

COMPARAÇÃO DOS MÉTODOS SUPPORT VECTOR MACHINE E MÁXIMA VEROSSIMILHANÇA GAUSSIANA PARA CLASSIFICAÇÃO DE IMAGEM DAS ÁREAS DE DUNAS COSTEIRAS DA PRAIA DO CASSINO, RIO GRANDE, RS

Lisandra Roldão Miranda ¹, Delnir Monteiro de Lemos ¹, Jean Marcel de Almeida Espinoza ¹

¹ IFRS - Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – Campus Rio Grande, Rua Engenheiro Alfredo Huch, 475, Bairro Centro, CEP 96201-460, Rio Grande, RS, Brasil

RESUMO

A classificação de imagens de satélites possui uma variedade de métodos para análise espectral, porém a escolha deles dependerá da área a ser estudada. Este trabalho apresenta, com as técnicas de Processamento Digital de Imagens, a classificação supervisionada da extensão das dunas costeiras do Balneário Cassino, no município de Rio Grande, utilizando os classificadores Support Vector Machines e Máxima Verossimilhança Gaussiana. Essas classificações proporcionaram implementar a análise de acurácia baseada na matriz erro e a comparação entre os classificadores. A partir da imagem de alta resolução espacial Geoeye, com o auxílio do software MatLab, as amostras de treinamento foram coletadas para treinar os algoritmos de classificação implementados na cena segmentada. Os resultados deste trabalho foram avaliados visualmente, assim comparando os métodos pela sua eficiência na classificação de imagem da área de estudo e trazendo subsídios para futuros interessados em classificadores para a área de dunas costeiras.

Palavras-chave — classificação, PDI, dunas, SVM, supervisionada.

ABSTRACT

The classification of satellite images has a variety of methods for spectral analysis, but the choice of them depends on the area being studied. This work presents, with the techniques of Digital Image Processing, supervised classification of the extension of the coastal dunes of the Balneário Cassino, in the city of Rio Grande, using the classifiers Support Vector Machines and Maximum Gaussian Likelihood. These classifications provided to implement the analysis of accuracy based on the matrix error and the comparison between the classifiers. From the high spatial resolution Geoeye image, with the aid of MatLab software, the training samples were collected to train the classification algorithms implemented in the segmented scene. The results of this work were evaluated visually, thus comparing the methods for their efficiency in the image classification of the study area and bringing subsidies to interested parties in classifiers for the coastal dune area.

Keywords — classification, DIP, dunes, SVM, supervised.

1. INTRODUÇÃO

A classificação de imagens leva em consideração as classes espectrais para extrair seus dados, mas dependendo da área de estudo essas divisões possuem valores de médias próximos entre si, necessitando de dimensões espectrais altas para separação apropriada dos dados ([1], 2009). Baseado nesta afirmação dentro das técnicas de Processamento Digital de Imagens (PDI) possuímos diversos métodos para classificação de imagens de satélites, como por exemplo a distância mínima, K-média, ISODATA e árvore de decisão ([2] et al, 2009).

Inseridos nesses métodos de classificação temos também o Support Vector Machine (SVM) e Máxima Verossimilhança Gaussiana (MVG), caracterizados como métodos supervisionados e paramétricos no de processamento de imagem deste trabalho. Os métodos supervisionados baseados na definição das classes e na coleta dos pixels de treinamento representativos de cada classe ([3], 2009), realizam cálculos a partir dessas amostras para serem referência nas demais classificações. Os métodos paramétricos baseados nas amostras de treinamento, pixels de valores conhecidos, estimam a densidade de probabilidade do comportamento dos dados das classes ([1], 2009), possuindo previamente seu valor.

