

Extração de Atributos Via Modelos Arma 1D/2D, Avaliação Estatística da Distância JM e Classificador Condicional Contextual.

Gilmar Rosa
Instituto Nacional de
Pesquisas Espaciais - INPE
gilmar@dpi.inpe.br

Luciano Vieira Dutra
Instituto Nacional de
Pesquisas Espaciais - INPE
dutra@dpi.inpe.br

Corina da Costa Freitas
Instituto Nacional de
Pesquisas Espaciais - INPE
corina@dpi.inpe.br

Resumo

O objetivo desta pesquisa é uma comparação dos métodos de extração de atributos via modelagem ARMA unidimensional e na sua forma Bidimensional; definir a significância estatística da distância JM uma vez que mesma cresce à medida que se adiciona mais atributo (critério da monotocidade) e ainda propor um classificador contextual condicional de tal forma que: Se a diferença da verossimilhança for estritamente diferente dos vizinhos logo se classifica pontualmente, caso contrário classifica-se usando a informação de contexto. Implementar em linguagem IDL, um software de classificação de imagens digitais que permita a agregação futura novas metodologias ou a implementação das já existentes.

Palavras-Chave: ARMA 1D, ARMA 2D, distância JM, Bootstrap, Cadeias de Markov, Classificação.

1. Introdução

Os métodos de extração de atributos via modelagem ARMA unidimensional foram largamente estudados por autores tais como: Dutra, 1990; Felgueira & Dutra, 1999. Estes métodos fornecem uma maneira alternativa de se estudar as propriedades texturais de uma imagem de maneira similar a abordagem bidimensional, que são mais difíceis de implementação e nem sempre determinadas estatísticas podem ser assumidas como por exemplo causalidade e computabilidade recursiva.

Os métodos bidimensionais dependem de conceitos tais como região de suporte e computabilidade recursiva e, além disso, tem um custo computacional muito alto, que de certa forma inviabiliza seu uso dependendo do tamanho da imagem a ser analisada. Embora os métodos unidimensionais tenham menos custo

computacional sua formulação não leva em consideração a dependência entre linhas de uma imagem o que pode causar um vício na extração de elementos texturais, desta forma um estudo comparativo entre essas duas metodologias é de grande importância dentro do contexto de extração de atributos.

Numa abordagem de seleção de atributos, um fenômeno que ocorre quando usamos as distâncias entre distribuições é chamado de *monotocidade* onde se observa que à medida que aumenta o número de atributos também o valor da distância entre distribuições também tendem a crescer, sendo o valor máximo atingido quando todos os atributos forem selecionados (Siedlecki & Sklansky, 1988). Desta forma responder a pergunta qual o número de atributos que discrimina com melhor exatidão um certo número de classes, não é uma questão fácil, se estamos usando uma abordagem filtro¹. Neste contexto estabelecer a significância estatística entre as distâncias JM do conjunto de atributo k e $k+1$, nos dá uma ferramenta ou critério de parada para seleção do menor conjunto de atributos k em um total de n atributos.

No contexto de classificação, esta pesquisa pretende desenvolver um classificador contextual condicional de tal forma que: Se a diferença da verossimilhança for estritamente diferente dos vizinhos logo se classifica pontualmente, caso contrário classifica-se usando a informação de contexto. Em princípio estaremos assumindo gaussianidade dos dados amostrais, o que torna a comparação das verossimilhanças pontuais menos complexas, mas outras distribuições podem ser assumidas. Desta forma se concluirmos que as verossimilhanças entre os vizinhos daquele ponto a ser classificado

¹ As abordagens filtros mais largamente utilizadas são baseadas nos índices de separabilidade entre classes ou distribuições

forem estatisticamente diferentes, estaremos, remetendo o problema de classificação para abordagem contextual, ou seja, *cadeias de Markov*, as quais poderão melhorar a classificação considerando uma certa confusão que a verossimilhança não consegue distinguir.

2. Extração de Atributos

Os métodos de extração de atributos via modelagem ARMA unidimensional foram largamente estudados por Dutra, 1990; Felgueira & Dutra, 1999. Estes métodos fornecem uma maneira alternativa de se estudar as propriedades texturais de uma imagem de forma aproximada da abordagem bidimensional. A abordagem bidimensional é mais difícil de implementação e dependem conceitos tais como região de suporte e computabilidade recursiva e, além disso, tem um custo computacional muito alto, que de certa forma inviabiliza seu uso dependendo do tamanho da imagem a ser analisada.

