

Uma nova abordagem na estimação de probabilidades *a priori* na classificação de imagens digitais em sensoriamento remoto

Celso Liczbinski^{1,2}
Vitor Haertel²

¹ Diretoria de Serviço Geográfico do Exército – DSG
1ª Divisão de Levantamento - 1ª DL
Rua Cleveland, 250, Menino Deus - 90850-240 - Porto Alegre - RS, Brasil
celso.liczbinski@terra.com.br

² Universidade Federal do Rio Grande do Sul - UFRGS
Centro Estadual de Pesquisa em Sensoriamento Remoto e Meteorologia - CEPSRM
Caixa Postal 15052 - 91501-970 - Porto Alegre - RS, Brasil
victor.haertel@ufrgs.br

Abstract. An adaptive classifier designed for high-dimensional image data classification has recently been proposed. The main feature in this proposed classifier consists in a novel method to acquire semi-labeled samples in an iterative way, during the classification process itself. The semi-labeled samples are then added to the set of available labeled samples, thus increasing the total number of training samples and consequently the reliability of the estimated values for the parameters in a parametric classifier. In this study we propose the insertion of an additional step into the iterative procedure implemented in the adaptive classifier, in order to produce more realistic estimates of the classes' *a priori* probability. The proposed procedure to estimate the classes' *a priori* probability makes use of the same weighted approach used by the adaptive classifier, is computationally efficient and can be easily implemented. Tests using hyperspectral image data have shown that the proposed approach can contribute to increased accuracies in the classification process.

Palavras-chave: *a priori* probability, semi-labeled samples, high-dimensional image data.

1. Introdução

Dados em alta dimensionalidade como em imagens hiperespectrais, abrem novas possibilidades no processo de classificação de imagens digitais em sensoriamento remoto. Na medida em que a dimensionalidade dos dados aumenta, amostras extraídas de populações normalmente distribuídas tendem a deslocar-se para as caudas das distribuições, deixando a parte central na qual ocorrem os valores máximos para a função densidade de probabilidade virtualmente vazia (Landgrebe, 2003). Este fato torna viável a separação de classes espectralmente muito semelhantes, isto é, classes que apresentam valores muito similares ou mesmo idênticos para os vetores de médias, desde que as respectivas matrizes de covariância difiram significativamente entre si (Fukunaga, 1990).

Dados desta natureza apresentam, entretanto alguns problemas, entre os quais a questão da estimação dos parâmetros do classificador a partir das amostras de treinamento, geralmente disponíveis em um número apenas limitado. Iniciando o processo de classificação com um número reduzido de bandas espectrais, verifica-se que a acurácia na imagem temática produzida pelo classificador tende a aumentar na medida em que novas bandas espectrais são adicionadas, isto é, na medida em que informações adicionais são acrescentadas ao classificador. Em um determinado ponto o valor estimado para a acurácia atinge um máximo para em seguida passar a diminuir, na medida em que bandas adicionais continuam sendo introduzidas, isto é, na medida em que a dimensionalidade dos dados continua a aumentar, enquanto o número de amostras permanece constante. Este padrão de comportamento, conhecido como o “fenômeno de Hughes”

resulta das incertezas nos valores estimados para um número crescente de parâmetros do classificador, a partir de um número limitado de amostras.

Três abordagens gerais têm sido investigadas com o objetivo de minimizar os efeitos do fenômeno de Hughes: (1) métodos de análise discriminante regularizada, (2) métodos de seleção e/ou extração de variáveis e (3) métodos de geração de amostras semi-rotuladas. Neste estudo é proposta e testada uma extensão à abordagem originalmente proposta por Jackson e Landgrebe (2001) na geração de amostras semi-rotuladas. Na presente proposta, amostras semi-rotuladas além de servirem para fins de estimação dos parâmetros do classificador, são empregadas ainda na estimação da probabilidade *a priori* associada a cada uma das classes em consideração.