Porém, é importante ressaltar que o Support Vector Machine (SVM) pode ser aplicado na classificação de imagens, como não supervisionado. O qual, une pela semelhança e separa pela diferença dos valores de cada pixel na imagem para discriminar as classes, não necessitando de pixels pré-definidos ([4], 2007). Também salientar que um método pode também ser considerado não paramétrico quando não é possível conhecermos a densidade de probabilidade do comportamento dos dados ([5], 2002).

A fim de comparar os semelhantes métodos, citados anteriormente, a área de estudo definida para este projeto foi a extensão de dunas costeiras da praia do Cassino, litoral do Rio Grande do Sul, pela sua função ambiental como Área de Preservação Permanente (APP), sendo assegurada pela Resolução do Conselho Nacional do Meio Ambiente ([6] n° 302/2002).

Sendo assim, os objetivos deste trabalho foram realizar a comparação dos métodos de classificação Support Vector Machine (SVM) e Máxima Verossimilhança Gaussiana (MVG) em uma imagem do satélite Geoeeye da extensão das dunas costeiras do Balneário Cassino, classificação da área de estudo com os métodos SVM e MVG, análise dos dados gerados, e a implementação da análise de acurácia baseada na matriz erro dos dados resultantes das classificações.

A fim de alcançar estes objetivos foram realizadas a segmentação pelo método de crescimento de regiões, a definição das classes para a classificação baseada no conhecimento da área de estudo e na análise visual da cena segmentada. Também, foram efetuados cálculos de textura para acrescentar atributos de diferenciação nas classes, sendo eles de desvio padrão e entropia. Isso possibilitou a coleta das amostras de treinamento, essa que é imprescindível para implementação dos métodos supervisionados comparados nesta pesquisa.

1.1. Área de Estudos

A área de estudo definida para realização desse trabalho é a extensão das dunas costeiras da praia do Cassino, localizada no balneário Cassino dentro do município de Rio Grande no estado do Rio Grande do Sul. As dunas costeiras são formadas pela interação dos ventos, areia e vegetação ([7], 2006), com a função ambiental de resguardo da costa, proteção dos lençóis freáticos e a preservação da fauna e flora. Sendo Áreas de Preservação Permanente (APP).

Essa área de preservação dentro da praia do Cassino caracteriza-se da mesma maneira, tendo em sua fauna o tuco-tuco, maçarico de colar, sapo das dunas, entre outros animais silvestres e na sua flora a catiporágua, margarida das dunas, capim das dunas, entre outras vegetações.

É importante salientar que houve a necessidade de realizar o recorte da área de estudo na imagem de alta resolução do satélite Geoeeye, pois não dispomos de recursos computacionais para manusear a Figura 1 (a), a qual mostra a extensão total da área de estudo presente na imagem Geoeeye, que apresenta a dimensão das dunas costeiras do final da localidade da Querência até o final da praia de encontro aos Molhes da Barra. Assim, optamos pela Figura 1 (b) que é o recorte da área de estudo da imagem Geoeeye para extensão do início da localidade do ABC no arroio do Gelo até o final da praia do Cassino.

Também, é relevante afirmar que as dunas costeiras são um patrimônio do município de Rio Grande. Sendo assim, as dunas da praia do Cassino foram escolhidas para realização dessa pesquisa e estão identificadas na Figura 2, elaborada no sistema de coordenadas geográficas com Datum SIRGAS 2000 e desenvolvido em setembro de 2017.

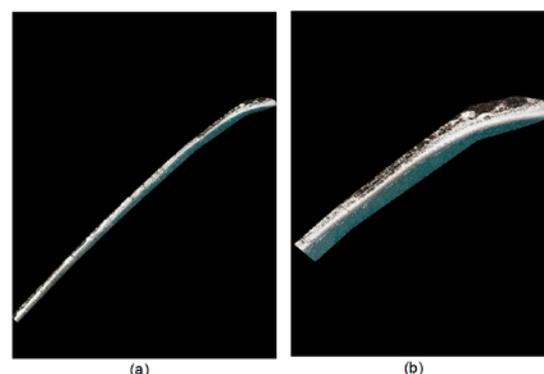


Figura 1. Imagem da área de estudo. Sensor Geoeeye (a) extensão total (b) recorte.

