

Classificação de áreas de café em Minas Gerais por meio do novo algoritmo QMAS em imagem espectral Geoeye-1

Priscila Pereira Coltri¹
Robson Leonardo Ferreira Cordeiro²
Tamires Tessarolli de Souza²
Luciana Alvim Santos Romani³
Jurandir Zullo Jr.^{1,4}
Caetano Traina Jr.^{2,4}
Agma Juci Machado Traina^{2,4}

¹Cepagri / Feagri – Unicamp
Cidade Universitária – 13083-970 - Campinas - SP, Brasil
{priscila, jurandir}@cpa.unicamp.br

²Universidade de São Paulo – USP/ICMC
Av.Trabalhador são-carlense, 400 - 13566-590 - São Carlos - SP, Brasil
{robson, caetano, agma}@icmc.usp.br
tamirests@grad.icmc.usp.br

³Embrapa Informática Agropecuária
Caixa Postal 6041 – 13083-886 - Campinas - SP, Brasil
luciana@cnptia.embrapa.br

⁴ Bolsista de Produtividade CNPq

Abstract. Although there exist different image processing techniques that can be used to discover knowledge from satellite images of coffee crops, there are still many issues to be addressed. One of them is that automatic image classification techniques usually have problems to recognize patterns from images of coffee crops, due to their spatial variability and planting characteristics. In this context, we present a comparison of two different methods for the task of classifying a Geoeye-1 image of coffee fields from the South of the state of Minas Gerais, in Brazil. The compared methods are: QMAS, a new algorithm for image classification, and MAXVER, a traditional method commonly used by agronomists to classify satellite images. The overall statistical results were reasonable for the traditional MAXVER method. Nevertheless, it has presented 30% in average of misclassification between the classes: Coffee and Forest. The majority of the areas in which the misclassification occurred refer to the middle of the coffee field, which complicates the process of post-classification. On the other hand, the QMAS algorithm presented better results, being more efficient especially for the coffee classification, since it did not present classificatory confusion in the middle of the coffee area. Between all the coffee fields classified by QMAS, only one was wrongly recognized as forest. In addition, the QMAS method was able to classify a forest fragment in the middle of the coffee plantation. Thus, we conclude that the QMAS algorithm is a viable alternative for the classification of remote sensing images from coffee producing regions.

Palavras-chave: coffee crops, spectral pattern, QMAS classification algorithm, cafeicultura, padrão espectral, algoritmo de classificação QMAS.

1. Introdução

O Brasil é o maior produtor e exportador de café em grão do mundo, com uma área plantada de 2,3 milhões de hectares e uma produção de 2,57 milhões de toneladas, de acordo com dados do IBGE, publicados em 2006. Em termos espaciais, o estado de Minas Gerais é responsável por mais de 50% da produção de café arábica do Brasil.

Moreira et al (2004) ressaltam que a cafeicultura mineira, apesar de sua importância econômica e social, carece de informações complementares para o seu sistema produtivo, principalmente em relação à sua extensão, distribuição espacial e o ambiente em que é cultivada. Os mesmos autores afirmam que essa lacuna de informações restringe o planejamento da cultura, dificulta o planejamento ambiental, as práticas de manejo, bem como as estimativas de produtividade que utilizam índices de penalização resultantes de adversidades climáticas e/ou fitossanitárias.

A maioria dos trabalhos de mapeamento é feita com imagens Landsat devido à sua qualidade espectral, resolução e, principalmente, facilidade de acesso. Recentemente, trabalhos com imagens de alta resolução, como IKONOS e QUICKBIRD, têm se destacado em estudos na cultura cafeeira. Nesse sentido, o satélite Geoeye-1 pode ser uma ótima opção para estudos de café.

Lançado em setembro de 2008, o satélite Geoeye se tornou inovador por usar uma tecnologia nova em satélites comerciais de sensoriamento remoto, que é a capacidade de diferentes detalhamentos e precisão. Ele é capaz de gerar imagens com detalhamento de 0,41 metros no modo pancromático (tons de cinza) e com 1,65 metros de detalhamento no modo multiespectral. A empresa Geoeye ainda destaca que tão importante quanto a sua resolução é a sua precisão de localização, pois o satélite Geoeye-1 é capaz de localizar qualquer objeto com precisão de localização de até 3 metros. Tal nível de precisão de georeferenciamento nunca havia sido alcançado anteriormente por qualquer outro sistema de geração de imagens orbitais. Dentre suas principais características, destaca-se, além da sua alta resolução, a capacidade de gerar até 700km² de imagens pancromáticas e 350km² de imagens pancromática e multiespectral (PSM) por dia, o que faz do satélite uma ferramenta importante para estudos sobre a cultura do café.

A extração de conhecimento relativo ao café, a partir de imagens de satélites, ainda apresenta desafios a serem vencidos. A grande variabilidade de padrões espectrais da cultura dificulta o reconhecimento de padrões por procedimentos automatizados e semi-automatizados (Bernades et al, 2007). Além disso, parâmetros importantes do meio físico, típicos de locais próprios para o plantio de café, dificultam ainda mais a classificação da cultura. Bernardes et al (2007) destacam a declividade e a exposição de vertentes, que alteram o sombreamento e a geometria de reflexão da radiação.