2.1 Modelos Arma Unidimensional.

Uma metodologia bastante utilizada na análise de series temporais é conhecida como abordagem de Box e Jenkins (1970) & Box, Jenkins e Reinsel (1994). Tal metodologia consiste em ajustar modelos auto-regressivos de médias móveis, ARMA(p,q), a um conjunto de dados. A estratégia para construção do modelo é baseada em ciclo iterativo, no qual a escolha da estrutura do modelo é baseada nos próprios dados. Os modelos discutidos neste trabalho são baseados na idéia de que uma série temporal em que seus valores sucessivos são altamente dependentes podem ser considerados como uma realização de um conjunto de valores independentes, "shocks" a_t . "Shocks" são realizações de uma distribuição de probabilidade conhecida, tendo média zero e variância (σ_a^2) constante. O conjunto destes "shocks" ou seqüências de variáveis aleatórias $a_t, a_{t-1}, a_{t-2}, \dots$, é denominado de *ruído branco (white noise)*.

Através do processo a_t , obtém-se o processo Z_t através do chamado *filtro linear* como mostrado na figura a seguir:

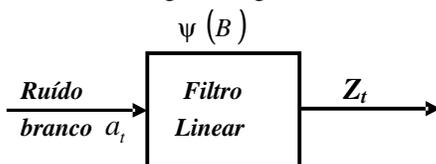


Figura 1: Representação de uma série como uma saída de *filtro linear*

O operador que transforma a_t em Z_t é chamado de *função de transferência* do filtro.

A operação de filtragem linear requer uma soma ponderada das observações anteriores do ruído branco a_t , tal que:

$$Z_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots \\ = \mu + \psi(B) a_t$$

Onde μ é o parâmetro que determina a media do processo, e

$$\psi(B) = 1 + \psi_1 B + \psi_2 B^2 + \dots$$

O operador que transforma a_t em Z_t é chamado de função de transferência do filtro.

Os filtros AR, MA e ARMA, são dados a seguir:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \dots + \phi_p \tilde{Z}_{t-p} + a_t \quad (\text{Filtro AR})$$

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \quad (\text{Filtro MA})$$

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \dots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (\text{Filtro ARMA})$$

$$\text{Sendo: } \tilde{Z}_t = Z_t - \mu$$

2.2 Modelos ARMA bidimensional

Uma seqüência complexa 2D ou matriz 2D $Z[m, n]$ é uma função definida sobre os inteiros $-\infty < m, n < \infty$ no plano 2D. A região sobre a qual a seqüência ou matriz 2D não é zero é chamada região de suporte. Uma típica matriz 2D finita é uma imagem a qual tem região de suporte $0 \leq m \leq M-1; 0 \leq n \leq N-1$.

A saída $Z[m, n]$ de um sistema linear 2D, considerando a entrada $a[m, n]$ pode ser expressa como

$$z[m, n] = \sum_{k=-\infty}^{\infty} \sum_{l=-\infty}^{\infty} h_{kl}[m, n] a[k, l]$$

Onde $h_{kl}[m, n]$ é a resposta do sistema linear através $[m, n]$ para uma unidade de impulso aplicado em $[k, l]$. De maneira bem sucinta os modelos 2D, AR, MA e ARMA, respectivamente são dados pelas seguintes equações diferenças:

$$\tilde{Z}[m, n] = - \sum_{[k,l] \neq [0,0]} \sum_k \phi[k,l] \tilde{Z}[m-k, n-l] + a[m,n]$$

$$\tilde{Z}[m, n] = \sum_{ki} \sum_l \theta[k, l] a[m-k, n-l]$$

$$\tilde{Z}[m, n] = - \sum_{[k,l] \neq [0,0]} \sum_k \phi[k,l] \tilde{Z}[m-k, n-l] + a[m,n]$$

$$+ \sum_k \sum_l \theta[k, l] a[m-k, n-l]$$

Sendo $\phi[0,0]$ e $\theta[0,0]=1$

3. Seleção de Atributos

Tem como principal objetivo selecionar um subconjunto de atributos relevantes, dentre todos os atributos disponíveis. De maneira simples, pode-se definir um atributo como relevante se ele é capaz de distinguir exemplos pertencentes a classes diferentes., ou seja, objetivo dos algoritmos de seleção de atributos é escolher o menor subconjunto que oferece a melhor classificação (acurácia global) , considerando os custos de processamento.

