relevantes para a classificação das imagens em estudo.

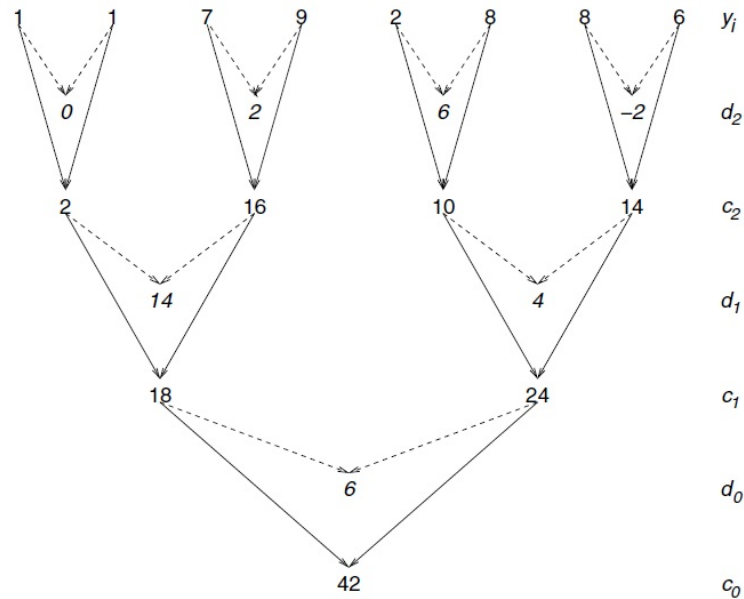
Para Oliveira (2007) as transformadas de ondaletas quase nunca são tratáveis analiticamente e invariavelmente os cálculos são numéricos. Apesar disso, os resultados são extremamente práticos e de rica interpretação.

2.2.1 Análise multirresolução de Mallat

Proposta por Stephane G. Mallat em 1989, esta metodologia faz uso de funções, chamadas filtros. Há dois tipos de filtros: passa-baixo (H), e passa-alto (G). Tais funções são chamadas filtros espelhos em quadratura. Neste trabalho as funções que serão os filtros passa-baixo e passa-alto serão a função escala e a ondaleta, respectivamente. O objetivo de tal metodologia é aproximar uma função previamente conhecida utilizando tais filtros. Mallat (1989) propõe a aplicação dos dois filtros a um vetor \mathbf{A} de tamanho $n = 2^J$, $J \in \mathbb{N}$, gerando dois outros vetores, \mathbf{B} e \mathbf{C} com metade do tamanho de \mathbf{A} , por unidade de tempo, método conhecido como transformada discreta decimada. Os componentes de \mathbf{B} e \mathbf{C} serão ditos de baixa e alta frequência, respectivamente. Na Figura 2.3 ilustra-se o método descrito acima aplicado ao vetor $\mathbf{y} = [1, 1, 7, 9, 2, 8, 8, 6]$.

Na Figura 2.3 as linhas pontilhadas representam uma subtração entre os elementos e a linha cheia uma soma. Os d 's são ditos coeficientes detalhe e os c 's resultantes da transformada são ditos coeficientes de escala. A cada novo passo, ou nova escala, tem-se novos coeficientes detalhe e de escala. Cada vetor resultante, de coeficientes, tem a metade do tamanho do vetor em que foi aplicada a transformada. Observa-se (FIGURA 2.3) que as somas e diferenças são tomadas sempre com os elementos das posições pares no primeiro fator da operação. Tal fato é denominado por Nason (2008) como decimação par, D_0 . Nota-se no procedimento utilizado na Figura 2.3 que alguns elementos do vetor \mathbf{Y} não estão associados através de uma adição ou subtração, como o y_2 e y_3 por exemplo. Segundo Nason (2008) este fato pode levar a perda de alguma informação.

Figura 2.3 – Algoritmo multiresolução de Mallat, ou transformada decimada



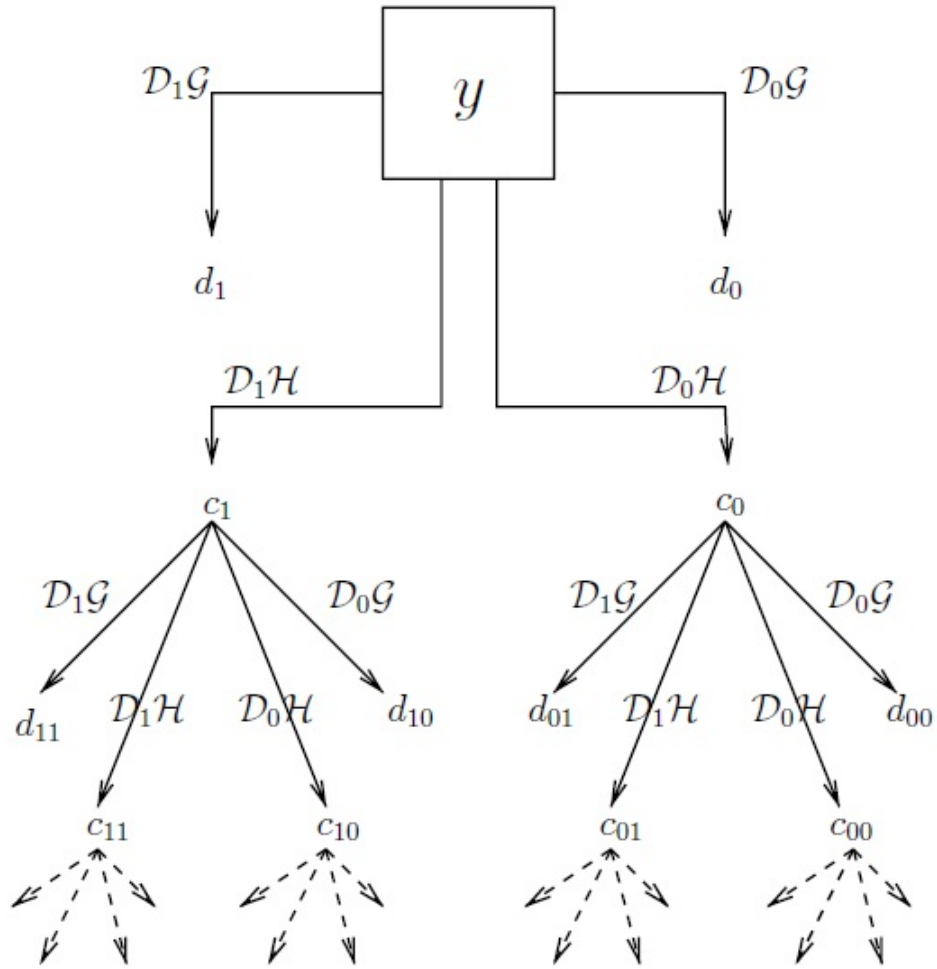
Fonte: Nason (2008)

2.2.2 Transformada não decimada

Na transformada não decimada, além da decimação par, D_0 , há também a decimação ímpar, D_1 , que é a aplicação dos mesmos filtros ao vetor original com uma rotação. Nesse caso, os elementos do vetor original serão deslocados uma posição para a esquerda. O primeiro elemento será deslocado para a última posição do vetor. A Figura 2.4 apresenta uma ilustração do procedimento executado numa transformada não decimada.

Uma característica importante dessa transformada é que o número de elementos do vetor resultante, com os coeficientes de escala e detalhe, da transformação é o mesmo que no vetor original. Kang e Vidakovic (2017) ressaltaram algumas vantagens e desvantagens ao utilizar a transformada discreta não decimada para a estimação do espectro de ondaleta e o expoente de Hurst (ambos conteúdos serão discutidos posteriormente neste texto). Segundo os autores a utilização da transformada discreta não decimada provoca um aumento na autocorrelação dos coeficientes de ondaletas e no viés dos estimadores de Hurst, quando comparada aos estimadores que utilizam a transformada ortogonal tradicional. Em contrapartida, a variância do estimador é menor, o tamanho do sinal ou imagem de entrada pode ser arbitrário, dentre outras vantagens.

Figura 2.4 – Diagrama de fluxo de uma transformada não decimada.



Fonte: Nason (2008)

Considerando um campo browniano fractal, os coeficientes de ondaletas são dados por

$$d_{j,\mathbf{k}}^i = 2^j \int B_H(\mathbf{u}) \psi^i(2^j \mathbf{u} - \mathbf{k}) d\mathbf{u} \quad (2.5)$$

em que o domínio da integral é o \mathbb{R}^2 . Esses coeficientes são variáveis aleatórias com média zero e variância

$$E \left[\left| d_{j,\mathbf{k}}^i \right|^2 \right] = 2^{2j} \int \int \psi^i(2^j \mathbf{u} - \mathbf{k}) \psi^i(2^j \mathbf{v} - \mathbf{k}) E[B_H(\mathbf{u}) B_H(\mathbf{v})] d\mathbf{u} d\mathbf{v}, \quad (2.6)$$

(NICOLIS; RAMIREZ-COBO; VIDA KOVIC, 2011). De (2.6) pode-se derivar

$$E \left[\left| d_{j,\mathbf{k}}^i \right|^2 \right] = \frac{\sigma_H^2}{2} V_{\psi^i} 2^{-(2H+2)j}, \quad (2.7)$$

onde $V_{\psi^i} = - \int \int \psi^i(\mathbf{p} + \mathbf{q}) \psi^i(\mathbf{q}) |\mathbf{p}|^{2H} d\mathbf{p}d\mathbf{q}$. Aplicando o logaritmo em ambos os lados da expressão (2.7) chega-se à seguinte equação

$$\log_2 E \left[\left| d_{j,\mathbf{k}}^i \right|^2 \right] = -(2H + 2)j + C_i, \quad (2.8)$$

onde $C_i = \frac{\sigma_H^2}{2} V_{\psi^i}(H)$, (NICOLIS; RAMIREZ-COBO; VIDAKOVIC, 2011). A equação (2.8) será usada para calcular o expoente de Hurst.

2.3 Expoente de Hurst (H)

Originalmente o expoente de Hurst, criado em 1951 por Harold Edwin Hurst, surgiu da observação e necessidade de se estudar séries temporais de dados de precipitação, temperatura, nível de rios, dentre outros. Seu objetivo era estabelecer uma relação entre o fluxo de um rio e a quantidade de água a ser armazenada de maneira que o reservatório não transbordasse nem esvaziasse, garantindo assim disponibilidade de água para as irrigações ou outras atividades. Seu estudo permite classificar séries temporais entre aleatórias e não aleatórias.