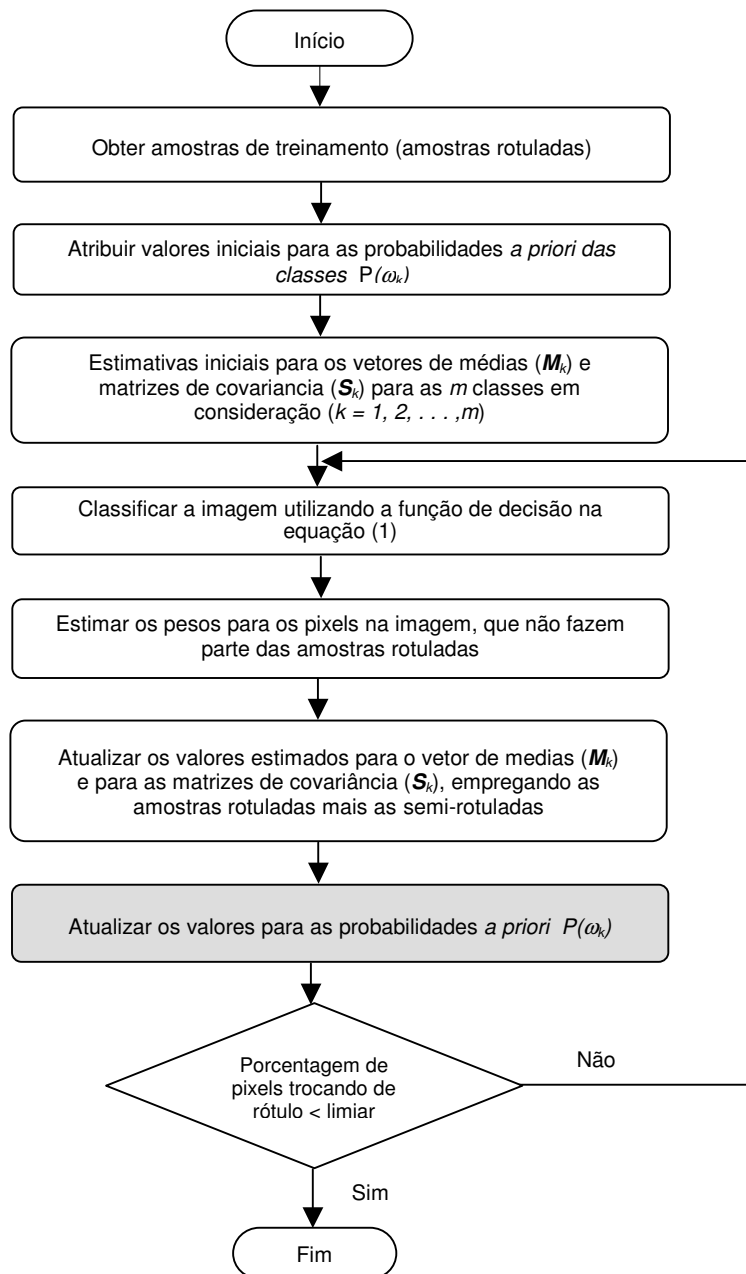


Figura 1. Fluxograma ilustrando a metodologia proposta. A caixa sombreada ilustra a contribuição adicional deste estudo.

2. Metodologia

A metodologia proposta está ilustrada de uma forma esquemática na Figura 1, onde o segmento atualizando os valores estimados para as probabilidades *a priori* se constitui na contribuição original deste estudo. O primeiro passo consiste na classificação inicial dos dados empregando o logaritmo natural da função de decisão de Bayes, na suposição de distribuição normal multivariada (1):

$$G_i(X) = -(X - \mu_i)^T \Sigma_i^{-1} (X - \mu_i) - \ln |\Sigma_i| + 2 \ln P(\omega_i) \quad (1)$$

onde μ e Σ representam os dois parâmetros (vetor de médias e matriz de covariância) associados à classe ω_i (1), $P(\omega_i)$ a probabilidade *a priori* e X o padrão (pixel) a ser classificado.

Em um passo inicial, os parâmetros são estimados utilizando somente as amostras disponíveis (amostras rotuladas) e assumindo valores idênticos para as probabilidades *a priori* das classes. A partir desta primeira iteração, os graus de pertinência de cada pixel a cada uma das classes podem ser estimados pela correspondente probabilidade *a posteriori* $p(\omega_i/X)$. Desta forma podem ser geradas as chamadas amostras semi-rotuladas. Estas amostras são constituídas por pixels extraídos da própria imagem e utilizados para incrementar o número de amostras disponíveis para fins de estimação dos parâmetros em (1).