Perceptivamente, quanto maior o número de atributos em uma base de dados, maior o poder discriminatório do classificador. No entanto, o tempo computacional do algoritmo aumenta agressivamente de acordo com o número de atributos presentes e, além disso, ocorre um decréscimo do valor da acurácia global. Este fenômeno é conhecido como maldição de dimensionalidade (Bishop, 1995).

3.1 Distância JM.

B_{ij} representa a distância de Bhattacharyya entre as classes i e j dada por:

$$B_{i,j} = \frac{1}{8}(\mu_i - \mu_j)^T \left[\frac{\Sigma_i + \Sigma_j}{2} \right]^{-1} (\mu_i - \mu_j) + \frac{1}{2} \ln \frac{\frac{|\Sigma_i + \Sigma_j|}{2}}{\sqrt{|\Sigma_i| |\Sigma_j|}}$$

onde Σ_i e Σ_j são as matrizes de covariância das classes i e j , μ_i e μ_j são os respectivos vetores de média.

Dado a distância de Bhattacharyya (B_{ij}), deriva-se à distância Jeffries-Matusita (JM), Dutra & Huber, (1999) ; Kavzoglu & Mather, (2002) ; Ndi Nyongui, A., Tonye & Akono, E., (2002); Serpico, B. S., Bruzzone, L., (2001), sendo dada por:

$$JM_{ij} = \sqrt{2 (1 - e^{-B_{ij}})}, \text{ onde } JM_{ij} \in [0, \sqrt{2}]$$

Alguns autores preferem usar a formulação: (Swain & King, 1973 ; Richards, 1999)

$$JM_{ij} = 2 (1 - e^{-B_{ij}}), \text{ onde: } JM_{ij} \in [0, 2]$$

A Distância B cresce sem saturação e pode causar distorções em classes muito separadas, logo o uso da distancia Jeffries-Matusita (JM) tem um comportamento saturante em $\sqrt{2}$ ou 2, dependendo da formulação.

A distância Jeffries-Matusita (JM) fornece unicamente uma medida de distância entre duas densidades de duas classes, o uso estendido para várias classes é feito através da média Jeffrey-Matusita de todas as classes, que é dado por:

$$JM = \frac{1}{m(m-1)} \sum_{i=1}^{m-1} \sum_{j=i+1}^m JM_{ij},$$

Onde m é o numero de classes.

Como função critério, a média das distâncias JM é uma das mais conhecidas e utilizadas pela comunidade de sensoriamento remoto para seleção de atributos em problemas com várias classes , Swain & Davis, (1978; Bruzzone, Roli & Serpico, (1995); Bruzzone & Serpico, (2000).

3.2 Estratégia de Busca: Busca Sequencial Progressiva (SFS) e Busca Sequencial Regressiva (SBS)

O algoritmo SFS (Seqüencial *Foward Feature Selection*) inicia com um conjunto vazio. A cada iteração é selecionada a característica que, unida ao conjunto determinado pela iteração anterior, produz o melhor resultado da função critério. Essa característica é adicionada ao conjunto de características anterior e uma nova iteração é realizada.

Este método tem menor custo computacional quando se deseja obter conjuntos pequenos em relação ao total de atributos (Jain e Zongker, 1997). Uma vez que um atributo tenha sido selecionado, ele não pode ser descartado do subconjunto, o que pode provocar o chamado efeito *nesting*. Este efeito ocorre quando o subconjunto ótimo não contém elementos do conjunto já selecionado, o que impossibilita que seja obtido o conjunto de atributos ótimo.

A seleção seqüencial Regressiva (SBS), ao contrário da seleção seqüencial para frente, inicia a busca por um subconjunto de atributos ótimos com uma solução representando todos os atributos, e a cada iteração um atributo é removido da solução atual, até que não se consiga melhorar a qualidade da solução encontrada.