Considere que a descarga de água de um rio no reservatório, em certo ano t , é dada por $X(t)$. A quantidade média de descarga de água que deve ser liberada do reservatório, num período de N anos, para que o seu volume permaneça inalterado é

$$\bar{X}_N = \frac{1}{N} \sum_{t=1}^N X(t).$$

Assim, esse valor médio será o volume a ser utilizado por ano. Considere ainda uma variável, $Z(t, N)$, que será a soma acumulada, até o ano t , da diferença entre a descarga no reservatório e o volume médio utilizado, ou seja,

$$Z(t, N) = \sum_{i=1}^t (X(i) - \bar{X}_N).$$

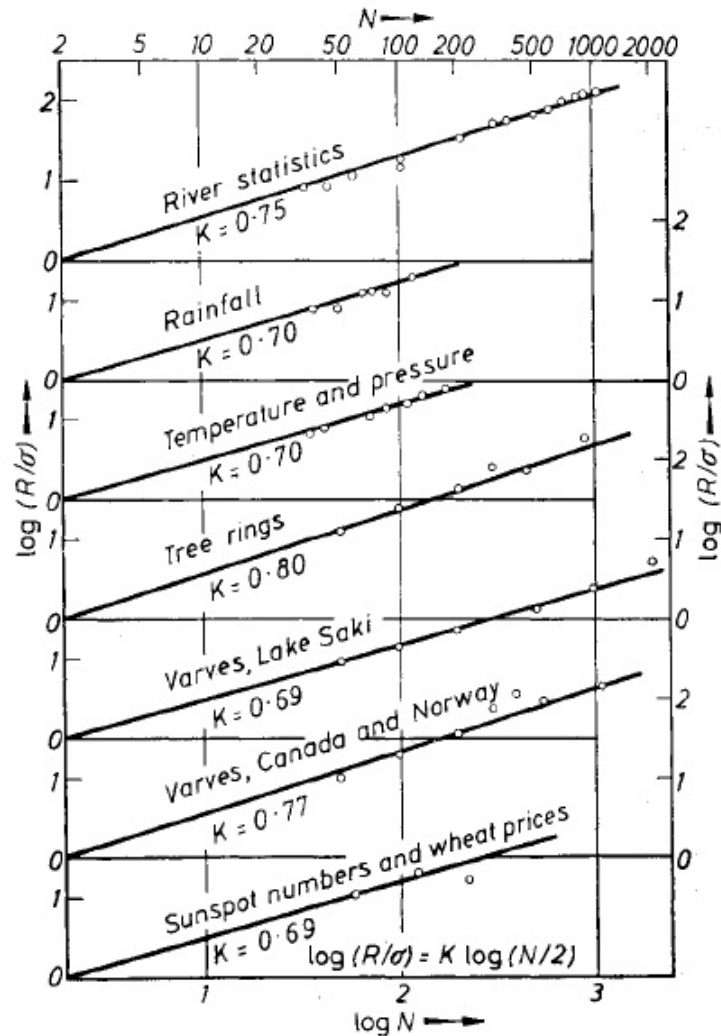
A diferença entre o valor máximo e mínimo de $Z(t, N)$ é o volume de água que deveria ser armazenada para que a descarga média do período se mantivesse. Esta é dada por

$$R(t) = \max\{Z(t, N)\} - \min\{Z(t, N)\},$$

com $1 \leq t \leq N$.

Finalmente, σ é o desvio padrão dos R 's obtidos no período de tempo N , considerado. Calcula-se, então, a estatística R/σ e estabelece-se uma relação com N . Mais especificamente, o autor mostra que há uma relação linear entre $\log(R/\sigma)$ e $\log(N)$ e o expoente de Hurst é o coeficiente angular da função. Na Figura 2.5 apresenta-se as relações existentes, para as variáveis mencionadas anteriormente, para vários fenômenos.

Figura 2.5 – Relação entre $\log(R/\sigma)$ e $\log(N)$ para sete grupos de fenômenos.



Fonte: Hurst (1956)

O expoente de Hurst é uma medida que indica a complexidade e dinâmica da estrutura de correlação de qualquer série temporal. Além disso, em uma decomposição de ondaletas ele conecta os coeficientes nos diferentes níveis de decomposição. De modo específico, há uma conexão entre o comportamento das magnitudes dos coeficientes de ondaletas e o expoente de singularidade (RAMIREZ-COBO; VIDA KOVIC, 2013). Tal expoente tem valores entre zero e

um, $0 < H < 1$, e é uma medida chave para fazer análise da duração de dependência de longo alcance (LYASHENKO et al., 2016).

Há várias interpretações para o expoente de Hurst. Quando seu valor é maior que 0,5 indica que o processo tem um comportamento de tendência estável. Para valores menores que 0,5 o processo é anti persistente. Para $H = 0,5$ os desvios do processo não dependem de valores prévios, são aleatórios. Outra possível interpretação, por exemplo, em virologia: $H < 0,5$ indica que o vírus está localmente confinado, $H \approx 0,5$ indica que o vírus comporta-se de acordo com um movimento browniano (aleatório) e $H > 0,5$ indica um movimento direcionado (LYASHENKO et al., 2016).

Assim como no caso unidimensional, para duas ou mais dimensões o expoente de Hurst continua sendo uma ferramenta muito importante na análise de similaridade. Segundo Nicolis, Ramirez-Cobo e Vidakovic (2011) o expoente de Hurst é a medida padrão de escala regular e pode estar conectada a medidas de dimensão, fractalidade, memória longa em sinais e imagens. Empiricamente, para o caso bidimensional, o expoente de Hurst pode ser calculado com o auxílio do coeficiente angular da regressão definida sobre os pares

$$\left(j, \log_2 \overline{|d_{j,\mathbf{k}}^i|^2} \right), i = h, v, d, \quad (2.9)$$

em que $\overline{|d_{j,\mathbf{k}}^i|^2}$ é a média dos quadrados dos módulos dos coeficientes detalhes na escala j . Nicolis, Ramirez-Cobo e Vidakovic (2011) mencionam que a média aritmética em (2.9) pode ser substituída por outras medidas de localização como a mediana, por exemplo. Considerando α o coeficiente angular da regressão linear nos pares em (2.9), a relação entre α e o expoente de Hurst é dada por:

$$H = -\frac{(\alpha + 2)}{2}. \quad (2.10)$$

Jeon, Nicolis e Vidakovic (2014) afirmaram que por estar estimando o expoente de Hurst empiricamente e quando há correlação dos resíduos, o espectro decai lentamente (com $\alpha > -2$) e pode gerar H negativo.

Kang e Vidakovic (2017) propuseram o uso da mediana como alternativa à média numa tentativa de reduzir a correlação. Os autores relatam que a mediana é mais robusta a *outliers* que possam ocorrer na amostra, devido à transformada logarítmica dos quadrados dos coeficientes.

Para o método proposto, os novos pares ordenados, como em (2.9), serão

$$\left(j, \text{mediana} \left(\log_2 \left| d_{j,k}^i \right|^2 \right) \right), i = h, v, d. \quad (2.11)$$

Usando os pares em 2.11, o expoente de Hurst será dado por

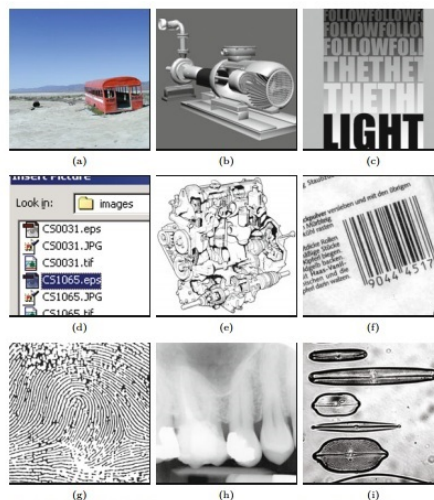
$$H = -\frac{\alpha}{2 \log 2} - 1, \quad (2.12)$$

em que α é a inclinação da reta de regressão sobre os pares em 2.11 estimada por mínimos quadrados ordinários. O conjunto de pontos obtidos, para os diferentes valores de j , nas ordenadas de 2.9 e 2.11 são chamados espectros de energia, $S(j)$.

2.4 Análise de imagens

Com o avanço tecnológico, a cada dia são propostas ferramentas mais potentes e sofisticadas, no que se refere à classificação de imagens. Algoritmos muito eficientes têm sido propostos para a melhoria da extração de características presentes nas imagens. Consequentemente, tem-se melhores condições de análise de tais imagens e, dessa forma, estabelecer conclusões mais confiáveis sobre algum interesse específico. Na Figura 2.6 apresentam-se diferentes imagens digitais.

Figura 2.6 – Exemplos de imagens digitais. (a) paisagem natural, (b) esboço gerado sinteticamente, (c) pôster, (d) captura de tela, (e) ilustração em preto e branco, (f) código de barras, (g) impressão digital, (h) raio X, (i) imagem obtida por microscópio

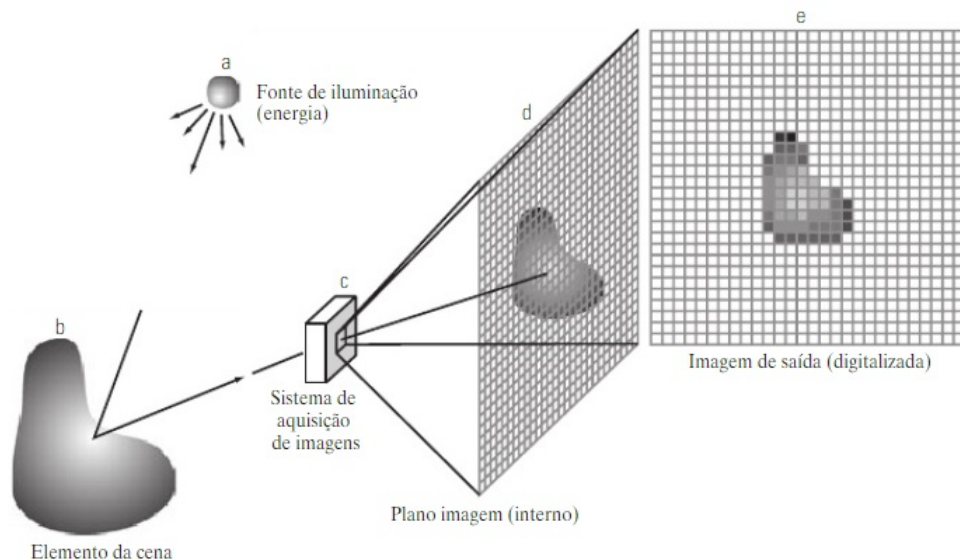


Segundo Burguer e Burgue (2016) são necessários três passos para se conseguir uma imagem digital:

- 1- a distribuição contínua da luz deve ser amostrada espacialmente;
- 2- esta função resultante deve então ser amostrada no tempo para criar uma imagem simples;
- 3- finalmente, os valores resultantes devem ser quantizados num conjunto finito de inteiros (ou valores de pontos flutuantes) de modo que ele possa ser representado por números digitais.

Visualmente, tem-se, genericamente falando, a situação apresentada na Figura 2.7.

Figura 2.7 – Processo de aquisição de uma imagem digital.