A probabilidade *a posteriori* pode ser estimada pelo teorema de Bayes:

$$p(\omega_i / X) = \frac{p(X / \omega_i) P(\omega_i)}{\sum_{j=1}^m p(X / \omega_j) P(\omega_j)} \quad (2)$$

onde m representa o número de classes em consideração.

Deve-se aqui observar que os pesos são estimados apenas para amostras (pixels) que não fazem parte do conjunto das amostras rotuladas, para fins de geração das chamadas amostras semi-rotuladas. Para as amostras rotuladas o peso é mantido igual à unidade para a classe correspondente e zero para as demais. A probabilidade *a posteriori* (2) pode ser estimada por meio da equação (1):

$$w_{k,i} = p(\omega_i / X_k) = \frac{\exp[G_i(X_k)]}{\sum_{j=1}^m \exp[G_j(X_k)]} \quad (3)$$

$w_{k,j}$ representa o peso associado ao pixel X_k e à classe ω_j .

A partir da segunda iteração, um determinado número de amostras semi-rotuladas é adicionado de uma forma ponderada, ao conjunto das amostras rotuladas para estimação dos parâmetros (M_i e S_i) e da probabilidade *a priori* $P(\omega_i)$:

$$M_i = \frac{1}{\sum_{k=1}^{N_i} w_{k,i}} \sum_{k=1}^{N_i} w_{k,i} X_k \quad (4)$$

$$S_i = \frac{1}{\left(\sum_{k=1}^{N_i} w_{k,i} \right)^{-1}} \sum_{k=1}^{N_k} w_{k,i} (X_k - M_i)(X_k - M_i)' \quad (5)$$

$$P(\omega_i) = \frac{\sum_{k=1}^N w_{k,i}}{N} \quad (6)$$

onde N representa o número total de pixels na imagem.

Neste estudo propõe-se, portanto, estimar a probabilidade *a priori* de cada classe pela fração do peso associado a esta classe, estimada sobre toda a imagem. Este valor é re-estimado a cada etapa no processo iterativo ilustrado na Figura 1. O processo iterativo é terminado quando o número de modificações (pixels trocando de classe) entre duas iterações sucessivas é menor do que um determinado limiar, pré-fixado.

3. Resultados e Discussão

A metodologia aqui proposta foi testada empregando dados hiperespectrais cobrindo uma área teste denominada de *Indian Pines*, localizada no noroeste do Estado de Indiana – EUA, apresentada na Figura 2. Nesta cena aparecem campos agrícolas com culturas de soja e milho empregando técnicas de cultivo distintas (cultivo tradicional, cultivo direto e cultivo mínimo). A imagem foi coletada no início da cultura, mostrando uma proporção grande de solo exposto. Este conjunto de circunstâncias torna estas classes espectralmente muito semelhantes entre si, constituindo-se em um verdadeiro desafio para um classificador. Dados tradicionais em baixa ou média dimensionalidade, como por exemplo, dados Landsat-TM, produzem imagens temáticas com muito baixo grau de acurácia. Foram utilizados neste estudo dados coletados pelo sistema sensor AVIRIS, que coleta dados em 224 bandas espectrais, no intervalo (0.4 – 2.5 μ m) do espectro eletromagnético (Landgrebe, 2003). Deste conjunto foram removidas bandas ruidosas, restando 190 bandas. Neste estudo foram utilizadas 6 classes de cobertura do solo: milho plantio direto, milho em cultivo mínimo, soja plantio direto, soja em cultivo mínimo, soja em cultivo tradicional e floresta. Como ocorre com frequência em situações reais, o número de amostras disponíveis é pequeno para uma estimativa confiável dos parâmetros do classificador. Desta forma, esta área experimental oferece as condições adequadas para testar a metodologia proposta.