O algoritmo SBS inicia com um conjunto de atributos completo ($Y_0 = X$) e, nas iterações do algoritmo, remove-se o atributo com o mínimo valor da função critério $J(Y_k + X)$, onde

Y_k é o conjunto de atributos já selecionados e X é o atributo a ser removido. O método SBS tem menor custo computacional, quando se deseja obter conjuntos grandes em relação ao total de atributos. Uma característica desse método é que uma vez eliminado o atributo, ele não retornará ao subconjunto.

A diferença entre SBS e SFS é que o SBS é iniciado com o conjunto de características completo (contendo todos os X atributos), eliminando um a um as menos importantes, ou seja, remove-se o atributo que maximize a função critério para os restantes.

Um fenômeno que ocorre quando usamos as distâncias entre distribuições é chamado de *monotocidade* onde se observa que à medida que aumenta o número de atributos também o valor da distância entre distribuições também tendem a crescer, sendo o valor máximo atingido quando todos os atributos forem selecionados (Siedlecki & Sklansky, 1988).

4. Bootstrap – Estimação de Parâmetros Via Método de Reamostragem.

Esse método consiste em uma técnica de reamostragem que permite aproximar a distribuição de uma função das observações a partir da distribuição empírica dos dados, baseado em uma amostra de tamanho finito. A amostragem é feita com reposição da amostra original (*bootstrap* não - paramétrico). Nessa situação, supomos que as observações são obtidas da função de distribuição empírica, \hat{F} , que designa uma massa de probabilidade igual a $1/n$ para cada ponto amostral. Já para o caso de *bootstrap* paramétrico, a amostragem é feita usando a distribuição da qual os dados são obtidos.

O *bootstrap* pode ser paramétrico ou não paramétrico. O *bootstrap* não paramétrico considera que a função de distribuição dos dados, F , é desconhecida e é estimada através da distribuição empírica \hat{F} . Atualmente, devido aos recursos computacionais disponíveis, o método *bootstrap*, além de ser usado para medir a precisão de estimadores, é aplicado também na construção de intervalos de confiança, testes de hipóteses, estimação do vício, etc. O algoritmo *bootstrap* pode ser resumido da seguinte forma:

a) A partir do conjunto de dados gere um grande número de amostra *bootstrap*, digamos B amostras *bootstrap*, $\{X^{\otimes 1}, X^{\otimes 2}, \dots, X^{\otimes B}\}$, cada

uma de tamanho n .

b) Para cada amostra *bootstrap* gerada, estime o parâmetro de interesse, ou seja, determine $\hat{\theta}^{\otimes(b)} = s(X^{\otimes b})$, com $b = 1, 2, \dots, B$.

O valor ideal para B depende da finalidade para qual a teoria esta sendo utilizada. Por exemplo, se desejamos somente estimar o erro padrão, Efron e Tibshirani (1993) sugerem valores de B entre 50 a 200.

5. Classificador de Máxima Verossimilhança e Contextuais Markovianos – ICM e MPM.

O classificador de máxima verossimilhança é um método de classificação pixel a pixel e pode ser definido por: cada um dos N pixels, a partir de seu vetor de atributos \mathbf{x} , será classificado como sendo da classe ω_j se

$$g_j(\mathbf{x}) > g_k(\mathbf{x}) \text{ para todo } j \neq k, \\ j, k = 1, \dots, N$$

Onde g é a função discriminante. Para o classificador de máxima verossimilhança essa função segue uma distribuição gaussiana multivariada sendo dada por:

$$g_j(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - m_j)' \Sigma_j^{-1} (\mathbf{x} - m_j) - \frac{1}{2} \ln |\Sigma_j| + \ln P(\omega_j)$$

Onde m_j e Σ_j são o vetor de média e a matriz de covariância e $P(\omega_j)$ é a probabilidade a priori da classe j . Richards, (1993).