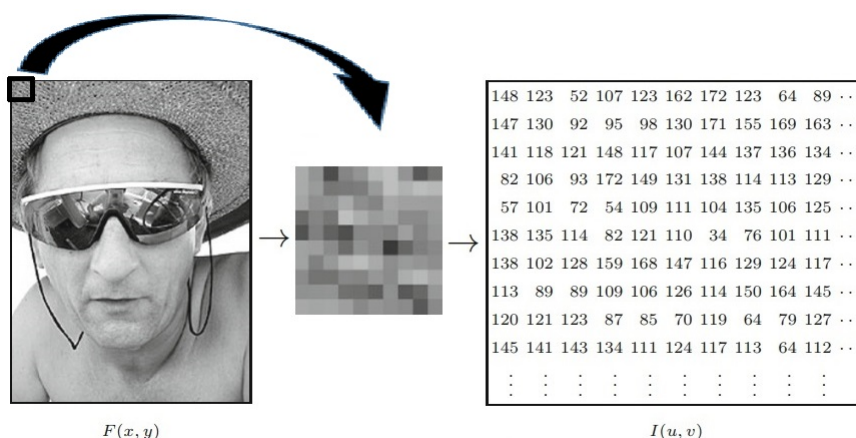


Fonte: Gonzalez e Woods, 2009

Após a finalização destes três passos tem-se uma imagem na forma de uma matriz bidimensional ordenada de números inteiros (FIGURA 2.8).

Cada elemento da matriz formada é chamado de pixel. Na Tabela 2.1 apresenta-se os possíveis valores que cada pixel pode assumir e também os usos mais comuns de cada tipo de imagem. Dessa forma, é necessário que se crie métodos efetivos para extração de informação contida em tais matrizes, possibilitando o conhecimento de padrões, ou características, que a olho nu são imperceptíveis.

Figura 2.8 – Representação computacional de uma imagem.



Fonte: Burguer e Burgue (2016)

Tabela 2.1 – Possíveis valores que os pixels podem assumir.

Imagens em escalas de cinza			
Canais	Bits/pixel	Valores	Uso
1	1	[0,1]	Imagens binárias: documentos, ilustrações, fax
1	8	[0,255]	Universal: fotos, escans, printes
1	12	[0,4095]	Alta qualidade: fotos, escans, printes
1	14	[0,16383]	Profissional: fotos, escaner's, printes
1	16	[0,65535]	Mais alta qualidade: medicina, astronomia
Imagens coloridas			
Canais	Bits/pixel	Valores	Uso
3	24	[0, 255] ³	RGB, Universal: fotos, escans, printes
3	46	[0, 4095] ³	RGB, Alta qualidade: fotos, escans, printes
3	42	[0, 16383] ³	RGB, Profissional: fotos, escaner's, printes
4	32	[0, 255] ⁴	CMYK, pré-impressão digital

Fonte: Burguer e Burgue (2016)

2.4.1 Análise de sementes

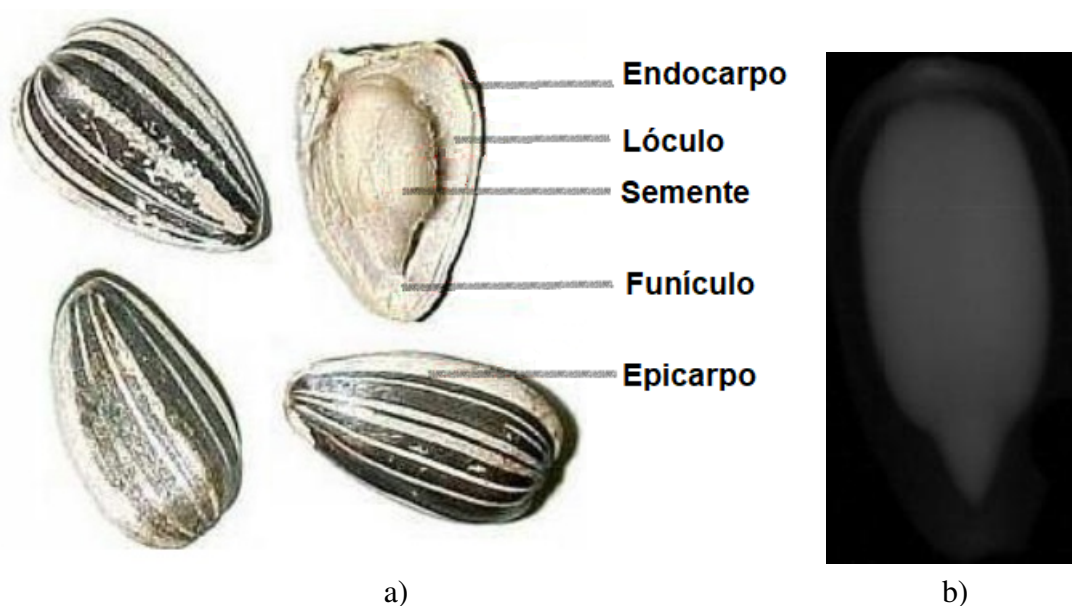
Cada planta exige um cuidado específico, tanto pela estrutura morfológica quanto pelos processos de plantio e cultivo. Para que se inicie bem um processo de cultivo é fundamental que se utilize materiais de qualidade, e um desses materiais é a semente. A necessidade de se atender produtores que estão constantemente preocupados em ter colheitas produtivas e de qualidade tem gerado o crescimento no comércio de sementes. Além disso, sementes mais resistentes e adaptadas aos diversos tipos de solos têm sido exigidas. Diante desse cenário,

torna-se necessário o desenvolvimento de métodos de produção e tratamento nesse setor para a seleção dos melhores espécimes.

Realizar uma análise de sementes é, a partir da verificação de parâmetros genéticos, físicos, fisiológicos e sanitários, determinar a qualidade de semeadura das sementes. Na análise em larga escala, normalmente toma-se uma amostra aleatória dentre alguns lotes, examina-se tal amostra e, qualifica-se todo o lote. Esse processo é parte essencial para se ter produções cada vez maiores e com mais qualidade. Há órgãos exclusivamente dedicados a definir as normas para a análise de sementes. Há, por exemplo, vários documentos emitidos pelo Ministério da Agricultura, Pecuária e Abastecimento no sentido de normatizar e orientar as ações na análise de sementes. Uma técnica muito utilizada nessa área é a análise através de imagens de raio X.

Lopes e Nascimento (2009) afirmaram que esse método era muito utilizado na França para determinar o valor cultural de lotes de sementes de hortaliças. Desde então a técnica se popularizou em todo o mundo. Uma característica bastante importante da análise de raio X é que ela é uma análise rápida e permite a avaliação morfológica de estruturas internas das sementes como do embrião, do endosperma sem a necessidade de destruir as sementes examinadas. Na Figura 2.9 apresenta-se um exemplo de uma semente de girassol e um raio X de uma semente da mesma espécie.

Figura 2.9 – Exemplo de sementes de girassol: a) Elementos da semente de girassol, b) Raio X de uma semente de girassol.



Diante desse cenário tecnológico, de busca por melhorias ou desenvolvimento de novas técnicas de análise de sementes, pesquisas que se dediquem a estes temas se tornam muito

relevantes. E é sobre esse assunto que dedica-se este trabalho. Outras pesquisas já foram feitas com a análise de raio X de sementes. Dentre outros, pode-se citar dois, que utilizaram o mesmo banco de dados que será usado neste trabalho.

Leite, Sáfadi e Carvalho (2013) utilizaram análise de componentes independentes juntamente com análise discriminante para fazer a classificação automática das sementes. Cada imagem, \mathbf{x}_i , $i = 1, \dots, n$, é decomposta como uma combinação linear de componentes independentes (imagens base) da seguinte forma

$$\mathbf{x}_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{in}s_n,$$

em que a_{ij} são os coeficientes da combinação linear e s_i componentes independentes. Os autores usaram, ainda, análise de componentes principais para fazer redução, $k < n$, no número de componentes independentes, de modo que cada imagem passou a ser

$$\mathbf{x}_i = a_{i1}s_1 + a_{i2}s_2 + \dots + a_{ik}s_k.$$

Nesse trabalho foi feita uma variação no número de componentes independentes e verificada qual a quantidade de componentes resultaria na melhor classificação. Utilizando uma parte das imagens como treinamento e outra como teste, com diferentes quantidades de componentes, para os diferentes tipos de sementes e diferentes análises, os autores afirmaram que a análise de componentes independentes é uma metodologia satisfatória para a classificação de imagens baseada nas características extraídas delas.

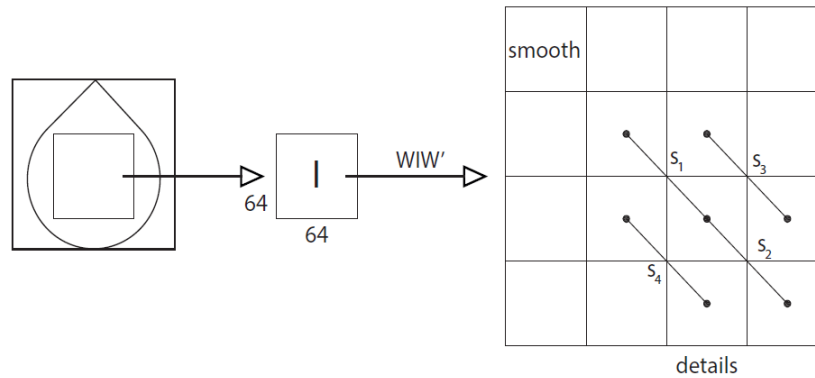
Sáfadi et al. (2016) classificaram as imagens baseados no espectro de ondaletas. Para isso os autores decompueram as imagens utilizando a transformada de ondaleta discreta não decimada de escala mista. O termo *escala mista* vem do fato de que nessa metodologia são combinados elementos de diferentes escalas ou subespaços de multirresolução. Matricialmente, a imagem transformada, \mathbf{B} , foi obtida da seguinte maneira

$$\mathbf{B} = \mathbf{WIW}',$$

em que \mathbf{I} é a imagem original e \mathbf{W} é uma matriz ondaleta. A matriz \mathbf{B} é dita transformada de ondaletas de escala mista e, a partir dela, vários índices foram derivados. A Figura 2.10 apresenta o procedimento descrito. Na Figura 2.10, s_1 e s_2 representam características de escala

mista sobre (ou dentro dos) os subespaços de multirresolução e s_3 e s_4 são medidas ligadas ao fluxo da energia da imagem entre resoluções de diferentes escalas. As quatro medidas foram calculadas para cada uma das imagens.

Figura 2.10 – Decomposição de ondaletas da imagem e obtenção de diferentes inclinações.