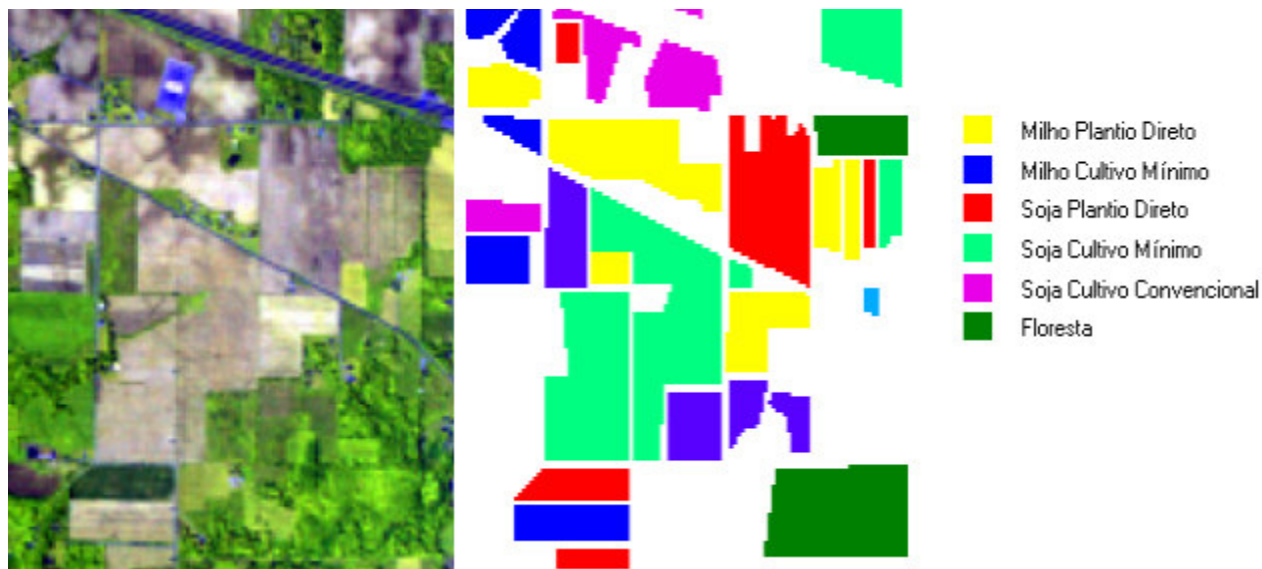


Figura 2. Composição colorida RGB com três bandas espectrais AVIRIS: 100(vermelho), 90(verde) e 30(azul), do segmento da região experimental de testes denominada *Indian Pine*, localizada a noroeste no Estado de Indiana – EUA, e a respectiva imagem temática da verdade terrestre, relativa às classes em estudo.

Para uma melhor avaliação da contribuição dos valores estimados para as probabilidades *a priori* das classes na acurácia da imagem temática produzida pela função de decisão (1), os seguintes experimentos foram feitos: (1) treinamento do classificador empregando apenas amostras rotuladas e adotando-se valores idênticos para as probabilidades *a priori* das classes (metodologia tradicional), (2) incluindo um determinado número de amostras semi-rotuladas na estimação dos parâmetros (método proposto por Jackson e Landgrebe (2001)), (3) adicionando ao processo, valores estimados para as probabilidades *a priori* das classes, conforme proposto e (4) empregando apenas amostras rotuladas, mas utilizando valores estimados para as probabilidades *a priori* das classes. Nestes experimentos foi estimada a acurácia na imagem temática produzida pela função de decisão (1) para valores crescentes na dimensionalidade dos dados. Procurou-se desta forma testar os efeitos desta metodologia no sentido de obter imagens temáticas mais acuradas através da minimização dos efeitos do fenômeno de Hughes com a introdução de amostras semi-rotuladas e da introdução de valores mais realistas para as probabilidades *a priori* das seis classes consideradas. Os resultados destes experimentos estão apresentados nas Figuras 3 e 4. A Figura 3 ilustra os valores médios para acurácia, estimados para as seis classes, com a utilização de 300 amostras rotuladas e 50 amostras semi-rotuladas por classe. Na Figura 4 encontra-se ilustrado os valores médios para a acurácia quando 500 amostras rotuladas e 50 amostras semi-rotuladas são empregadas. Várias conclusões podem ser obtidas destes resultados. Inicialmente, os resultados do fenômeno de Hughes são claramente visíveis. O valor estimado para a acurácia média tende inicialmente a crescer, refletindo a contribuição resultante da inclusão de informações adicionais (novas bandas espectrais) no processo de classificação. A acurácia atinge um máximo para em seguida decrescer, refletindo a dificuldade na estimativa de um número crescente de parâmetros. Coerentemente com as predições teóricas, este efeito é retardado quando o número de amostras rotuladas aumenta. As Figuras 3 e 4 servem também para ilustrar a contribuição das amostras semi-rotuladas e dos valores estimados para as probabilidades *a priori* conforme proposto neste estudo.