Ecosistema de dunas costeiras do Balneário Cassino

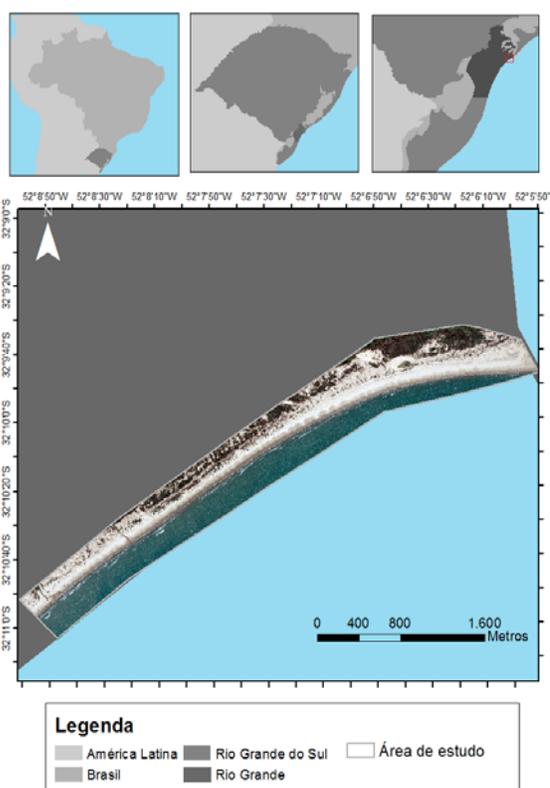


Figura 2. Área de estudo.

2. MATERIAIS E MÉTODOS

O trabalho foi desenvolvido com a utilização de uma imagem de alta resolução espacial Geoeeye, adquirida em 5 de dezembro de 2015, via projeto do curso técnico de Geoprocessamento do Instituto Federal do Rio Grande do Sul em colaboração com a Prefeitura Municipal do Rio Grande, a qual possui 4 bandas espectrais e com o software MatLab, e a partir de classes pré-definidas, sendo elas: corpos d'água

1, corpos d'água 2, corpos d'água 3, sedimentos 1, sedimentos 2, sedimentos 3, vegetação 1 e vegetação 2, que são referentes, respectivamente, a arroio, mar, corpos d'água entre a vegetação, sedimentos a beira mar, sedimentos entre a beira mar e as dunas, sedimentos do mar, vegetação clara e vegetação escura.

As amostras de treinamento foram coletadas sobre a imagem Geoeeye para treinar os algoritmos do Support Vector Machines (SVM) e Máxima Verossimilhança Gaussiana (MVG) para serem implementados na cena Geoeeye segmentada por crescimento de regiões. Em que, o SVM utilizou o kernel Radial Basis Function (RBF), enquanto o MVG usou o limiar segundo análise de cada um dos pixels.

A fim de verificar os resultados obtidos nas classificações dos métodos SVM e MVG, foi realizada uma saída de campo no dia 13 de novembro de 2017 com o objetivo de melhor identificar as classes na área de dunas costeiras da praia do Cassino, e também foi realizada a análise da matriz erro que representou a uniformidade entre os resultados das classificações possibilitando o cálculo dos erros de omissão e comissão de cada método de classificação.

3. RESULTADOS

Os resultados deste trabalho foram à segmentação por crescimento de regiões na imagem Geoeeye da área de estudo, (figura 3), a qual teve por objetivo dividir a imagem em regiões ou objetos para facilitar a sua análise, a classificação da segmentação pelo método de classificação Support Vector Machine (SVM), e a classificação da mesma cena pelo método Máxima Verossimilhança Gaussiana (MVG).