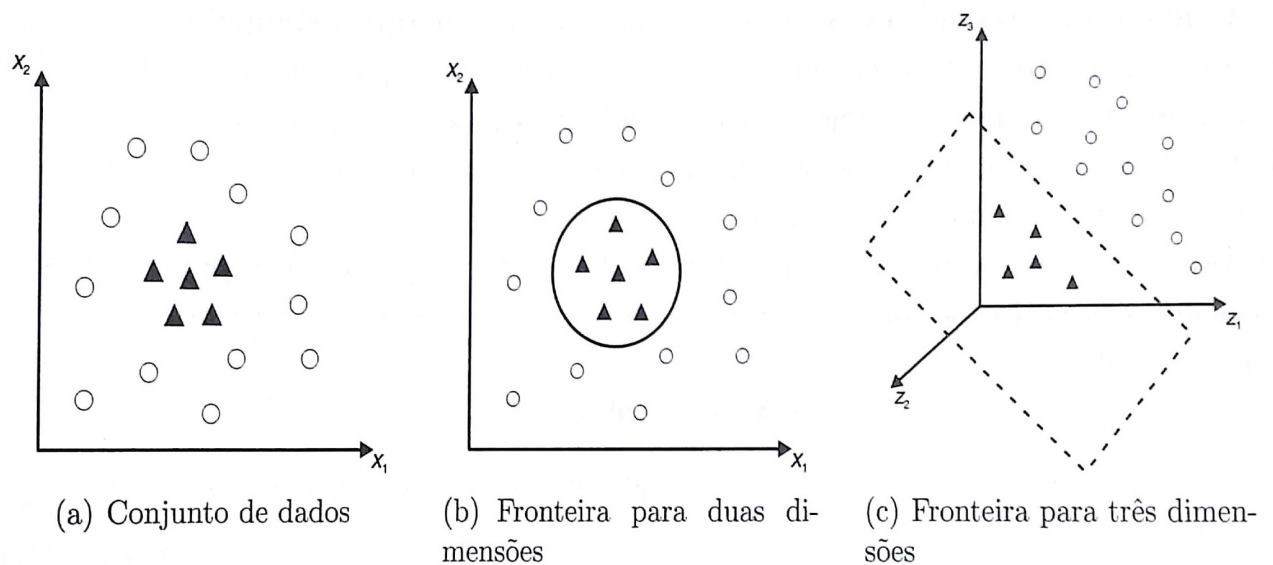
Algoritmos de aprendizagem de máquina foram utilizados para a validação da metodologia. Os índices de uma parte do conjunto de imagens foram usados como treinamento, dentro do algoritmo, e a outra parte ficou para a validação. A metodologia foi utilizada considerando dois cenários: classificadores binários (sementes cheias/ danificadas) e multinomial (cheia/pouco danificadas/danificadas). Nesse trabalho, para os classificadores binários os autores encontraram uma taxa de acerto de 82% e para classificadores multinomiais a taxa média de acerto foi 57%.

2.5 Support Vector Machines

O Suport Vector Machine (SVM) é uma família de algoritmos de aprendizagem de máquina. Aprendizagem de máquina, de maneira geral, é a construção de modelos a partir de uma amostra de dados e a utilização desses modelos para caracterizar novos dados. Há formas diferentes de aprendizado. Aprendizado supervisionado é aquele em que o conjunto de dados para aprendizagem está etiquetado com a resposta correta para cada elemento. O SVM é um algoritmo de aprendizagem de máquina supervisionado que trabalha com a construção de hiperplanos em espaços multidimensionais. A ideia principal dos SVM's é usar uma função núcleo para produzir um hiperplano ótimo de maneira que os dados sejam separados linearmente. As funções núcleo são utilizadas para projetar os dados de um espaço de entrada para um espaço característico (feature space) de maior dimensão, no qual, posteriormente, é realizada a classificação dos dados.

Existem várias funções núcleo dentre as quais pode-se citar as funções gaussianas (mais conhecidas como funções de bases radiais), funções quadráticas e polinomiais (SÁFADI et al., 2016). Há funções núcleo não lineares, que são usadas quando os dados de treinamento não são linearmente separáveis (FACELI et al., 2017). Fernández-Delgado et al. (2014) fizeram a avaliação de 17 famílias de classificadores e a família SVM com núcleo gaussiano obteve o segundo melhor resultado geral, e a diferença entre os dois primeiros colocados foi estatisticamente insignificante. Na Figura 2.11 apresenta-se um esquema do funcionamento dos SVM's com função núcleo não linear. Percebe-se que inicialmente, (a), os dados não poderiam ser separados por nenhum método linear. Com a aplicação de uma transformação não linear para um espaço de maior dimensão os dados ficaram agrupados de tal forma que uma separação linear tornou-se possível.

Figura 2.11 – Mapeamento de um espaço de entrada para um espaço característico em um SVM.



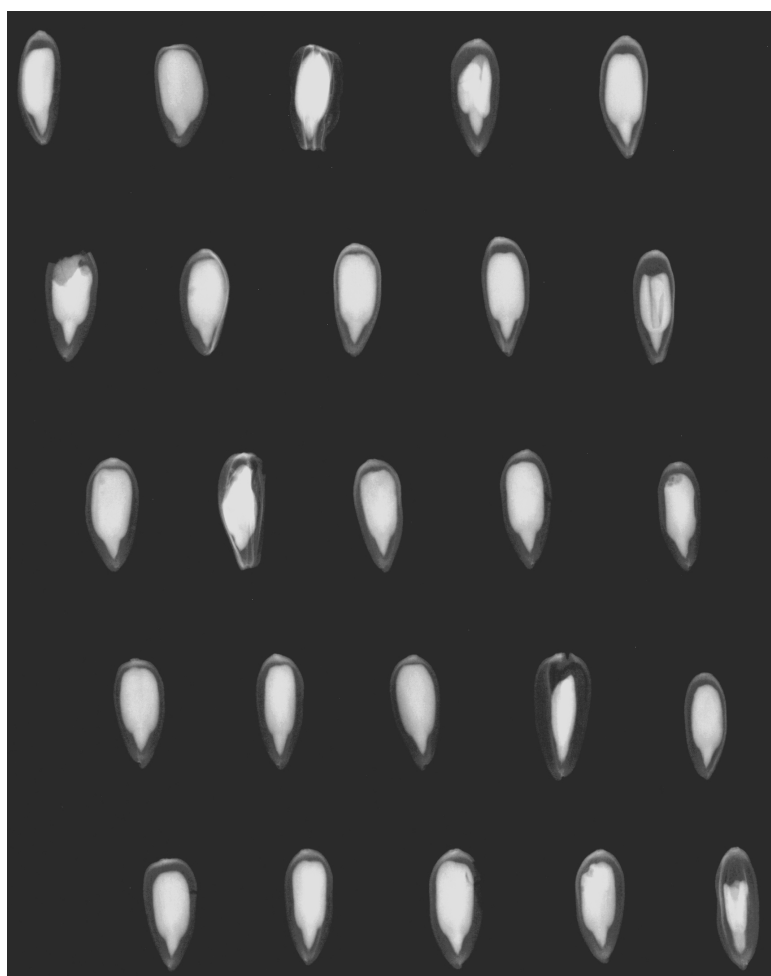
Fonte: Faceli et al. (2017)

A precisão do algoritmo é verificada aplicando o modelo pré-ajustado a um novo conjunto de dados e posteriormente, com a construção de uma matriz de confusão, são aferidas as taxas de acerto. A matriz de confusão é onde se armazenam as classificações corretas e os erros de classificação. Nessa matriz, os números na diagonal principal são as classificações corretas, os demais elementos são erros de classificação.

3 MATERIAL E MÉTODOS

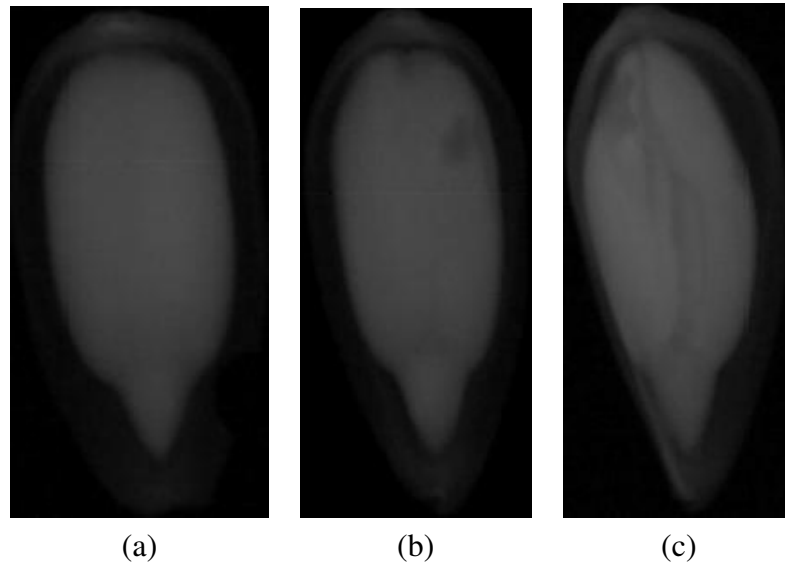
Os dados utilizados neste trabalho são imagens de raio X de sementes de girassol (*Helianthus annuus* L.) obtidas no Laboratório de Análises de Sementes do Departamento de Agricultura, na Universidade Federal de Lavras. A amostra de sementes de girassol, cultivar Hélio-250 foi produzida em Uberlândia, estado de Minas Gerais, colheita 2010/2011. As sementes foram separadas em três tipos: cheias, com danos leves e danificadas. Há um total de 175 imagens de sementes cheias, 130 imagens de sementes levemente danificadas e 140 imagens de sementes danificadas. A classificação foi feita por profissional especializado na área de sementes após análise das imagens de raio X das sementes. Na Figura 3.1 apresenta-se uma chapa de raio X utilizada para análise e classificação.

Figura 3.1 – Exemplo de chapa de raio X utilizada para classificação das sementes como cheias, com danos leves e danificadas.



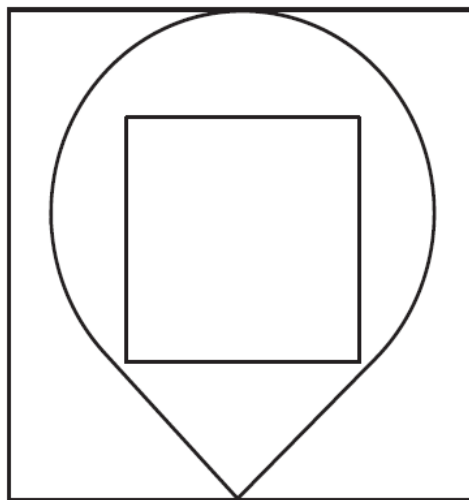
A Figura 3.2 apresenta um exemplo de imagem para cada tipo de semente.

Figura 3.2 – Imagens de sementes classificadas como: a) semente cheia, b) semente com danos leves e c) semente danificada.



Pelo fato de que as imagens têm tamanhos diferentes, optou-se por trabalhar apenas com a parte central de cada uma delas. Isso possibilita que se tenha a “mesma parte” de todas as imagens, a saber, uma matriz de 64x64. Esse corte foi realizado de forma automática, com a construção de uma função no software **R**. Na Figura 3.3 apresenta-se o procedimento descrito.

Figura 3.3 – Representação de como foram extraídas as partes internas de cada imagem para a decomposição.