Nas análises de imagens de sensoriamento remoto para identificação de culturas, as áreas cafeeiras costumam apresentar confusões com diferentes classes de ocupação agrícola, principalmente culturas anuais, como citrus. Adicionalmente, fragmentos de vegetação nativa, como cerrado e mata também são comumente confundidos com áreas de café (Souza et al, 2009).

A classificação de imagens orbitais pode ser feita visualmente (método mais trabalhoso) ou por meio de classificadores automáticos ou semi-automáticos, os quais nem sempre retornam resultados precisos. Dentre as técnicas de classificação supervisionada disponíveis, a máxima verossimilhança (MAXVER) é a mais utilizada para o tratamento de dados de satélite (Shiba et al, 2005). Diversos métodos e algoritmos vêm sendo testados em imagens orbitais com a finalidade de classificar café, como o uso de rede neurais artificiais (Andrade et al, 2009) e análise de textura (Souza et al, 2009), entre outras. No entanto, ainda há necessidade de aperfeiçoamento das técnicas a fim de obter resultados cada vez mais satisfatórios. Nesse contexto, Cordeiro et al., 2010 propuseram o algoritmo QMAS que combina técnicas de extração de características de imagens e estratégias de mineração de dados para a classificação semi-automática de grandes coleções de imagens, com base em poucos exemplos de imagens previamente classificadas, dadas como entrada ao algoritmo.

Diante do grande desafio que é classificar imagens de sensoriamento remoto de café, o objetivo deste trabalho foi aplicar o novo algoritmo QMAS para classificar áreas de café

comparando os resultados com o método tradicional de Classificação Supervisionada MAXVER, em imagens Geoeye-1. Os resultados indicam que o algoritmo QMAS obteve mais êxito na classificação das áreas de café do que o MAXVER, configurando-se em uma alternativa viável a classificação de imagens de satélite.

2. Metodologia do Trabalho

Nessa seção é brevemente apresentada a área de estudo, bem como detalhes das duas técnicas de classificação utilizadas no trabalho: QMAS e MAXVER, além dos procedimentos utilizados para a aplicação delas.

2.1 Área de Estudo

A área de estudo encontra-se no sul de Minas Gerais, no município de São Sebastião do Paraíso, na fazenda experimental da EPAMIG (Empresa Agropecuária de Minas Gerais), com altitude de 991m, latitude de 20°55'S e longitude de 46°55'W. A Figura 1 mostra uma imagem Geoeye-1 de 24 de maio de 2009, com detalhe (Figura 1a) para o recorte utilizado na aplicação dos algoritmos de classificação, denominado como área de estudo, no presente trabalho.

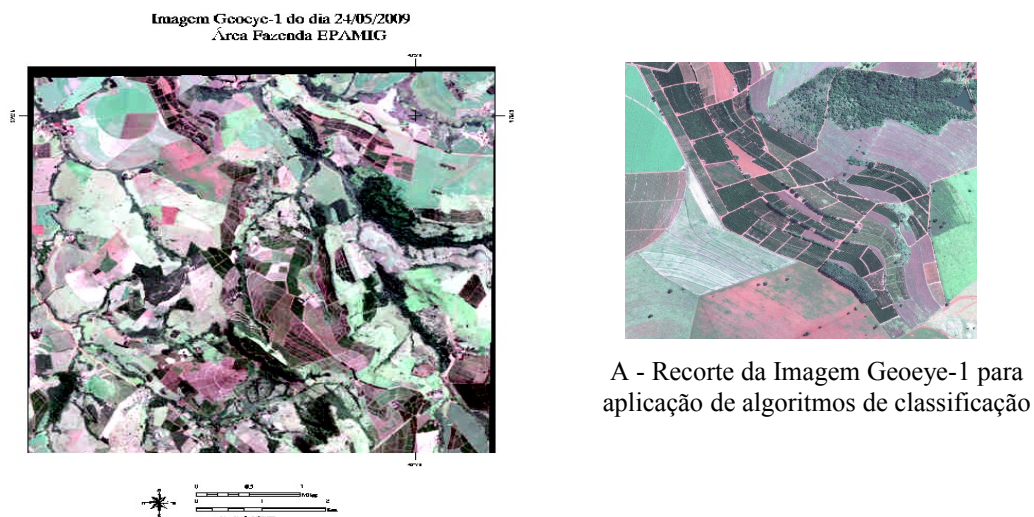


Figura 1 – Imagem Geoeye 1 - Composição RGB dia 24/05/2009. Em A.: Área selecionada: recorte para aplicação dos algoritmos de classificação

2.2. Métodos de Classificação:

Os dois métodos utilizados MAXVER e QMAS, são descritos a seguir:

(a) Classificação Supervisionada por Máxima Verossimilhança (MAXVER)

O classificador por Máxima Verossimilhança (MAXVER) é do tipo paramétrico (pois assume distribuição normal para cada uma das classes em cada uma das bandas) e calcula os parâmetros médios e a matriz de covariância (Jensen, 1986). Ele computa a probabilidade de um determinado pixel pertencer a cada classe determinada pelo usuário na fase de treinamento do classificador. O classificador MAXVER está presente em diversos softwares de SIGs (Sistemas de Informações Geográficas), daí a sua ampla utilização em sensoriamento remoto.