A abordagem contextual assume que a imagem tem propriedades Markovianas, ou seja, de que o rótulo de um *pixel*, Besag, (1974), w depende dos rótulos dos vizinhos locais. Os dois algoritmos de classificação mais largamente usados nesta abordagem são: *Iterated Conditional Modes* (ICM) e *Maximizer Of Posterior Marginals* (MPM). (Tso & Mather, 2001)

6. Desenvolvimento

Este trabalho visa a comparação dos métodos de extração de atributos via modelagem ARMA unidimensional e na sua forma Bidimensional, definir a significância estatística da distância JM uma vez que mesma cresce a medida que se adiciona o mais atributos (critério de monotocidade) e ainda propor um classificador contextual condicional de tal forma que: Se a diferença da verossimilhança for estritamente diferente dos vizinhos logo se classifica pontualmente, caso contrário classifica-se usando a informação de contexto.

Para Alcançar os objetivos propostos esta sendo construído um software baseado na

linguagem IDL (*Interactive Data Language*) denominado de Sistema de Classificação de Imagem Digital, SCID, Figura 2,a seguir, onde se pode agregar vários módulos independentemente. Os módulos implementados serão extração de atributos usando a metodologia unidimensional e bidimensional, neste caso também será executado uma avaliação das duas metodologias. As propriedades estatísticas da distância JM serão avaliadas usando-se o método de reamostragem *bootstrap* em sua formulação não paramétrica, com implementação correspondente no SCID, uma vez que mesma cresce à medida que se adiciona mais atributo (critério da monotocidade), logo definir ate que ponto este crescimento é significativo nos remetera um critério para finalizar o algoritmo de busca sub-otimo, como por exemplo, SFS.

Na proposição do classificador de máxima verossimilhança condicional contextual, alem da implementação no SCID, o estudo das propriedades estatísticas das verossimilhanças entre o ponto e seu vizinho será feita considerando uma distribuição paramétrica para o classificador, no caso mais geral será considerada a distribuição normal, logo isso permitira a seguinte regra de decisão: Se a diferença da verossimilhança do ponto estudado, for estritamente diferente dos vizinhos logo se classifica pontualmente, caso contrário classifica-se usando a informação de contexto, neste caso algoritmos tais como MPM e ICM.

7. Comentários Finais.

A extração de atributos via modelagem ARMA unidimensional tem mostrado boa performance em relação ao custo computacional, no entanto, seu desempenho, em termos de extração de atributos, em relação a metodologia bidimensional, ainda não foi avaliada na literatura, considerando os modelos AR, MA e ARMA e ainda não se conhece uma implementação em software comercial, dessa abordagem, o que em estudos preliminares tem trazido bons resultados.

Na abordagem seleção de atributos, o estudo das propriedades estatísticas da distância JM, trará uma metodologia de definição automática do numero de atributos a ser

selecionado no caso, do uso da abordagem filtro para seleção de canais. Atualmente esta decisão é um critério subjetivo do pesquisador.

Considerando a abordagem de classificação de imagem, o classificador ora pretendido, será uma ferramenta de classificação mista entre classificador de máxima verossimilhança tradicional e os métodos de classificação contextual. Espera-se que este classificador contextual condicional, por ser oriundo da combinação de dois classificadores produza resultados bem aceitáveis uma vez que somente a abordagem contextual, tem custo computacional elevado, por este algoritmo ser iterativo.

8. Referências

- [1] Aksasse, B., Radoune, L., *Two-dimensional autoregressive (2-D AR) model order estimation.*, IEEE Transactions on signal processing, vol. 47, nr 7, july 1999.
- [2] Besag, J. *Spatial interaction and the statistical analysis of lattice systems.* Journal of the Royal Statistical Society, 36:192-236, 1974.
- [3] Box, G., E., P., Jenkins, G., M., *Time series analysis: Forecasting and Control*, 2nd ed. Holden-Day, San Francisco, 1970.
- [4] Bruzzone L., Roli, F., and Serpico, S. B., "An extension to multiclass cases of the Jeffreys-Matusita distance," IEEE Trans. Geosci. Remote Sensing, vol. 33, pp. 1318–1321, Nov. 1995.
- [5] Duda, R. O., Hart, P. E., Stork, D. G. *Pattern Classification*, 2nd edition, John Wiley & Sons, 2001. 654p.
- [6] Dutra, L.V. *Classificação de texturas usando modelos ARMA e distâncias da função de autocorrelação.* São José dos Campos. (INPE-5067-TDL/406). Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, 1990, 136 P.
- [7] Dutra, L.V.; Huber, R. *Feature Extraction and Selection for ERS-1/2 InSar Classification*, International Journal Remote Sensing, 1999, vol. 20 no. 5, 993-1016.



