

USO DE GRAUS DE CONFIANÇA DAS CLASSES EM CLASSIFICADORES BAYESIANOS

ORLANDO ALVES MÁXIMO^{1,2}
DAVID FERNANDES²

¹Instituto de Estudos Avançados (IEAv) – Centro Técnico Aeroespacial (CTA)
Rodovia dos Tamoios Km 5,5 – 12228-840 – São José dos Campos – SP, Brasil
Tel.: (12) 3947-5353, Fax: (12) 3944-1445
maximo@ieav.cta.br

²Instituto Tecnológico de Aeronáutica (ITA) – Centro Técnico Aeroespacial (CTA)
Praça Mal. Eduardo Gomes, 50 – 12228-900 - São José dos Campos - SP, Brasil
Tel.: (12) 3947-6807, Fax: (12) 3947-5878
{david, maximo}@ele.ita.br

Abstract. This paper evaluates the class reliability factors influence in a multidimensional Bayesian classifier. To evaluate the best reliability factors values, an accuracy index based on five well known accuracy coefficients and a data set reliability factor based on the class reliability factors were used. The results for a two-dimensional classifier show that there are several combinations of class reliability factors that achieve the best results. These combinations, once converted to data set reliability factors show that there is a linear relation between the best choices of reliability factors. This is consistent to previous works.

Keywords: Classification, reliability factors, segmentation, bayesian classifier

1. Introdução

Nas últimas décadas, o número de plataformas operando sistemas de sensoriamento remoto, disponibilizando seus produtos em bases comerciais ou não, cresceu bastante. Muitos destes sistemas possuem características multi ou hiperespectrais, multifrequência ou, até mesmo, multipolarização, o que fez aumentar significativamente a quantidade de dados disponíveis.

Tal quantidade de dados, por sua vez, fez crescer a necessidade do uso do computador para sua manipulação, e os processos de fusão e classificação assistidos por computador têm sido bastante utilizados. Diversas abordagens podem ser seguidas neste processo. Mauris e Bolon (2001) apresentam uma compilação de trabalhos nesta área.

Na maioria das vezes, estas abordagens consideram todos os conjuntos de dados disponíveis da mesma forma e todos têm o mesmo impacto no resultado final, independente de sua qualidade. Recentemente, alguns pesquisadores introduziram um novo conceito, denominado “grau de confiança”, que tem como objetivo controlar a influência de cada conjunto de dados (imagens) no processo de fusão e/ou classificação (Benediktsson et al., 1990 ; Jeon e Landgrebe, 1999, por exemplo). Esta nova abordagem, porém, não leva em consideração a influência da qualidade das classes presentes nas imagens.

Este trabalho tem por objetivo apresentar uma extensão do conceito de grau de confiança das imagens, avaliando a influência do grau de confiança das classes num processo de classificação com abordagem bayesiana. Os resultados para um caso bi-dimensional mostraram que não há uma única combinação de graus de confiança que maximizam a performance do classificador. Combinações mal escolhidas impactaram a precisão do classificador em até 4%, para o caso estudado. Além disso, a utilização de um estimador para a obtenção do grau de confiança da imagem a partir do grau de confiança das classes, baseado na média dos graus de confiança, preservou as relações entre os graus de confiança das

imagens. Tais resultados são consistentes com análises recentes da influência do grau de confiança das imagens no resultado final de um classificador Bayesiano.

2. Classificadores Bayesianos e a Função de Adesão

Num processo de classificação, diversas estratégias podem ser seguidas para a tomada de decisão. Uma destas estratégias é baseada na minimização do Risco de Bayes para teste de hipóteses. Segundo Scharf (1991), o Risco de Bayes é definido como uma média da Função de Perda. A Função de Perda, \tilde{e}_{ij} , corresponde ao custo de classificar uma região como pertencente à classe j , quando, na realidade, ela pertence à classe i .