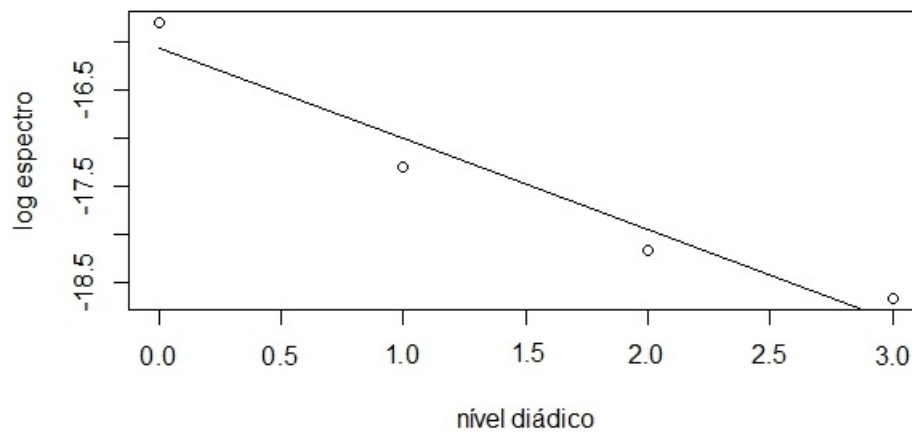


Inicialmente cada um dos três tipos de imagens foram utilizados para a decomposição de ondaletas. Os coeficientes de ondaletas foram obtidos por meio da transformada de ondaletas discreta não decimada, usando a função proposta por Daubechies (1992). Nicolis, Ramirez-Cobo e Vidakovic (2011) estudaram varias famílias de ondaletas, dentre elas a Daubechies, e

concluíram que as diferenças entre os valores paramétricos encontrados não eram significantes, exceto para a base Haar. Escolheu-se trabalhar com a Daubechies com quatro momentos nulos e utilizou-se quatro níveis de decomposição. Posteriormente foi calculado o espectro de ondaleta, $S(j)$, que é definido como a ordenada apresentada em (2.9) e em (2.11). Por meio do método de mínimos quadrados ordinários foi ajustada uma regressão linear sobre os pares ordenados em (2.9).

Na Figura 3.4 apresenta-se um exemplo do ajuste da reta por mínimos quadrados.

Figura 3.4 – Representação da obtenção do expoente de Hurst por meio do coeficiente angular de uma regressão linear ajustada no log espectro.



Então as relações (2.10) e (2.12) foram usadas para o cálculo do expoente de Hurst. O objetivo da decomposição foi verificar características do expoente direcional de Hurst para os diferentes tipos de sementes. Esses expoentes formaram, então, um conjunto de dados para treinamento e validação em um algoritmo de aprendizagem de máquina. Mais especificamente: Suport Vector Machine com função núcleo do tipo *linear*. Outras funções núcleo foram testadas, a gaussiana por exemplo, porém o melhor resultado foi encontrado com o núcleo linear. Tanto para treinamento como para a validação foram utilizadas as decomposições na direção horizontal, vertical e diagonal de cada imagem. As decomposições foram efetuadas usando o pacote *waveslim* (WHITCHER, 2015), do software *R*. Foram considerados nesta análise quatro grupos: os três tipos de sementes; sementes cheias e levemente danificadas; cheias e danificadas e levemente danificadas e danificadas.

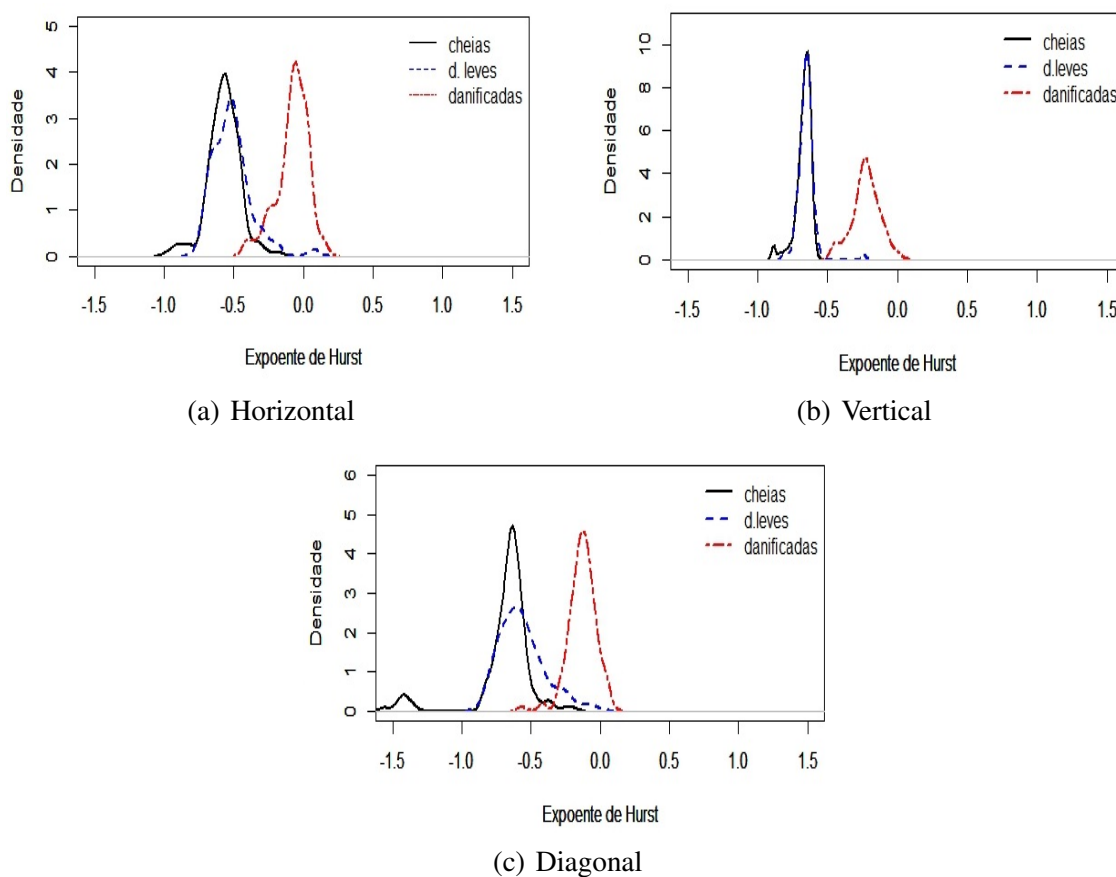
Para cada grupo de sementes foi feita uma seleção aleatória de 75% das sementes para treinamento do algoritmo. Os 25% restantes foram usados para verificação do aprendizado, ou acurácia do algoritmo, por meio da construção de uma matriz de confusão. Esse procedimento foi repetido, em cada grupo de semente, 1000 vezes. Outras proporções para treinamento e validação foram testadas porém a que proporcionou melhor resultado foi 75% e 25%, respectivamente. Os pacotes do software *R* que auxiliaram nessa simulação foram o *caTools* (TUSZYNSKI, 2014), o *e1071* (MEYER et al., 2017) e o *caret* (KUHN, 2017). Desejou-se, com isso, verificar a utilidade do método proposto e comparar a sua acurácia com a de outras metodologias existentes na literatura.

4 RESULTADOS E DISCUSSÕES

Nesta seção apresenta-se os principais resultados obtidos neste trabalho. Inicialmente apresenta-se estatísticas descritivas dos expoentes de Hurst obtidos com a decomposição de on-daletas, para a média e mediana como medidas de localização. Posteriormente apresenta-se os resultados encontrados em mil simulações, com o objetivo de validar o método proposto. Os resultados das simulações serão apresentados por grupo de sementes, a saber: sementes cheias e danificadas, sementes levemente danificadas e danificadas, sementes cheias e levemente danificadas e por fim o grupo de todas as sementes. Para cada grupo de sementes os resultados com a média e a mediana serão apresentados e comparados.

Na Figura 4.1 tem-se as distribuições do coeficiente de Hurst, com a média como medida de localização, nas direções horizontal, vertical e diagonal, respectivamente. Cada um dos casos contém a distribuição para as sementes cheias, levemente danificadas e danificadas. As distribuições das sementes cheias e levementes danificadas se sobrepõem em todos os casos. No entanto, na decomposição diagonal das sementes levementes danificadas há um aumento na variabilidade dos valores dos expoentes, causando um achatamento na curva da distribuição. Há também uma leve mudança para a direita no valor da média dos coeficientes de Hurst para as sementes levementes danificadas. Nas três direções, a distribuição dos expoentes das sementes danificadas é bem diferente das distribuições das demais sementes. Percebe-se que há uma mudança considerável para a direita na média dos coeficientes de Hurst das sementes danificadas. Nota-se, então, que a técnica é eficaz na extração de características em cada tipo de imagem. Ainda nessa figura, verifica-se que a maioria dos valores dos coeficientes de Hurst, para todas as sementes, são negativos. Tal fato pode estar relacionado com o que afirmaram Jeon, Nicolis e Vidakovic (2014), que nesse caso indicam a presença de correlação nos resíduos.

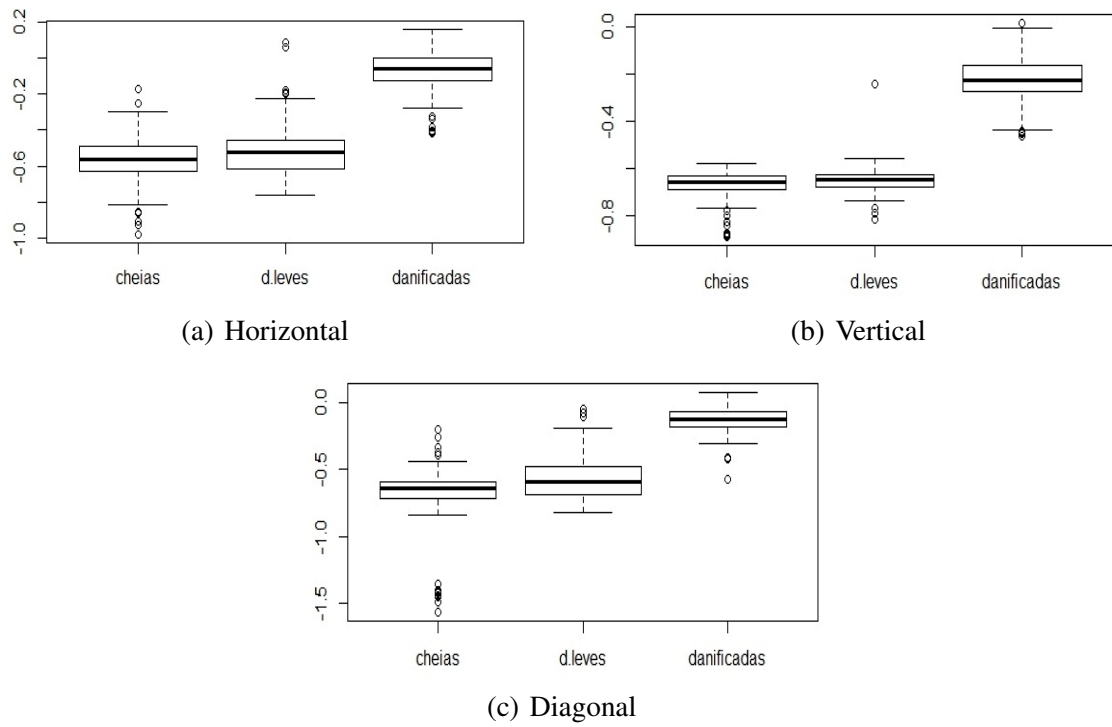
Figura 4.1 – Distribuição empírica dos coeficientes de Hurst para as direções horizontal, vertical e diagonal, considerando a média como medida de localização.