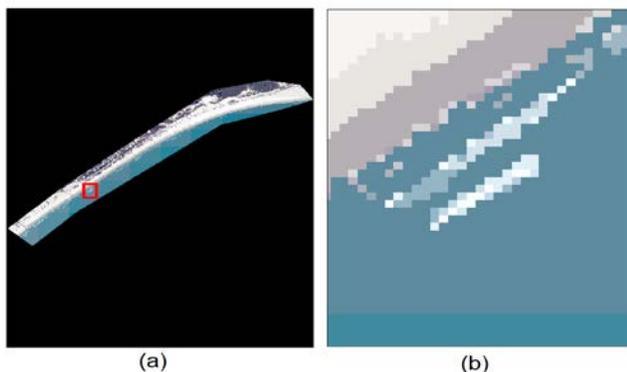


Figura 3. Imagem da segmentação área de estudo da imagem Geoeeye (a) extensão total (b) recorte.

Na implementação do algoritmo se optou pelo Kernel *default* do software *MatLab* o *Radial Basis Function* (RBF) que é comumente usado na classificação com método SVM. O RBF diminui as distâncias entre amostras, assim variando entre zero e um nos espaços abstratos, ou seja, difere se sim ou não é daquela classe o *pixel* analisado. É essencial ressaltar o diferencial do SVM, que consiste em minimizar expectativa

de erros, sendo assim através do risco empírico é definida a média da taxa de erro nas amostras de treinamento.

4. DISCUSSÃO

Na análise visual dos resultados das classificações foi possível observar a confusão do classificador SVM entre as classes corpos d'água 2 e sedimentos 1 com sedimentos 2, essas confusões que podem ser esclarecidas pela segmentação ter sido dividida em retângulos, desse modo segmentos de mesma classe ficaram com padrões de cores diferentes. Enquanto, o classificador MVG teve maior confusão entre corpos d'água 1, sedimentos 1 e vegetação 1 com sedimento 2, essas confusões que também podem ser atribuídas à semelhança de cor entre os pixels de treinamentos coletados nas classes. Em comparação aos resultados dos métodos SVM e MGV, encontramos erros similares de classificação, como a confusão da classificação entre corpos d'água 1 e sedimentos 3, que acontece pela diferença de cor da classe sedimentos 3 na área em questão dentro da segmentação provocando o erro nos dois métodos (figura 4).

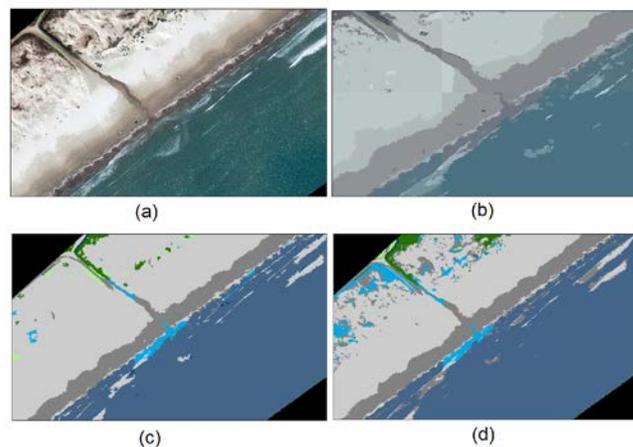


Figura 4. Confusão entre as classes corpos d' água 1 com sedimentos 3 (a) imagem Geoeeye (b) segmentação da imagem (c) classificação SVM (d) classificação MVG da segmentação.

Outro equívoco dos classificadores Support Vector Machine (SVM) e Máxima Verossimilhança Gaussiana (MVG) na classificação foi um local que na imagem original e na segmentada representava a vegetação 1, mas cada um dos métodos optou em classificar relacionado a outra classe, em que o SVM classificou como sedimentos 2 e o MVG como sedimentos 3, isso ocorreu pela imagem de coleta possuir resquícios de sedimentos, assim cada classificador optou por designar, essa vegetação para uma das classes de sedimentos que acreditou ser mais correta (figura 5).