Considerando a existência de uma observação, a estratégia de decisão para a classificação consiste em minimizar a Perda Condicional Média. Neste caso, Swain (1978) define a Perda Condicional Média como:

$$L_X(w_j) = \sum_{i=1}^m \mathbf{I}_{ij} p(w_j | X), \quad (1)$$

onde $\{w_j, j = 1, \dots, m\}$ é o conjunto de m classes; X é a observação (que é a realização de um processo estocástico no qual está presente uma das classes w_j) e $p(w_j | X)$ é a probabilidade condicional de a classe ser w_j , dada a observação X .

Definindo o custo \tilde{e}_{ij} como:

$$\mathbf{I}_{ij} = \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases}, \quad (2)$$

então (1) torna-se:

$$L_X(w_j) = 1 - p(w_j | X), \quad (3)$$

e $L_X(w_j)$ será mínimo, quando a probabilidade *a posteriori* $p(w_j | X)$ for máxima.

Desse modo, a regra de decisão pode ser expressa por:

$$X \in w_i \Leftrightarrow p(w_i | X) = \max_j p(w_j | X), \quad (4)$$

ou seja, para os custos dados em (2), a maior probabilidade *a posteriori* indica a classe a ser escolhida.

Partindo de (4), Lee et al. (1987) definem a Função de Adesão para um conjunto de n observações, ou seja $X = \{X_1, \dots, X_n\}$.

Desse modo:

$$p(w_j | X) = p(w_j | X_1, \dots, X_n). \quad (5)$$

Mas

$$p(w_j | X) = \frac{p(X | w_j)p(w_j)}{p(X)} = \frac{p(X_1, \dots, X_n | w_j)p(w_j)}{p(X_1, \dots, X_n)}.$$

Assumindo a condição de independência entre as observações, temos que:

$$p(w_j | X) = p(w_j) \prod_{k=1}^n \frac{p(X_k | w_j)}{p(X_k)},$$

mas, $p(X|w) = p(w|X)p(X)/p(w)$, então:

$$p(w_j | X) = p(w_j)^{1-n} \prod_{k=1}^n p(w_j | X_k) .$$

Desta forma, uma regra decisão alternativa pode ser dada por:

$$X \in w_i \Leftrightarrow p(w_i)^{1-n} \prod_{k=1}^n p(w_i | X_k) = \max_j \left[p(w_j)^{1-n} \prod_{k=1}^n p(w_j | X_k) \right] . \quad (6)$$

Lee et al. (1987) chamam o primeiro termo da igualdade em (6) de Função de Adesão. A regra de decisão torna-se então:

$$X \in w_i \Leftrightarrow F_i(X) = \max_j F_j(X) , \quad (7)$$

onde $F_j(X)$ é a função de adesão da classe j , dado o conjunto de observações X e pode ser reescrita da forma:

$$F_j(X) = p(w_j) \prod_{k=1}^n \frac{p(w_j | X_k)}{p(w_j)} \quad (8)$$

3. Graus de confiança das observações

O classificador da função de adesão definido em (6) considera todas as observações com o mesmo peso no resultado final, independente da qualidade das mesmas. Tal consideração pode prejudicar o desempenho dos classificadores que operam com observações “não-confiáveis”. Para solucionar esta situação, pode ser utilizado um sistema de graus de confiança, capaz de aumentar a influência das fontes mais confiáveis e diminuir a influência das fontes não-confiáveis, melhorando o desempenho do classificador.

Uma boa razão para a adoção de um sistema de graus de confiança torna-se clara pelo fato de o classificador da Função de Adesão ser, essencialmente, um produtório de probabilidades relacionadas com as observações disponíveis. Se qualquer uma das probabilidades gerar valores próximos de zero, o resultado final será, também, próximo de zero, mascarando a influência das outras fontes mais confiáveis, mesmo que esta observação com probabilidade próxima de zero seja uma observação não-confiável.

Os graus de confiança devem atuar de maneira tal que quando a confiança de uma observação é baixa, o efeito desta observação no resultado final seja diminuído; e quando a confiança de uma observação é alta, seu efeito é realçado. Uma maneira de conseguir tal efeito é a introdução de graus de confiança na forma de expoentes.