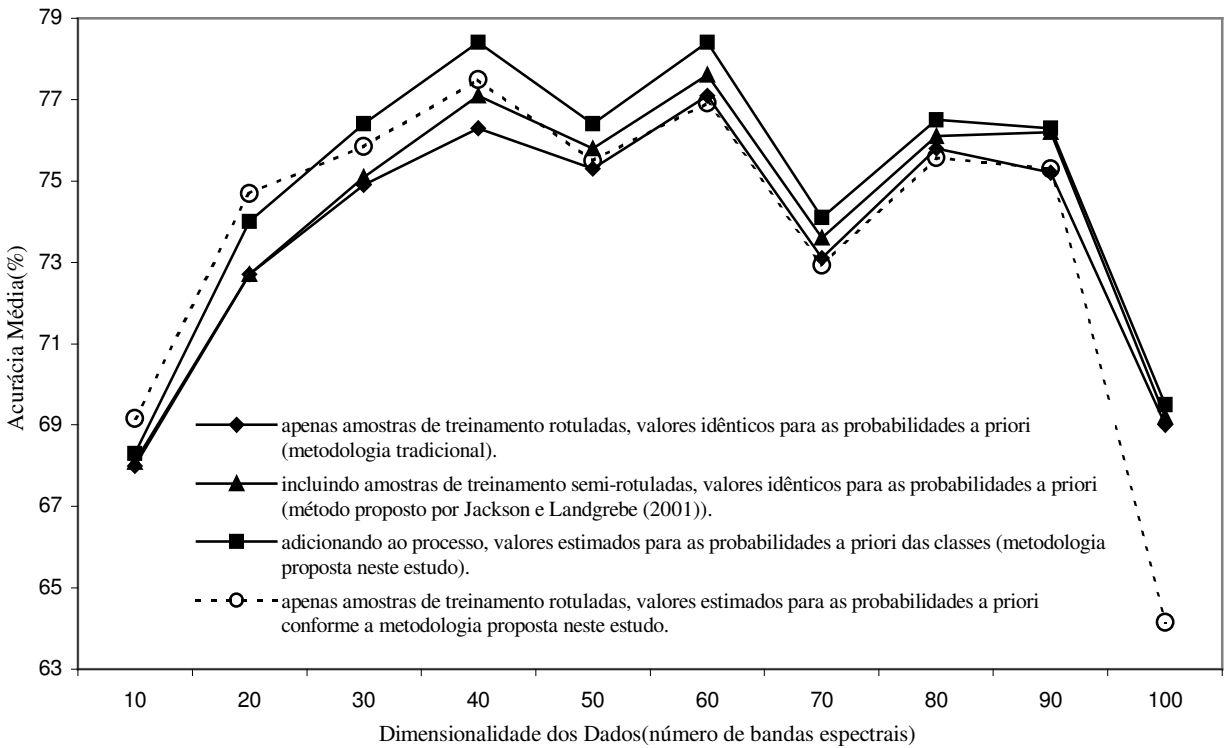


Figura 3. Valores médios para a acurácia, estimados para as seis classes, com a utilização de 300 amostras de treinamento rotuladas e 50 amostras semi-rotuladas por classe.