Analisando os box plots (FIGURA 4.2), percebe-se que há outliers nas distribuições de todas as sementes nas três direções. Nota-se que há semelhança entre as distribuições das sementes cheias e levemente danificadas. A distribuição das sementes danificadas no entanto, é muito distinta das demais. As sobreposições das distribuições são causadas, em grande parte dos casos, nos locais onde estão os valores extremos.

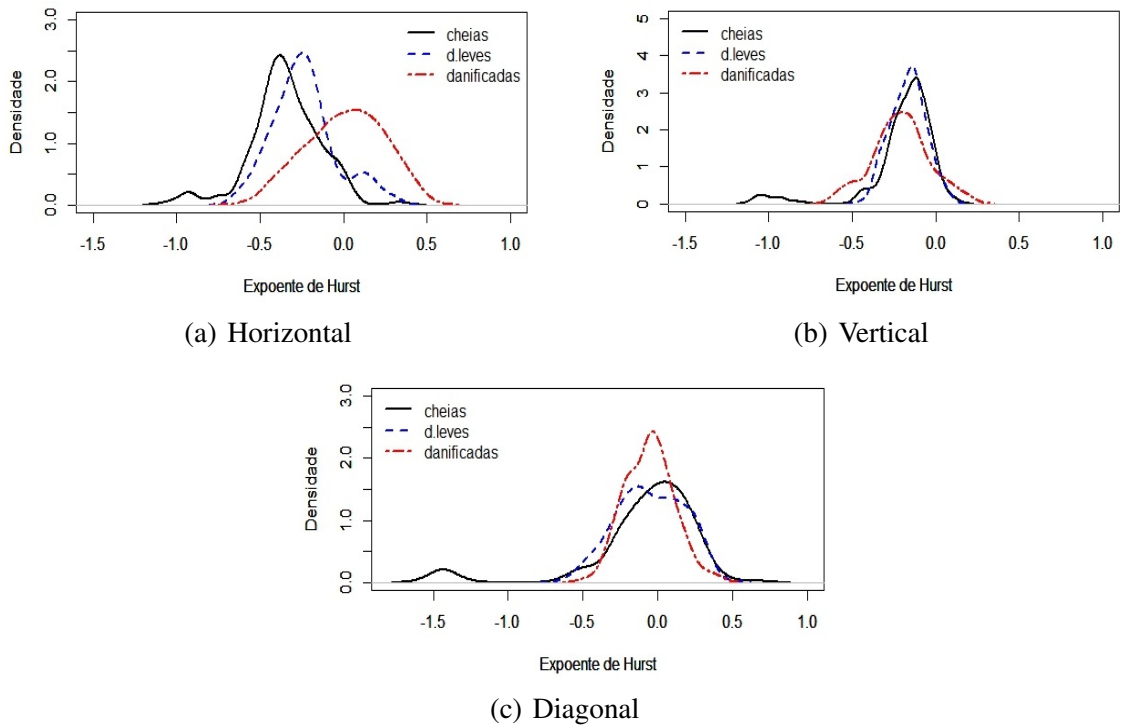
A substituição da média pela mediana provocou mudanças nos resultados, porém menos que o esperado. Na Figura 4.3 são apresentados os gráficos das distribuições amostrais dos coeficientes de Hurst nas três direções consideradas. A substituição provocou uma “sobreposição” nas distribuições direcionais empíricas dos coeficientes para todas as sementes. Na direção horizontal, porém, essa sobreposição é um pouco menor devido a um deslocamento gradual para a direita da média das distribuições das sementes levemente danificadas e danificadas, respectivamente. No geral, houve um deslocamento para a direita na distribuição dos coeficientes para todos os tipos de sementes, o que indica a diminuição na correlação dos resíduos, como

Figura 4.2 – Boxplot dos coeficientes de Hurst para as direções horizontal, vertical e diagonal, considerando a média como medida de localização.



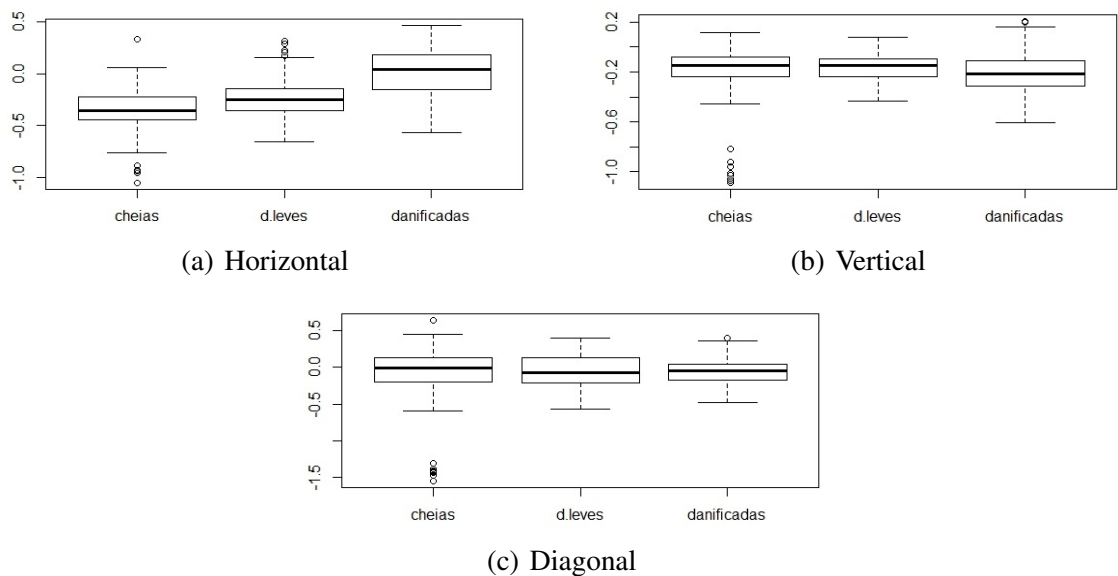
mencionou Jeon, Nicolis e Vidakovic (2014).

Figura 4.3 – Distribuição dos valores dos expoentes de Hurst na direção horizontal, vertical e diagonal considerando a mediana como medida de localização.



Para o caso da mediana como medida de localização, Figura 4.4, também há a presença de outliers, porém os demais valores das distribuições, nas três direções, são mais próximos. Isso dificulta a discriminação dos diferentes tipos de sementes consideradas.

Figura 4.4 – Boxplot dos coeficientes de Hurst para as direções horizontal, vertical e diagonal, considerando a mediana como medida de localização.



Para o grupo das sementes cheias e danificadas, a taxa média de classificação correta, com a média como medida de localização, foi de 99,76%. Já com a mediana essa taxa foi de 80,93%. Para este grupo de sementes a superioridade no poder de classificação com o uso da média ao invés da mediana é evidente. Além disso a taxa de classificação correta, com o uso da média, foi bem maior que as taxas encontradas na metodologia proposta por Sáfadi et al. (2016). Nesse trabalho os autores encontraram 64% de acurácia utilizando as inclinações s_1 e s_2 e 82% usando quatro inclinações, a saber s_1 , s_2 , s_3 e s_4 . Percebe-se que a taxa encontrada, com a mediana como medida de localização, é levemente inferior ao resultado encontrado em Sáfadi et al. (2016), para as quatro inclinações. Nas Figuras 4.5 e 4.6 apresentam-se os histogramas das taxas de acurácia das 1000 simulações. É interessante observar que para este grupo, usando a média como medida de localização, em 813 das 1000 simulações todas as classificações foram feitas de forma perfeita.

Figura 4.5 – Histograma da taxa de classificação correta para as sementes cheias (c) e danificadas (d), em 1000 simulações, considerando a média como medida de localização.

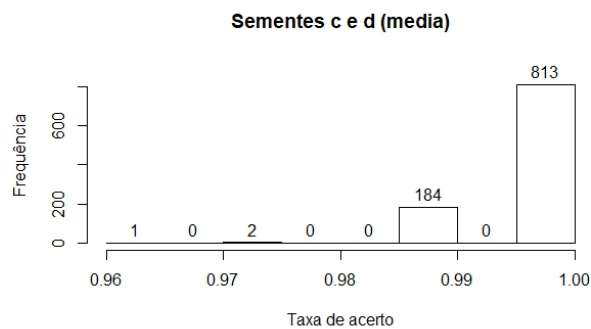
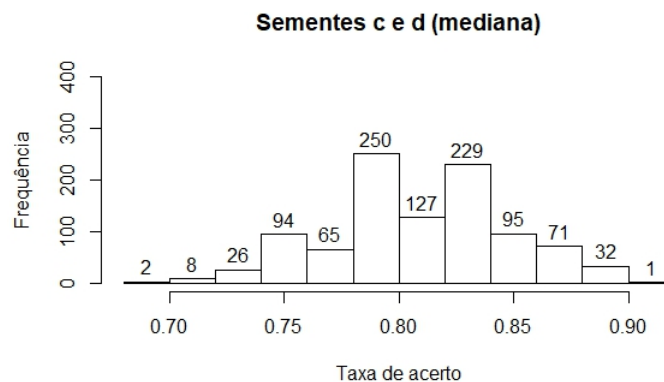


Figura 4.6 – Histograma da taxa de classificação correta para as sementes cheias (c) e danificadas (d), em 1000 simulações, considerando a mediana como medida de localização.



Na Tabela 4.1 apresentam-se as matrizes de médias de confundimento para as 1000 simulações.

Tabela 4.1 – Confundimento médio em 1000 simulações para sementes cheias (c) e danificadas (d), usando a média e a mediana como medidas de localização.

		Classificada como			Classificada como		
		cheia	danificadas	Total	cheia	danificadas	Total
Classificação real	cheia	43.79	0.00	43.79	37.05	6.81	43.86
	danificadas	0.19	35.02	35.21	8.26	26.89	35.15
	Total	43.98	35.02	79	45.31	33.70	79
				(a) Média	(b) Mediana		

Na Tabela 4.1 os elementos da diagonal principal representam o número médio de classificações corretas e os demais elementos no corpo da tabela representam o número médio de erros de classificação (confusão). Observa-se que quase não houve confusão na classificação das imagens utilizando a média como medida de localização. Uma média de 0,19 sementes danificadas foram classificadas como sementes cheias pelo método proposto. Usando a mediana como medida de localização houve confusão tanto para as sementes cheias que foram classificadas como danificadas quanto para as sementes danificadas que foram classificadas como cheias. Um número médio de 8,26 sementes danificadas foram classificadas como cheias e 6,81 sementes cheias foram classificadas como danificadas.