Benediktsson et al. (1990) analisaram a influência de cada observação dentro da Função de Adesão. Para isso, calcularam a razão entre a Função de Adesão para $X = \{X_1, \dots, X_{n+1}\}$ e $X = \{X_1, \dots, X_n\}$ obtendo:

$$\frac{F_j(X_{1,L}, X_{n+1})}{F_j(X_{1,L}, X_n)} = \frac{p(w_j | X_{n+1})}{p(w_j)} . \quad (9)$$

De posse desta razão, os autores chegaram à conclusão de que a Função de Adesão definida em (8) representa melhor a influência de cada observação no resultado final do classificador.

Assim, para controlar a influência de cada observação dentro do processo de classificação, graus de confiança a_i são definidos como expoentes das contribuições

individuais de cada observação, criando a Função de Adesão Modificada, $F_j'(X)$, que é dada por:

$$F_j'(X) = p(w_j) \prod_{i=1}^n \left[\frac{p(w_j | X_i)}{p(w_j)} \right]^{\alpha_i}, \quad (10)$$

onde os α_i 's são selecionados no intervalo [0,1]. Assim, se uma fonte é totalmente não-confiável, seu grau de confiança é zero e ela não terá nenhuma influência no resultado final da classificação, independente do valor de $p(w_i|X_i)$. No extremo oposto, caso a fonte seja totalmente confiável, seu grau de confiança será 1 e sua influência no resultado final da classificação será totalmente preservada.

Utilizando o Teorema de Bayes, (10) pode ser reescrita em função de $p(X_i|w_j)$ como:

$$F_j'(X) = p(w_j) \prod_{i=1}^n \left[\frac{p(X_i | w_j)}{p(X_i)} \right]^{\alpha_i} \quad (11)$$

Como o termo $\prod_{i=1}^n [p(X_i)]^{\alpha_i}$ está presente em todos $F_j'(X)$, ele não influi no processo de decisão e $F_j'(X)$ pode ser escrita como:

$$F_j'(X) = p(w_j) \prod_{i=1}^n [p(X_i | w_j)]^{\alpha_i}. \quad (12)$$

Máximo e Fernandes (2002) testaram a influência do grau de confiança da observação na precisão de um classificador baseado na Função de Adesão Modificada, utilizando um par de imagens SAR (bandas X e L), com diversas janelas de filtragem. Foi constatado que a utilização dos graus de confiança apresenta um ganho de eficiência em relação ao Classificador de Máxima Verossimilhança e que existe uma relação linear entre os graus de confiança das duas imagens que propiciam os melhores resultados de classificação. A **Figura 1** apresenta a relação entre os graus de confiança que obtiveram os melhores resultados de classificação para as janelas de normalização estudadas.

4. Graus de confiança das classes

Os graus de confiança das observações, no entanto, são influenciados pela qualidade dos dados das classes presentes nas mesmas. Deste modo, é necessário avaliar a influência da qualidade, ou do grau de confiança, das classes na composição do grau de confiança das observações.

Seja \mathbf{a}_{ij} o grau de confiança da classe j na imagem i , com $i = 1, 2, \dots, n$; $e j = 1, 2, \dots, m$. Consideremos as seguintes premissas básicas: 1) Tanto o grau de confiança das classes (\mathbf{a}_{ij}) quanto o grau de confiança das observações (\mathbf{a}_i) devem estar contidos no intervalo [0,1]; 2) $\mathbf{a}_i = f(\mathbf{a}_{ij})$ com $j = 1, 2, \dots, m$; 3) Se $\mathbf{a}_{i1} = \dots = \mathbf{a}_{im} \Rightarrow \mathbf{a}_i = \mathbf{a}_{i1}$; 4) $\mathbf{a}_i = 0 \Leftrightarrow \mathbf{a}_{i1} = \dots = \mathbf{a}_{im} = 0$; 5) $\mathbf{a}_i = 1 \Leftrightarrow \mathbf{a}_{i1} = \dots = \mathbf{a}_{im} = 1$; 6) \mathbf{a}_i deve refletir as variações de qualquer \mathbf{a}_{ij} ($j = 1, 2, \dots, m$) numa razão direta de proporcionalidade.

