



**11ª Jornada Científica e
Tecnológica do IFSULDEMINAS**

**& 8º Simpósio de
Pós-Graduação**

DESENVOLVIMENTO DE FERRAMENTA PARA AVALIAR A QUALIDADE DAS CLASSIFICAÇÕES DE IMAGENS MULTIESPECTRAIS E HIPERESPECTRAIS

Rodrigo C. RIBEIRO¹; Camila S. dos ANJOS²

RESUMO

No decorrer das últimas décadas tem crescido o uso dos sensores imageadores, que registram informações sobre os objetos presentes na superfície terrestre. Tal imageamento tem o objetivo de monitorar continuamente o meio ambiente e os recursos naturais. Inúmeros processamentos podem ser realizados nas imagens, dependendo da finalidade do trabalho requerido, e um dos métodos de processamento é a classificação de imagens, que busca identificar os materiais encontrados na superfície da terra. Todas as classificações apresentam erro, e o presente trabalho tem como finalidade a implementação de um algoritmo que apresenta as controversas entre verdade de campo e a classificação. Para avaliar a qualidade do resultado do processo de classificação de imagens foram implementados os índices de acurácia global, kappa, erro de omissão e inclusão, que são essenciais para validar a confiabilidade das imagens classificadas.

Palavras-chave: Matriz Confusão; Acurácia; Índice de validação.

1. INTRODUÇÃO

O Sensoriamento Remoto é explicado como sendo um conjunto de processos que possibilita a aquisição de informações dos alvos e objetos que representa a superfície terrestre sem o contato com mesmo (MORAES, 2002).

A matriz de confusão é uma das maneiras de demonstrar as controversas entre a verdade de campo e o resultado do processo de classificação, onde as linhas representam à classe correta (Verdade de Campo) e as colunas à classe obtida na classificação. Existem dois erros, os de omissão que é em relação à verdade de campo, ou seja, são pixels pertencentes a uma determinada classe que foi classificada em outra classe, e os erros de inclusão que é em relação à classificação, ou seja, correspondem aos pixels que sendo de uma classe, foram incluídas em outra. Assim podendo avaliar quanto uma classe foi subestimada ou superestimada, e ainda obter a acurácia do ponto de vista do usuário (Classificação) e do produtor (Verdade de Campo) (CENTENO, 2003).

Um dos procedimentos usado para mensurar a exatidão é o índice Kappa. Sendo recomendado como uma medida apropriada da exatidão por representar inteiramente a matriz de confusão, tomando todos os elementos da matriz e não só a diagonal principal, como outros índices (MOREIRA, 2001).

1 Bolsista PIBIC/CNPq, IFSULDEMINAS – Campus Inconfidentes. E-mail: rodrigo-ribeiro14@hotmail.com.

2 Orientador, IFSULDEMINAS – Campus Inconfidentes. E-mail: camila.lacerda@ifsuldeminas.edu.br.

2. MATERIAL E MÉTODOS

A ferramenta foi implementada no *software FreeMat*, onde apresenta ao usuário a precisão de sua classificação, posteriormente foram criadas duas matrizes de tamanho 6x7 muito parecidas, contendo cinco classes, onde a primeira é a suposição de uma verdade de campo e a segunda é a suposição de uma classificação, onde foi calculado no *software LibreOffice Calc* todos os índices e resultados para validar o algoritmo que foi criado. A verdade de campo do município de Inconfidentes foi adquirida e a classificação da cena foi gerada, para fazer a comparação entre uma validação por amostras aleatórias e uma validação pixel a pixel.

Primeiramente o algoritmo busca as duas imagens (Matrizes) sendo a verdade de campo e a classificação que se deseja validar. Logo em seguida o código verifica o tamanho da imagem inserida, sendo de suma importância para todo o projeto. Então a verdade de campo passa por um *looping* para buscar o número de classes, sendo este valor usado para criar o tamanho da matriz confusão, com todas as posições da mesma com o número zero.

Para a realização da validação pixel a pixel foi implementada uma rotina para construir uma matriz confusão, verificando cada pixel da verdade de campo com a classificação, onde o valor encontrado em cada posição da verdade de campo e valor encontrado na mesma posição na classificação irá ser respectivamente a posição de linha e coluna da matriz confusão que somará o número um nesta posição da matriz confusão, sendo que o *looping* percorre as imagens da esquerda para a direita e cima para baixo. O caso o usuário deseje realizar a validação por amostras aleatórias, o código passará por uma rotina para criar a matriz confusão do mesmo modo da rotina anterior, mas em vez de percorrer as duas imagens totalmente, irá buscar posições das imagens randomicamente, dependendo do número de amostras que o usuário deseja.