No grupo das sementes levemente danificadas e danificadas os resultados encontrados foram semelhantes. Considerando a média como medida de localização a taxa média de acurácia foi de 99,26% e com a mediana 76,22%. Usando as quatro inclinações, para esse grupo de sementes, Sáfadi et al. (2016) encontraram uma taxa de acurácia de 63%. Aqui, tanto a média como a mediana proporcionaram melhores resultados, ainda que o uso da média tenha proporcionado resultado consideravelmente superior ao da mediana. Os histogramas das taxas de acurácia, para as 1000 simulações, são apresentados nas Figuras 4.7 e 4.8.

Figura 4.7 – Histograma da taxa de classificação correta para as sementes levemente danificadas (dl) e danificadas (d), em 1000 simulações, considerando a média como medida de localização.

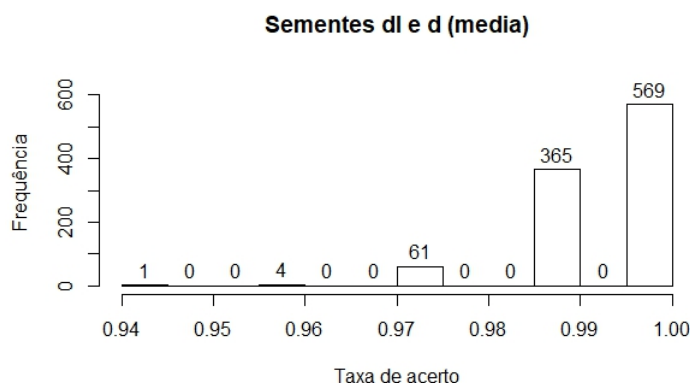
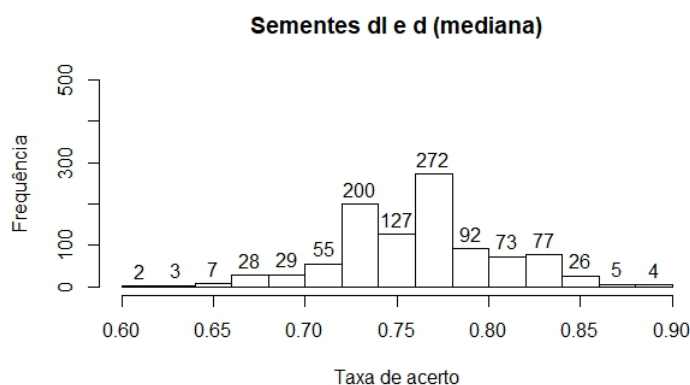


Figura 4.8 – Histograma da taxa de classificação correta para as sementes levemente danificadas (dl) e danificadas (d), em 1000 simulações, considerando a mediana como medida de localização.



Na Tabela 4.2 apresentam-se as matrizes de médias de confundimento para as 1000 simulações. Observa-se que para este grupo o comportamento do grupo formado pelas sementes cheias e danificadas se mantém. Há pouca confusão na classificação quando a média é utilizada. Um número médio de 0,47 sementes levemente danificadas foram classificadas como cheias e 0,03 sementes danificadas, em média, foram classificadas como levemente danificadas. Para a mediana como medida de localização esses números são de 7,27 e 8,90, respectivamente.

Quando o grupo das sementes cheias e levemente danificadas foi considerado houve uma queda considerável na acurácia do método, tanto com a média quanto com a mediana. Com a média como medida de localização a taxa média de acurácia foi de 63,09% e com a mediana foi de 58,28%. Este resultado é bem inferior ao encontrado por Sáfadi et al. (2016), que foi de 89%. É importante destacar aqui a subjetividade e dificuldade na caracterização das sementes como cheias e levemente danificadas, tratando-se de grupos muito semelhantes (LEITE; SÁ-

Tabela 4.2 – Confundimento médio em 1000 simulações para sementes danificadas e levemente danificadas, usando a média e a mediana como medidas de localização.

		Classificada como			Classificada como		
		danificadas	leve. dani.	Total	danificadas	leve. dani.	Total
Classificação real	danificadas	35.10	0.03	35.13	26.38	8.90	35.28
	leve. dani.	0.47	32.40	32.87	7.27	25.45	32.72
	Total	35.57	32.43	68	33.65	34.35	68
				(a) Média	(b) Mediana		

FADI; CARVALHO, 2013). Interessantemente o método proposto por Sáfadi et al. (2016) foi superior na detecção de tal diferença. Nas Figuras 4.9 e 4.10 são apresentados os histogramas das distribuições da taxa de acurácia usando a média e a mediana como medidas de localização.

Figura 4.9 – Histograma da taxa de classificação correta para as sementes cheias (c) e levemente danificadas (dl), em 1000 simulações, considerando a média como medida de localização.

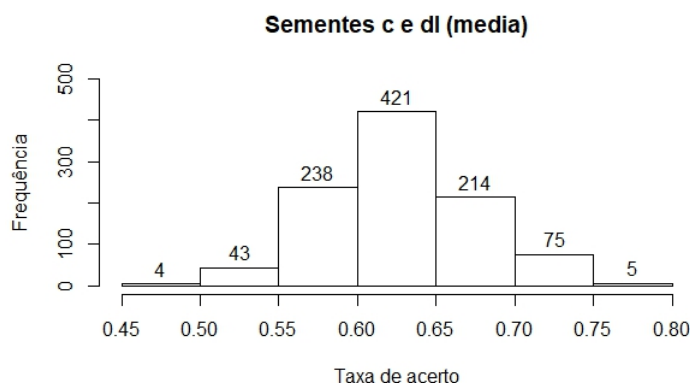
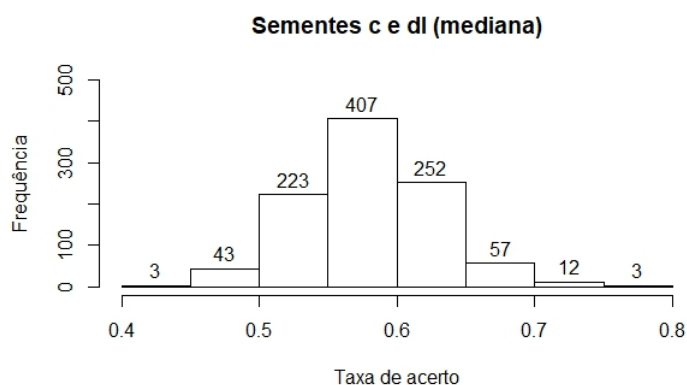


Figura 4.10 – Histograma da taxa de classificação correta para as sementes cheias (c) e levemente danificadas (dl), em 1000 simulações, considerando a mediana como medida de localização.



Na Tabela 4.3 apresentam-se as matrizes de médias de confundimento para as 1000 simulações para o grupo de sementes cheias e levemente danificadas.

Tabela 4.3 – Confundimento médio em 1000 simulações para sementes cheias e levemente danificadas, usando a média e a mediana como medidas de localização.

		Classificada como			Classificada como		
		cheias	leve. dani.	Total	cheias	leve. dani.	Total
Classificação real	cheias	41.39	2.99	44.38	38.13	6.01	44.14
	leve. dani.	25.43	7.19	32.62	26.11	6.75	32.86
	Total	66.82	10.18	77	64.24	12.76	77
				(a) Média	(b) Mediana		

Observa-se(TABELA 4.3) que há uma grande confusão na classificação de sementes levemente danificadas que foram consideradas como cheias. Usando a média como medida de localização, um número médio de 25,43 sementes levemente danificadas foram classificadas erroneamente. Com a mediana como medida de localização esse número médio foi de 26,11.

Na Tabela 4.4 apresenta-se um resumo dos resultados encontrados até o momento com este trabalho e a comparação entre estes resultados e os resultados encontrado por outros autores.

Tabela 4.4 – Taxa média de acurácia para cada grupo de semente e metodologia

Grupo de sementes	Média	Mediana	Sáfadi et al. (2016)
Cheias e danificadas	99,76%	80,93%	64% e 82%
Levemente danificadas e danificadas	99,26%	76,22%	63%
Cheias e levemente danificadas	63,09%	58,28%	89%

Para o grupo de todas as sementes, a taxa média de classificação correta, utilizando a média como medida de localização foi de 74,15% e utilizando a mediana foi 57,05%. Percebe-se aqui a superioridade na acurácia do algoritmo de classificação provocada pelo uso da média ao invés da mediana. O resultado encontrado para a mediana é semelhante ao encontrado em Sáfadi et al. (2016), que foi de 57%. Neste grupo de sementes é, também, possível comparar os resultados com o trabalho de Leite, Sáfadi e Carvalho (2013). Esses autores, usando componentes independentes, conseguiram uma taxa de aproximadamente 80% de acurácia na classificação das sementes. De acordo com a variação no número de componentes independentes essa taxa variou levemente. Porém, nesse trabalho, os autores não utilizaram um processo de simulação para se ter condições de estimar a taxa de acurácia média. Isso dificulta a comparação entre

a metodologia proposta neste trabalho e a utilizada em Leite, Sáfy e Carvalho (2013). Nas Figuras 4.11 e 4.12 apresentam-se os histogramas das taxas de classificação correta nas 1000 simulações feitas.

Figura 4.11 – Histograma da taxa de classificação correta para os três tipos de sementes, em 1000 simulações, considerando a média como medida de localização.

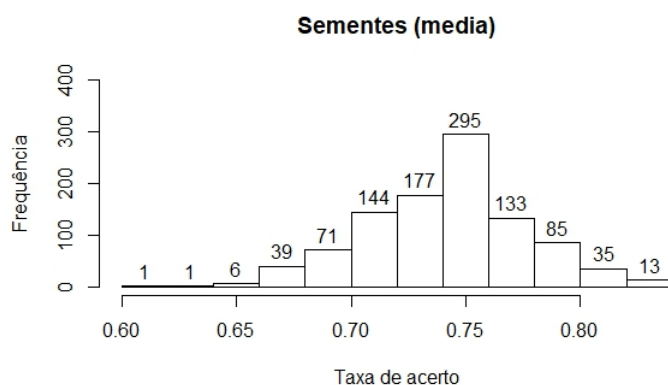
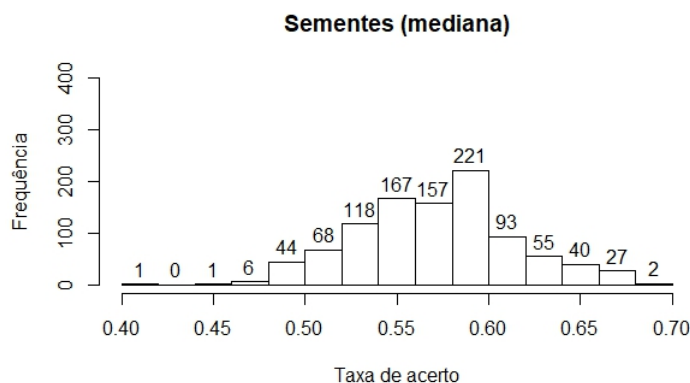


Figura 4.12 – Histograma da taxa de classificação correta para os três tipos de sementes, em 1000 simulações, considerando a mediana como medida de localização.