Um estimador de \mathbf{a}_i que segue todas as premissas estabelecidas pode ser dados por:

$$\mathbf{a}_i = \frac{1}{m} \sum_{j=1}^m \mathbf{a}_{ij}. \quad (13)$$

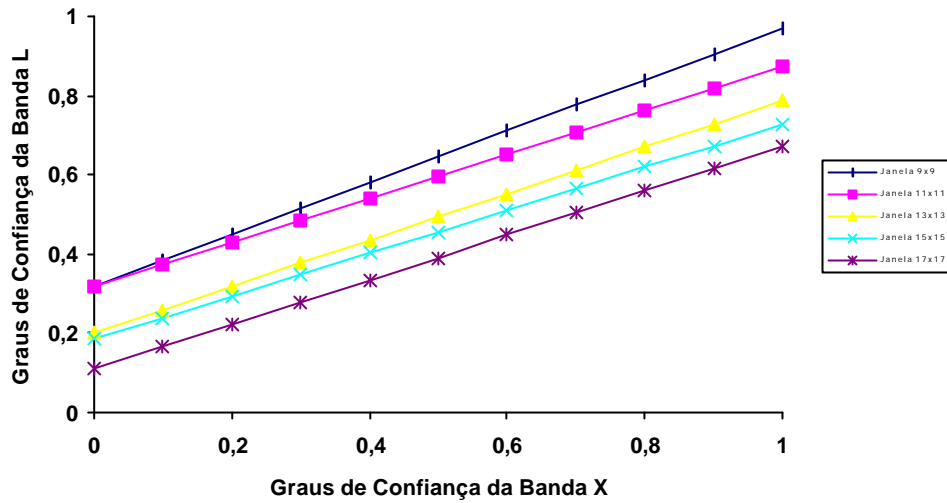


Figura 1 – Relação entre grau de confiança que proporcionam a melhor precisão do classificador baseado na Função de Adesão. Adaptado de Máximo e Fernandes (2002).

Assim, inserindo o grau de confiança das classes presentes nas observações, (11) pode ser reescrita como:

$$F_j^i(X) = p(w_j) \prod_{i=1}^n \left[\frac{p(w_j | X_i)}{p(w_j)} \right]^{a_{ij}} \quad (14)$$

ou então,

$$F_j^i(X) = p(w_j) \prod_{i=1}^n \left[\frac{p(X_i | w_j)}{p(X_i)} \right]^{a_{ij}} \quad (15)$$

onde $p(X_i) = \sum_{j=1}^m p(w_j) p(X_i | w_j)$.

5. Descrição do teste de desempenho

Para a avaliação da influência do grau de confiança, foi realizado um teste em duas etapas. Na primeira, um par de imagens do SAR580 (bandas X e L) de uma região da Alemanha foi classificado utilizando a Função de Adesão Modificada definida em (15), com uma janela de filtragem de 15x15. Nesta etapa, foram definidas quatro classes em cada imagem e três valores possíveis de grau de confiança para cada uma delas (0,3, 0,6 e 0,9). Foram utilizadas todas as combinações de graus de confiança, totalizando 6561 classificações.

A precisão de cada uma destas classificações foi avaliada utilizando-se o Índice de Precisão, $D(X)$, definido por Máximo e Fernandes (2002) como sendo: $D(X) = 1 - |P - P(X)| / \sqrt{5}$. Onde $P(X)$ é um vetor de classificação composto pelos seguintes coeficientes de acurácia: Precisão Global, Precisão Média por Classes, Coeficiente Kappa e dois coeficientes baseados na informação de Kullback-Leibler (as formulações para estes coeficientes podem ser encontradas em Nishii - 1999). $P = [1,1,1,1,1]$ é o vetor da classificação perfeita.