Figura 2: Tela inicial do SCID – Sistema de Classificação de Imagens Digitais

- [8] Efron, B., Tibshirani, R. J., *An Introduction to the Bootstrap*, Chapman Hall, 1993.
- [9] Ferreira, A., C., Soares, J., F., Cruz, F., R., B., *Reconhecimento de Padrões em Estatística: Uma abordagem comparativa*, Proceedings of the Brazillian Conference on Neural Networks – pp 409-411, April, 2001- Rio de Janeiro- Brazil.
- [10] Gonzalez, R.C.; Woods, R.E. *Digital image processing*. New York, Addison-Wesley, 1992. 716p.
- [11] L. Bruzzone., Serpico, S. B., “A technique for feature selection in multiclass cases,” *Int. J. Remote Sensing*, vol. 21, pp. 549–563, 2000.
- [12] Kimura, H., Honoki, Y., A hybrid approach to high resolution two-dimensional spectrum analysis, *IEEE Transactions on acoustics, speech, and signal processing*, Vol. ASSP-35, Nr 7, July 1987.
- [13] Marzetta, T., Two-dimensional linear prediction: autocorrelation arrays, minimum-phase predictions error filters and reflection coefficient arrays., *IEEE Transactions on signal processing*, vol. ASSP-28 Nr 6, December, 1980.
- [14] Marroquin, J., Mitter, S., Poggio, T., 1987. *Probabilistic solution of ill-posed problems in computational vision*. Journal of the American Statistical Association, no. 82, pp. 76-89.
- [15] Kavzoglu, T., Mather, P.M., *The Role of Feature in Artificial Neural Network Applications*, *Int. J. Remote Sensing*, vol. 23, pp. 2919–2937, 2002.
- [16] Kudo, M., Sklansky., *Comparison of algorithms that select features for pattern classifiers*, *Pattern Recognition*, Vol 33, pp. 25-41, 2000
- [17] Liu, H., Dash, M., *Feature selection for classification*, *Intelligent Data Analysis 1* (1997) pp. 131-156
- [18] Nyongui, A., Tonye, E., Akono, A., *Evaluation of Speckle Filtering and Texture Analysis Methods for Land Cover Classification from SAR Images*, *Int. J. Remote Sensing*, vol. 23, pp. 1895–1925, 2002.
- [19] Jain, A. K., and Zongker, D., *Feature-selection: Evaluation, application, and small sample performance*. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(2):152-157, February 1997.
- [20] Piramuthu S., *Evaluating Feature Selection Methods for Learning in Data Mining Applications*, *IEEE Conference on Systems Science*, 1998, pp. 294-301
- [21] Ojeda, S.; M.; Vallejos, O., R., Lucini, M., M., *Performance of robust RA estimator for bidimensional autoregressive models*, *J. Statist. Comput. Simul.*, 2002, Vol 72(1), pp. 47-62.
- [22] Richards, J.A. *Remote sensing digital image analysis: an introduction*. Revised and Enlarged Edition, 1993.
- [23] Ripley, B. D. (1996), *Pattern Recognition and Neural Networks*, Cambridge University Press, Cambridge.
- [24] Somol, P., Pudil, P., Novovicová, J., and Paclík, P., *Adaptive floating search methods in feature selection*. *Pattern Recognition Letters*, 20:1157-1163, 1999.
- [25] Swain P. H. and S. M. Davis, *Remote sensing: the quantitative approach*. New York: McGraw-Hill, 1978.
- [26] Theodoridis, S., Koutroumbas, K.: *Pattern Recognition*. Academic Press (1999).
- [27] Tso B, Mather P M. *Classifications methods for remotely sensed data*, New York, Taylor & Francis, 2001.
- [28] Zhang, D.; Cheng, J., *High resolution two-dimensional ARMA spectral estimation*, *IEEE Transactions on signal processing*, Vol 39, Nr 3, March, 1991
- [29] Zhang, D.; Cheng, J., *High resolution two-dimensional ARMA spectral estimation*, *IEEE Transactions on signal processing*, Vol 39, Nr 3, March, 1991