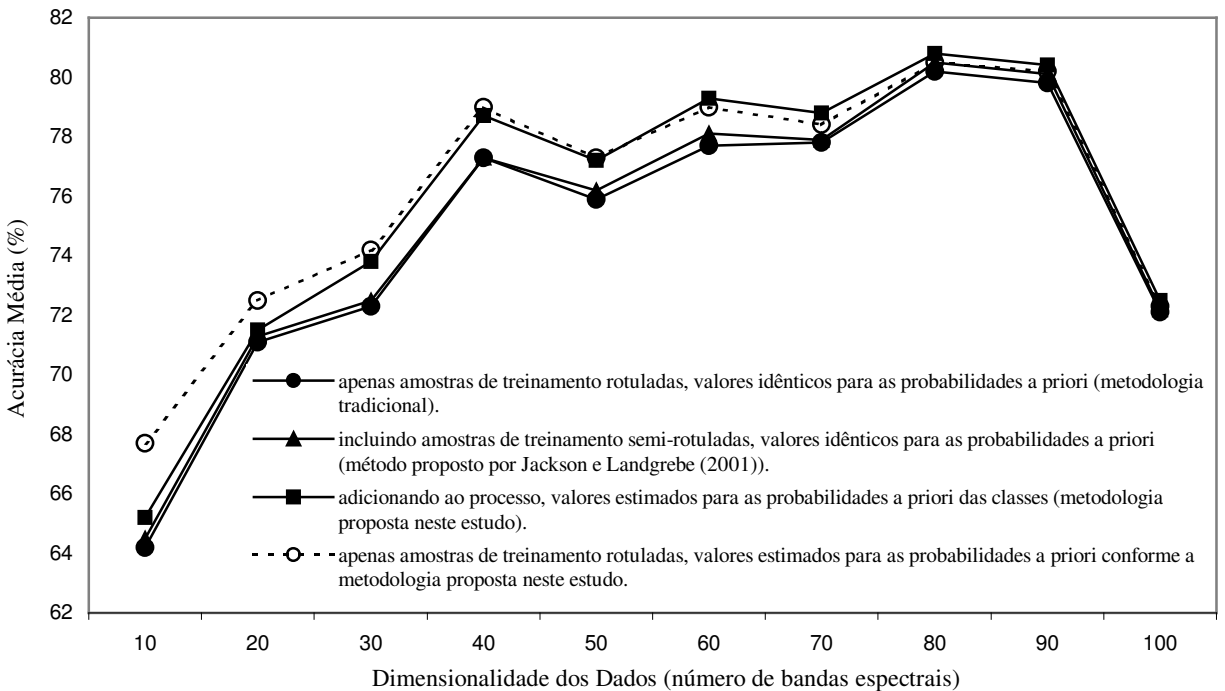


Figura 4. Valores médios para a acurácia, estimados para as seis classes, com a utilização de 500 amostras de treinamento rotuladas e 50 amostras semi-rotuladas por classe.

4. Conclusões

Neste estudo é investigada uma nova abordagem para estimação de valores para as probabilidades *a priori* em um classificador Bayesiano. A metodologia aqui proposta faz uso de amostras semi-rotuladas conforme proposto por Jackson e Landgrebe (2001). Neste estudo um passo adicional é incluído no processo iterativo que re-estima a cada etapa os parâmetros na função de decisão de Bayes, com o objetivo de obter valores mais realistas para as probabilidades *a priori* das classes envolvidas. A metodologia proposta é de fácil implementação e computacionalmente eficiente. Os experimentos desenvolvidos mostram a validade desta proposta, pelo incremento resultante na acurácia da imagem temática.

Referências Bibliográficas

Chelsea, C. White, III. A Posteriori Representations Based on Linear Inequality Descriptions of *A Priori* and Conditional Probabilities. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. SMC-16, No. 4, pp. 570-573, 1986.

Chittineni, C.B. Maximum Likelihood Estimation of Label Imperfection Probabilities and Its Use in the Identification of Mislabeled Patterns. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. GE-20, No. 1, pp. 99-111, 1982.

Fukunaga, K. **Introduction to Statistical Pattern Recognition**. 2nd edition, Academic Press, 1990.

Ibrahim, M.A.; Arora, M. K.; Ghosh, S. K.; Chen, H. Approaches to Improve Accuracy of Neural Network Classification of Images Dominated by Mixed Pixels. *Proceedings of the Geoscience and Remote Sensing Symposium, IGARSS*, Vol. 1, pp. 568-571, 2004.

Jackson, Q.; Landgrebe, D. An Adaptive Classifier Design for High-Dimensional Data Analysis With a Limited Training Data Set. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 39, No. 12, pp. 2664-2679, 2001.

Landgrebe, D. **Signal Theory Methods in Multispectral Remote Sensing**. Wiley-Interscience, 2003.

Shashahani, B.M.; Landgrebe, D. The effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon. *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 32, No. 5, pp. 1087-1095, 1994.

Törmä, M.; Lumme, J.; Patrikainen, N.; Luojus, K. **Fusion of Low Resolution Optical and High Resolution SAR Data for Land Cover Classification**, pp. 2680-2683, 2004.

Tran, D. Estimation of Prior Probabilities in Speaker Recognition. *Proceedings of 2004 International Symposium on Intelligent Multimedia, Video and Speech Processing*, pp. 141-144, 2004.