Também, através da saída de campo foi possível avaliar os resultados das classificações dos dois métodos, sendo uma forma de observar, como estão dispostas em campo cada classe e compreender a dificuldade que os classificadores encontram pela homogeneidade das áreas. Analisando cada

uma das matrizes erro percebemos que o método de classificação Support Vector Machine obteve menos erros de omissão e comissão em seu resultado, sendo assim sua matriz é mais acurada. Enquanto o método MVG apresentou grande número de erros em sua matriz, que pode ser explicado pela variância das classes serem similares, o que confundiu o método ao definir sua função decisão pela ponderação em relação às médias (figuras 6 e 7).

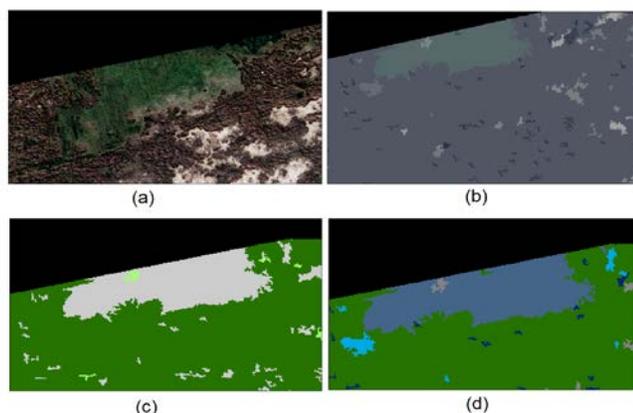


Figura 5. Confusão de classificação (a) imagem Geoeye (b) segmentação da imagem (c) classificação SVM (d) classificação MVG da segmentação.

| Classes | Classes | | | | | | | | | TOTAL |
|---------------|---------------|------|-------|------------|--------|--------|-----------|-------|-------|---------------|
| | Corpos d'água | | | Sedimentos | | | Vegetação | | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | |
| Corpos d'água | 1 | 4305 | 0 | 0 | 14 | 2 | 0 | 0 | 0 | 4321 |
| | 2 | 0 | 54927 | 0 | 0 | 0 | 2991 | 0 | 0 | 57918 |
| | 3 | 0 | 0 | 1024 | 0 | 0 | 0 | 0 | 0 | 1024 |
| Sedimentos | 1 | 345 | 0 | 0 | 155122 | 122 | 1 | 0 | 0 | 155590 |
| | 2 | 3 | 0 | 0 | 442 | 469743 | 0 | 7 | 78 | 470273 |
| | 3 | 0 | 6 | 0 | 1 | 0 | 151438 | 0 | 0 | 151445 |
| Vegetação | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 16731 | 0 | 16731 |
| | 2 | 2 | 0 | 0 | 0 | 17 | 0 | 1 | 73499 | 73519 |
| TOTAL | | 4655 | 54933 | 1024 | 155579 | 469884 | 154430 | 16739 | 73577 | 930821 |

Figura 6. Matriz erro da classificação com o método SVM.

| Classes | Classes | | | | | | | | | TOTAL |
|---------------|---------------|------|-------|------------|--------|--------|-----------|-------|-------|---------------|
| | Corpos d'água | | | Sedimentos | | | Vegetação | | | |
| | 1 | 2 | 3 | 1 | 2 | 3 | 1 | 2 | 3 | |
| Corpos d'água | 1 | 2491 | 898 | 0 | 12122 | 5658 | 7525 | 4323 | 9177 | 42394 |
| | 2 | 0 | 42593 | 0 | 0 | 0 | 48485 | 1401 | 80 | 92559 |
| | 3 | 64 | 1 | 981 | 57 | 34 | 1 | 298 | 4889 | 6325 |
| Sedimentos | 1 | 481 | 395 | 0 | 129406 | 30547 | 2566 | 844 | 388 | 164827 |
| | 2 | 5 | 27 | 0 | 13929 | 431728 | 140 | 142 | 17 | 445988 |
| | 3 | 210 | 10713 | 0 | 2 | 6 | 95475 | 1339 | 650 | 108395 |
| Vegetação | 1 | 229 | 257 | 0 | 13 | 351 | 22 | 1070 | 3481 | 5423 |
| | 2 | 1175 | 49 | 43 | 50 | 1360 | 216 | 7322 | 54895 | 65110 |
| TOTAL | | 4855 | 54933 | 1024 | 155579 | 469884 | 154430 | 16739 | 73577 | 930821 |