De acordo com Centeno (2003) após a construção da matriz confusão, a acurácia global e o índice kappa podem ser obtidos pela seguinte expressão:

$$K = \frac{\theta_1 - \theta_2}{1 - \theta_1} \quad (1)$$

Sendo:

$$\theta_1 = \frac{\sum_{i=1}^c x_{ii}}{n} \quad \theta_2 = \frac{\sum_{i=1}^c x_{i!} * x_{!i}}{n}$$

$x_{i!}$ = Somatório parcial da linha

$x_{!i}$ = Somatório parcial da coluna

c = Tamanho da matriz confusão

n = Total de pixels que entrou na matriz

θ_1 = Acurácia Global

K = Índice kappa

Ainda sobre o mesmo autor para obter a acurácia do ponto de vista do usuário é dado pelo valor da diagonal da classe dividido pelo somatório parcial da coluna da matriz confusão da mesma classe, a acurácia do ponto de vista do produtor é dado pelo valor da diagonal da classe dividido pelo somatório parcial da linha da matriz confusão da mesma classe. O erro de omissão é obtido dividindo a diferença entre a soma parcial da linha e o valor da diagonal pela soma parcial da linha da mesma classe, o erro de inclusão é obtido dividindo a diferença entre a soma parcial da coluna e o valor da diagonal pela soma parcial da coluna da mesma classe. Foi implementado ainda uma rotina comparando se o somatório parcial da coluna da matriz confusão for maior que o somatório parcial da linha de cada classe, o classificador superestimou a classe, ou se for menor o classificador subestimou a classe.

Após o código implementado, foram realizadas cinco validações com 10%, 20% e 30% de amostras aleatórias de uma mesma classificação com o objetivo de demonstrar a aleatoriedade das amostras, que melhor representa a classificação em comparação com a validação pixel a pixel.

3. RESULTADOS E DISCUSSÕES

Tanto na acurácia global quanto no índice kappa, quanto maior o percentual de amostras aleatórias, menor o desvio padrão entre as cinco validações de uma mesma porcentagem, além da diferença entre a validação completa (pixel a pixel) com a validação por amostras aleatórias da mesma classificação diminuir gradativamente, como mostra a última linha da tabela 1.

Tabela 1: Comparação de validação entre porcentagens de amostras aleatórias.

	10%		20%		30%	
	Acurácia Global	Índice Kappa	Acurácia Global	Índice Kappa	Acurácia Global	Índice Kappa
1	33,04	26,22	33,01	26,19	32,99	26,15
2	32,98	26,15	33,13	26,28	33,01	26,20
3	33,22	26,39	32,89	26,05	32,90	26,07
4	32,85	26,04	32,88	26,06	33,03	26,22
5	33,03	26,24	32,93	26,12	33,08	26,23
Média	33,02	26,21	32,96	26,14	33,00	26,17
Desvio Padrão	0,13	0,13	0,10	0,10	0,07	0,07
Δ pixel a pixel	0,03	0,04	0,02	0,03	0,01	0,00

Fonte: Autoral.

Usando as duas matrizes 6x7 criadas na metodologia, obteve-se como resultado a validação da implementação do código, que mostra a matriz confusão e todos os produtos de interesse para o usuário (Quadro 1).

Quadro 1: Resultados da implementação do código.

		Classificação					Somatório das Linhas	Acurácia do Produtor (%)	Erro de Omissão (%)
		Classe 1	Classe 2	Classe 3	Classe 4	Classe 5			
Verdade de campo	Classe 1	9	0	0	0	0	9	100,00	0,00
	Classe 2	2	6	0	1	0	9	66,67	33,33
	Classe 3	0	2	6	1	0	9	66,67	33,33
	Classe 4	0	0	0	7	2	9	77,78	22,22
	Classe 5	0	1	0	0	5	6	83,33	16,67
	Somatório das Colunas	11	9	6	9	7	Acurácia Global = 78,57 % Índice Kappa = 73,13 %		
	Acurácia do Usuário (%)	81,82	66,67	100,00	77,78	71,43			
	Erro de Inclusão (%)	18,18	33,33	0,00	22,22	28,57			

Fonte: Autoral.

Ainda sobre o quadro 1, a análise demonstra que a classe 1 foi superestimada, a 2 balanceada, a 3 foi subestimada, a 4 balanceada e a classe 5 superestimada, comparando os somatórios parciais das linhas e colunas.

4. CONCLUSÕES

Conclui-se que para uma validação por amostras aleatórias, quanto maior o percentual de amostras mais próxima de uma validação pixel a pixel será a mesma classificação. Ressalta-se a coleta de amostras aleatórias sem a intervenção do usuário na decisão de quais amostras se deseja coletar, proporciona uma validação mais realista sem a inserção de erros pelo usuário.

O código foi implementado de maneira simples e eficiente, onde qualquer usuário possa realizar a validação de suas classificações de imagem de maneira precisa.

AGRADECIMENTOS

Agradeço pelo apoio ao Instituto Federal de Educação, Ciência e Tecnologia do Sul de Minas Gerais – Campus Inconfidentes pela oportunidade da pesquisa.

REFERÊNCIAS

CENTENO, J A. S.. **Sensoriamento Remoto e Processamento de Imagens Digitais**. Curitiba: Departamento de Geomática - Ufpr, 2003. 219 p.

MOREIRA, M. A., **Fundamentos de Sensoriamento Remoto e Metodologias de Aplicação**. São José dos Campos: INPE, 2001.

MORAES, E. C.. **Fundamentos de Sensoriamento Remoto**. São José do Campos: Instituto Nacional de Pesquisas Espaciais-INPE, 2002.