Após o cálculo do Índice de Precisão, foram selecionados aqueles que possuíam uma diferença máxima de 2% em relação ao maior Índice de Precisão obtido. Os valores dos graus

de confiança das classes utilizados nestas classificações foram, então, utilizados para o cálculo do grau de confiança das imagens, baseado na expressão (13).

Na segunda etapa, os resultados foram comparados com aqueles obtidos por Máximo e Fernandes (2002).

6. Resultados

A **Figura 2** mostra os resultados do cálculo de $D(X)$ para a primeira etapa. Nela, observa-se que as variações nos graus de confiança das classes podem provocar uma variação de até 4% na precisão do classificador. Com relação aos resultados da classificação (**Figura 3**), a variação dos graus de confiança das classes tem maior impacto nas fronteiras de classes com médias mais distantes, provocando um alargamento ou um estreitamento das mesmas.

A **Figura 4a** apresenta o resultado do cálculo do grau de confiança da imagem baseado na expressão (13), para as classificações que possuíam precisão com diferença máxima de 2% em relação ao maior Índice de Precisão obtido. Verifica-se que não existe uma única combinação de valores de graus de confiança das classes que maximiza o resultado do classificador.

Da análise da **Figura 4a**, verificou-se que poderia ser calculada a tendência dos dados através de uma regressão linear. A **Figura 4b** apresenta uma comparação entre esta regressão e a apresentada em Máximo e Fernandes (2002) para a mesma janela de filtragem. Verifica-se uma boa correlação entre os resultados, indicando que o cálculo dos graus de confiança das imagens baseado na média dos graus de confiança das classes preservou a tendência dos mesmos.

A reta representativa da tendência dos graus de confiança das imagens calculada a partir dos graus de confiança das classes apresenta uma inclinação que confere à Banda L maior confiança em relação à Banda X, quando comparada à reta de Máximo e Fernandes (2002). Neste trabalho os autores verificaram que, apesar da Banda L apresentar melhores resultados de classificação que a Banda X, quando usadas isoladamente, as combinações que apresentavam os melhores Índices de Precisão privilegiavam a Banda X (**Figura 4b**).

Esta nova inclinação, que aumenta a confiança na Banda L, reflete melhor os resultados das classificações individuais e pode indicar um caminho para a estimação dos graus de confiança a partir da imagem.

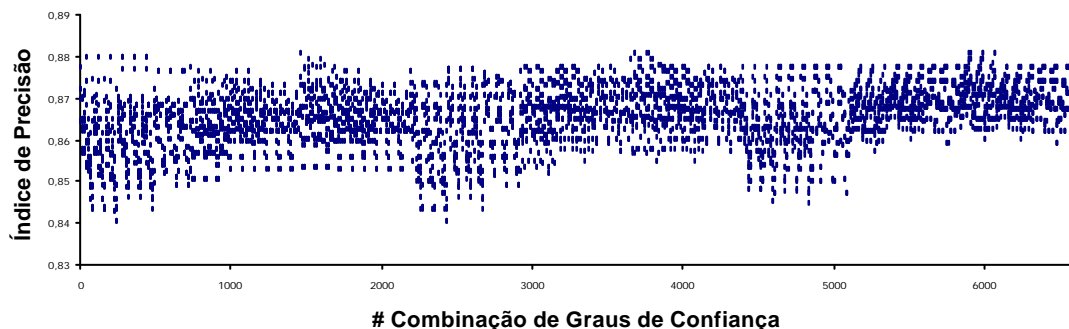


Figura 2 – Cálculo do Índice de Precisão da etapa 1.