Figura 7. Matriz erro da classificação com o método MVG.

5. CONCLUSÕES

Ao longo da presente pesquisa foi realizada a comparação entre os métodos de classificação Support Vector Machine (SVM) e Máxima Verossimilhança Gaussiana (MVG) na área de dunas costeiras do Balneário Cassino. Ambos os métodos atingiram os resultados esperados que foi classificar as dunas. Porém, baseado nos resultados obtidos de cada método é visível que para escolher qual deles é mais adequado para a área de dunas costeiras, depende diretamente do objetivo do estudo.

Visto que o classificador Support Vector Machine (SVM) obteve o melhor resultado pela matriz erro alcançando menos

erros na classificação, esse método se torna uma melhor escolha quando o objetivo do estudo é a realização da detecção de mudanças, pois sua classificação possui áreas mais definidas para todas as classes, assim possibilitando analisar as áreas de mudanças positivas e negativas nas dunas costeiras.

Já o classificador Máxima Verossimilhança Gaussiana (MVG) não alcançou bons resultados na sua matriz erro possuindo mais erros de confusão entre as classes. Mas ainda é uma boa escolha quando o objetivo do trabalho é a comparação da imagem classificada com a imagem original das dunas costeiras, pois esse método é mais autêntico nas formas das classes, sendo sua classificação mais semelhante com a imagem original.

Também, é importante ressaltar que a imagem de alta resolução escolhida, acabou trazendo dificuldades para aplicação da metodologia, pois apresenta grande quantidade de pixels, assim necessitando de um computador adequado para o seu processamento o que, entretanto, não evitou a utilização de um tempo considerável para a sua execução.

6. REFERÊNCIAS

- [1] Andreola, R. "Support Vector Machines na classificação de imagens hiperespectrais". Porto Alegre, UFRGS, 2009. Dissertação – Programa de Pós-Graduação em Sensoriamento Remoto do Centro Estadual de Pesquisas em Sensoriamento Remoto e Meteorologia, UFRGS.
- [2] Nascimento, R. F. F; Alcântara, E. H; Kampel, M; Stech, J. L; Novo, E. M. L. M; Fonseca, L. M. G. "O algoritmo SVM: avaliação da separação ótima de classes em imagens CCD-CBERS-2". Natal, XIV Simpósio Brasileiro de Sensoriamento Remoto, p. 2079-2086, 2009.
- [3] Mather, P.; Tso, Bt. "Classification Methods for Remotely Sensed Data". 2. ed. London, CRC Press, 2009.
- [4] Lorena, A.; Carvalho, A. "Uma introdução às Support Vector Machines". Revista de informática teórica e aplicada. São Paulo, v. 14, n. 2, p. 43-67, 2007.
- [5] Semolini, R. "Support Vector Machines, Inferência Transdutiva e o Problema de Classificação". Dissertação (Mestrado em Engenharia de Computação) - UNICAMP. Campinas, SP, 2002.
- [6] CONAMA. Resolução n° 302, de 20 de março de 2002.
- [7] NEMA - Núcleo de Educação e Monitoramento Ambiental. "Plano de Manejo das Dunas Costeiras do Município de Rio Grande". Rio Grande, 2006.