Na Tabela 4.5 apresentam-se as matrizes de confundimento médio para as 1000 simulações para o grupo de todas as sementes. Novamente há confusão na classificação de sementes levemente danificadas. Um número médio de 25,58 sementes levemente danificadas foram classificadas como cheias, quando a média foi usada como medida de localização. Empregando a mediana como medida de localização esse número foi de 24,36. Em termos percentuais, nesse mesmo grupo de sementes 80,1% foram classificadas de forma errada, usando a média. Usando a mediana a porcentagem de erros de classificação foi de 95%. Essa alta taxa de confusão é fruto da sobreposição das distribuições dos coeficientes de Hurst para o grupo de sementes levemente danificadas e cheias, nas três direções de decomposição (FIGURAS 4.1 e 4.3).

Tabela 4.5 – Confundimento médio para todas as sementes em 1000 simulações, considerando a média e a mediana como medidas de localização.

		Classificada como			
		cheia	danificadas	leve. dani.	Total
Classificação real	cheia	41.56	0.00	2.68	44.24
	danificadas	0.17	35.01	0.03	35.21
	leve. dani.	25.58	0.49	6.48	32.55
	Total	67.31	35.5	9.19	112
(a) Média					
		Classificada como			
		cheia	danificadas	leve. dani.	Total
Classificação real	cheia	36.53	5.66	1.71	43.9
	danificadas	7.29	25.82	2.07	35.18
	leve. dani.	24.36	7.01	1.55	32.92
	Total	68.18	38.49	5.33	112
(b) Mediana					

Na Tabela 4.6 apresenta-se um resumo dos resultados encontrados até o momento com este trabalho e a comparação entre estes resultados e os resultados encontrado por outros autores. Cabe ressaltar que no trabalho de Leite, Sáfyadi e Carvalho (2013) os autores não fizeram estudo de simulação, a taxa de acurácia encontrada é resultado de uma análise com todas as sementes.

Tabela 4.6 – Taxa média de acurácia para cada grupo de semente e metodologia

Grupo de sementes	Média	Mediana	Sáfyadi et al. (2016)	Leite, Sáfyadi e Carvalho (2013)
Todas as sementes	74,15%	57,05%	57%	80%

5 CONCLUSÕES

O uso do expoente de Hurst, juntamente com a transformada de ondaletas não decimada foi eficiente para a classificação das sementes. O emprego da média como medida de localização, no cálculo do coeficiente de Hurst, produziu um resultado superior comparado ao uso da mediana, para o conjunto de imagens considerado. A técnica proposta neste trabalho apresentou, na maioria dos cenários, bons resultados. A dificuldade humana na distinção entre sementes cheias e levemente danificadas foi refletida no método proposto, que teve sua acurácia bastante reduzida quando este grupo de sementes foi considerado.

5.1 Trabalhos futuros

Pretende-se mudar a forma usada para obtenção do expoente de Hurst e verificar se há melhoria na acurácia da identificação de características das imagens sementes. Os estimadores trimean propostos por Feng e Vidakovic (2017) para o cálculo do expoente de Hurst parecem ser um bom ponto de partida. Segundo os autores, esses métodos reduzem a variância dos estimadores e aumentam a precisão da predição na maioria dos casos estudados por eles. Outra pretensão é acrescentar bancos de dados de diferentes tipos de sementes e verificar o método mais eficiente na classificação das imagens na maioria dos casos.

REFERÊNCIAS

- ANTONINI, M. et al. Image coding using wavelet transform. **IEEE Transactions on image processing**, Charlottesville, v. 1, n. 2, p. 205–220, 1992.
- BURGUER, W.; BURGUE, M. J. **Digital image processing: An algorithmic introduction using java**. 2. ed. New York: Springer, 2016. 811 p.
- BURRUS, C. S.; GOPINATH, R. A.; GUO, H. **Introduction to wavelets and wavelets transforms: A primer**. 1. ed. New Jersey: Prentice Hall, 1998. 257 p.
- CHANG, S. G.; YU, B.; VETTERLI, M. Adaptive wavelet thresholding for image denoising and compression. **IEEE Transactions on image processing**, Charlottesville, v. 9, n. 9, p. 1532–1546, 2000.
- DAUBECHIES, I. **Ten Lectures on Wavelets**. 2. ed. Philadelphia: SIAM, 1992. 357 p.
- DELOUILLE, V.; JANSEN, M.; SACHS, R. von. Second-generation wavelet denoising methods for irregularly spaced data in two dimensions. **Signal Processing**, Amsterdam, v. 86, n. 7, p. 1435–1450, jul. 2006.
- EMBRAPA. **Girassol**. [S.l.], 2018. Disponível em: <<https://www.embrapa.br/soja/cultivos/girassol>>. Acesso em: 17 jul. 2018.
- FACELI, K. et al. **Inteligência Artificial: Uma abordagem de aprendizado de máquina**. 1. ed. Rio de Janeiro: LTC, 2017. 378 p.
- FENG, C.; VIDAKOVIC, B. Estimation of the hurst exponent using trimean estimators on nondecimated wavelet coefficients. **JOURNAL OF LATEX CLASS FILES**, v. 14, n. 8, 2017. Disponível em: <<https://arxiv.org/pdf/1709.08775.pdf>>. Acesso em: 24 mai. 2018.
- FERNÁNDEZ-DELGADO, M. et al. Do we need hundreds of classifiers to solve real world classification problems? **Journal of Machine Learning Research**, New York, v. 15, n. Oct, p. 3133–3181, 2014.
- GOSWAMI, J. C.; CHAN, A. K. **Fundamental of wavelets: theory, algorithms and applications**. 2. ed. Hoboken: WILEY, 2011. 359 p.
- HURST, H. E. The problem of long-term storage in reservoirs. **Hydrological Sciences Journal**, London, v. 1, n. 3, p. 13–27, 1956. Disponível em: <<https://doi.org/10.1080/02626665609493644>>.
- JEON, S.; NICOLIS, O.; VIDAKOVIC, B. Mammogram diagnostics via 2-d complex wavelet-based self-similarity measures. **Journal of Mathematical Sciences**, São Paulo, v. 8, n. 2, p. 265–284, 2014.
- KANG, M.; VIDAKOVIC, B. MEDL and MEDLA: Methods for assessment of scaling by medians of log-squared nondecimated wavelet coefficients. 2017. Disponível em: <<https://arxiv.org/pdf/1703.04180.pdf>>.
- KUHN, M. caret: Classification and regression training. **R**, Package Version 6.0-78, 2017. Disponível em: <<https://cran.r-project.org/web/packages/caret/caret.pdf>>. Acesso em: 17 jul. 2018.

- LEITE, I. C. C.; SÁFADI, T.; CARVALHO, M. L. M. de. Evaluation of seed radiographic images by independent component analysis and discriminant analysis. **Seed Science and Technology**, Boston, v. 41, n. 2, p. 235–244, 2013.
- LIMA, P. C. de. Wavelets: uma introdução. **Matemática Universitária**, São Paulo, v. 33, n. 1, p. 13–44, 2002.
- LOPES, A. C. A.; NASCIMENTO, W. M. **Análise de Sementes de Hortaliças**. Embrapa hortaliças. Circular Técnica, 83. Brasília, 2009. 9 p. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/783114/4/ct83.pdf>>. Acesso em: 24 mai. 2018.
- LYASHENKO, V. V. et al. Hurst exponent as a part of wavelet decomposition coefficients to measure long-term memory time series based on multiresolution analysis. **American journal of systems and software**, San Vicente del Raspeig, v. 4, n. 2, p. 51–56, 2016.
- MALLAT, S. G. Multiresolution approximations and wavelet orthonormal bases of $L^2(\mathbb{R})$. **Transactions of the American Mathematical Society**, New York, v. 315, n. 1, p. 69–87, 1989.
- MEYER, D. et al. e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), tu wien. **R**, Package Version 1.6-8, 2017. Disponível em: <<https://cran.r-project.org/web/packages/e1071/e1071.pdf>>. Acesso em: 17 jul. 2018.
- MORETTIN, P. A. **Ondas e ondaletas: da análise de fourier à análise de ondaletas de séries temporais**. 1. ed. São Paulo: EdUSP, 2014. 276 p.
- NASON, G. P. **Wavelet methods in Statistics with R**. 1. ed. New York: Springer, 2008. 257 p.
- NICOLIS, O.; RAMIREZ-COBO, P.; VIDAKOVIC, B. 2d wavelet-based spectra with applications. **Computational statistics and data analysis**, Amsterdam, v. 55, n. 1, p. 738–751, jan. 2011.
- OLIVEIRA, H. M. de. **Análise de Fourier e Wavelets: sinais estacionários e não estacionários**. 1. ed. Recife: EdUFPE, 2007. 342 p.
- PIMENTEL, E. A.; SILVA, J. F. da. Decomposição de ondaletas, análise de volatilidade e correlação de índices financeiros. **Estudos Econômicos**, São Paulo, v. 41, n. 2, p. 441–462, 2011.
- RAMIREZ-COBO, P. et al. A wavelet-based spectral method for extracting self-similarity measures in time-varying two-dimensional rainfall maps. **Journal of time series analysis**, Hoboken, v. 32, n. 4, p. 351–363, jul. 2011.
- RAMIREZ-COBO, P.; VIDAKOVIC, B. A 2d wavelet-based multiscale approach with applications to the analysis of digital mammograms. **Computational statistics and data analysis**, Amsterdam, v. 58, n. 2, p. 71–81, 2013.
- REMENYI, N. et al. Image denoising with 2d scale-mixing complex wavelet transforms. **IEEE Transactions on image processing**, Charlottesville, v. 23, n. 12, p. 5165–5174, dez. 2014.
- RODRIGUES, L. L. M. **Análise de variância em séries temporais: uma abordagem usando ondaletas**. 100 p. Tese (Doutorado) — Universidade Federal de Lavras, Lavras, 2015.

SÁFADI, T. et al. Wavelet-based spectral for descriptors of detection of damage in sunflower seeds. **International Journal of Wavelets, Multiresolution and Information Processing**, London, v. 14, n. 4, p. 235–244, 2016.

SHAPIRO, J. M. Embedded image coding using zerotrees of wavelets coefficients. **IEEE Transactions on image processing**, Charlottesville, v. 41, n. 12, p. 3445–3462, 1993.

TUSZYNSKI, J. Tools: moving window statistics, gif, base64, roc auc, etc. **R**, Package Version 1.17.1, 2014. Disponível em: <<https://cran.r-project.org/web/packages/caTools/caTools.pdf>>. Acesso em: 17 jul. 2018.

WHITCHER, B. Waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing. **R**, Package Version 1.7.5, 2015. Disponível em: <<https://cran.r-project.org/web/packages/waveslim/waveslim.pdf>>. Acesso em: 17 jul. 2018.