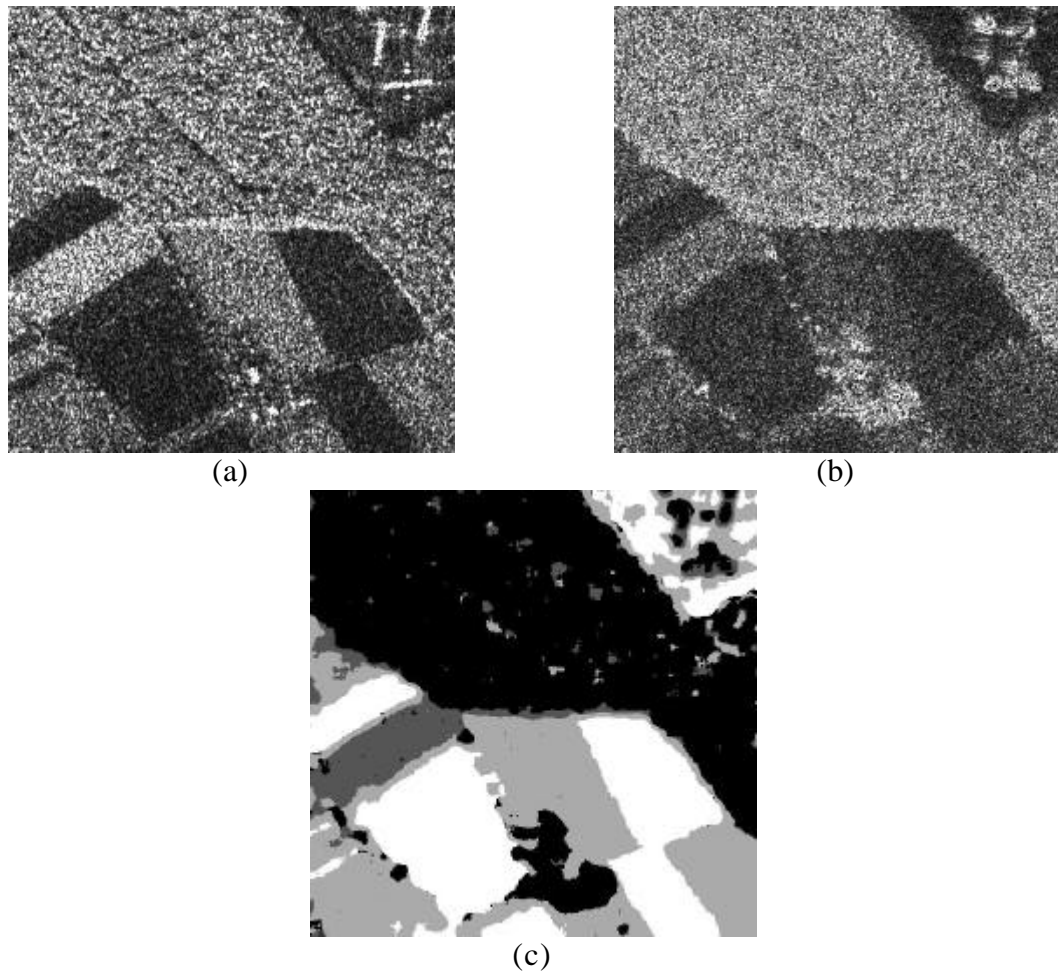


Figura 3 – a) Banda X original; b) Banda L original; c) Classificação utilizando graus de confiança para as classes: $\mathbf{a}_{11} = 0,3$; $\mathbf{a}_{12} = 0,3$; $\mathbf{a}_{13} = 0,6$; $\mathbf{a}_{14} = 0,9$; $\mathbf{a}_{21} = 0,6$; $\mathbf{a}_{22} = 0,6$; $\mathbf{a}_{23} = 0,3$; $\mathbf{a}_{24} = 0,9$ (obs.: As classes da Banda X receberam a numeração \mathbf{a}_{1*} e as classes da Banda L receberam a numeração \mathbf{a}_{2*}).

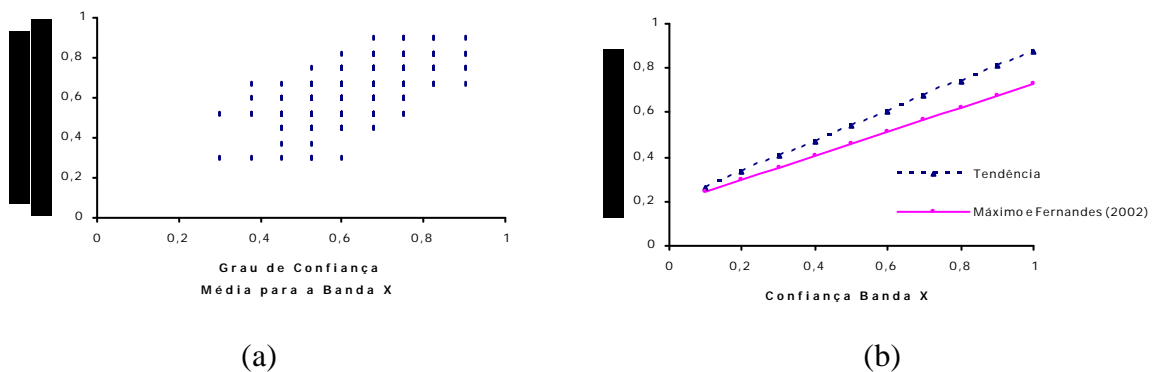


Figura 4 – a) Cálculo do grau de confiança da imagem; b) Comparação entre a tendência dos dados e os resultados obtidos com o grau de confiança das imagens.

7. Conclusão

Os resultados obtidos indicam que não há uma única combinação de graus de confiança das classes que maximizam o resultado do classificador.

O cálculo do grau de confiança da imagem baseado na média do grau de confiança das classes resultou numa relação linear entre os valores obtidos, de acordo com resultados obtidos anteriormente. Os valores obtidos também indicaram a necessidade de se aumentar os valores dos graus de confiança da Banda L em relação à Banda X, o que reflete melhor o fato da Banda L apresentar melhores resultados de classificação que a Banda X, quando usadas isoladamente.

Trabalhos futuros incluem o estudo da estimação dos graus de confiança das classes através de parâmetros da imagem, tais como: distância média entre classes, distância de Bhattacharyya etc.

8. Agradecimentos

Os autores agradem à Força Aérea Brasileira (FAB), à Fundação Aplicações de Tecnologias Críticas (ATECH) e à Comissão para Coordenação do Projeto do Sistema de Vigilância da Amazônia (CCSIVAM) pelo seu apoio na realização deste trabalho.

9. Referências

- Benediktsson, J.A.; Swain, P.H.; Ersoy, O.K. Neural network approaches versus statistical methods in classification of multisource remote sensing data, *IEEE Transactions on Geoscience and Remote Sensing*, v. 28, n. 4, p. 540-552, July 1990.
- Jeon B., Landgrebe D.A., "Decision fusion approach for multitemporal classification", *IEEE Transactions on Geoscience and Remote Sensing*, v. 37, p. 1227-1233, May 1999.
- Lee, T.; Richards, J.A.; Swain, P.H. Probabilistic and evidential approaches for multisource data analysis, *IEEE Transactions on Geoscience and Remote Sensing*, vol. GE-25, n. 3, p. 283-293, May 1987.
- Mauris L.V.G. and Bolon P.H., "A statistical overview of recent literature in information fusion", *IEEE MESS Sys. Mag.*, p. 7-14, March 2001.
- Máximo, O.A.; Fernandes, D. The use of data set reliability factors in a multidimensional Bayesian classifier. In: International Geoscience and Remote Sensing Symposium, Toronto, CA, June 2002. *Proceedings*. Publicação em CD.
- Nishii .R, Tanaka S., "Accuracy and inaccuracy assessment in land-cover classification", *IEEE Transactions on Geoscience and Remote Sensing*, v. 37, n. 1, p. 491-498, Jan. 1999.
- Scharf, L.L. *Statistical signal processing – detection, estimation, and time series analysis*, Addison-Wesley Publishing Company, Massachusetts, MA, 1991, 524 p.
- Swain, P.H. Bayesian classification in a time-varying environment, *IEEE Trans. Sys. Man. And Cyber.*, SMC-8 v. 12, p. 879-883, Dec. 1978.