A classificação supervisionada MAXVER foi realizada no software ENVI com a imagem Geoeye-1 no formato RGB. Utilizou-se a classificação caracterizada como do tipo “pixel a pixel” porque a mesma utiliza apenas a informação espectral, isoladamente, de cada pixel para encontrar regiões homogêneas. A definição das classes de uso foi feita por interpretação

visual da imagem, a qual, devido sua alta resolução, também foi considerada como referência para avaliar a exatidão da classificação. Foram identificadas as seguintes classes de uso do solo: Café, Mata, Solo Exposto e Milho.

As amostras para treinamento do classificador foram definidas com base em padrão característico de cada classe de uso do solo, na própria imagem. Para avaliação da classificação, foi calculada uma matriz de erros, que aponta a distribuição dos erros e acertos (Jensen, 1986). A partir desta matriz, é possível verificar as classes com desempenhos melhores e piores, além de permitir o cálculo de parâmetros de exatidão geral da classificação quando se utilizam como base de extração dos dados, imagens de sensores orbitais. A partir desta matriz de erros é possível fazer uma análise multivariada, sendo que a mais utilizada é a geração do índice Kappa (K) (Gong e Howarth, 1990). Os mesmos autores sugerem que o índice K é uma medida de precisão importante a ser associada à matriz de erro, por representá-la inteiramente, isto é, considera todos os elementos da matriz e não apenas aqueles que se situam na diagonal principal.

A classificação foi avaliada segundo o valor do Índice Kappa (K), de acordo com a classificação de Fonseca (2000) onde: péssimo, para $K=0$; mau, para K entre 0 e 0,2 ; razoável, para K entre 0,2 e 0,4; bom, para K entre 0,4 e 0,6, muito bom, para K entre 0,6 e 0,8 e excelente, para K acima de 0,8. Além disso, aproveitando a alta resolução do satélite, foi feita a avaliação visual.

(b) Classificação por uso do Algoritmo QMAS

O algoritmo QMAS (Cordeiro et al., 2010) utiliza técnicas de extração de características de imagens e algoritmos de mineração de dados a fim de realizar a classificação semi-automática de grandes coleções de imagens. O algoritmo recebe como entrada dois conjuntos de imagens. O conjunto de treinamento provê poucos exemplos de imagens previamente classificadas, tipicamente entre dois a cinco exemplos por classe, enquanto o segundo conjunto, o conjunto de teste, consiste em uma grande coleção de imagens, não classificadas. A classificação realizada é dita semi-automática, pois as imagens do conjunto de treinamento são utilizadas como exemplos, os quais são generalizados de forma automática, possibilitando a classificação das imagens no conjunto de teste. De forma geral, o algoritmo QMAS consiste nos seguintes passos:

- (i) **extração de características** das imagens de entrada, que permite a representação dos dados em um espaço multi-dimensional;
- (ii) **agrupamento** das imagens de entrada;
- (iii) **criação de um grafo** em três níveis (grafo tripartido), composto de vértices, os quais representam os grupos encontrados no passo anterior, as imagens de entrada e as classes estipuladas na base de treinamento e, arestas, que representam as ligações conhecidas entre as imagens de entrada e seus respectivos grupos e classes;
- (iv) **análise do grafo** para definir as classes mais apropriadas para cada uma das imagens não classificadas. A classe mais apropriada para uma dada imagem é identificada da seguinte maneira: varreduras aleatórias são feitas no grafo, as quais são sempre iniciadas no nodo que representa a imagem a ser classificada; dentre os nodos que representam as classes, o nodo com maior probabilidade de ser visitado durante as varreduras indica a classe mais apropriada para a imagem de interesse.

Para utilização do algoritmo, foram consideradas as classes Café, Mata, Solo exposto e Milho, assim como foi feito para o método MAXVER. A área de estudo foi dividida em 4.096 retângulos de tamanhos iguais (64x64 pixels). Cada retângulo foi considerado como uma

imagem a ser classificada por QMAS. Apenas 13 exemplos classificados manualmente foram passados ao algoritmo, cinco exemplos de retângulos contendo Café, três exemplos de Solo Exposto, três exemplos de Mata e dois exemplos de Milho. Dessa forma, para classificar o conjunto com 4.096 imagens, foram utilizadas como exemplo apenas 13 imagens classificadas manualmente, cerca de 0,3% dos dados de entrada, enquanto QMAS classificou de forma automática os demais 99,7% dos dados.

As seguintes características foram extraídas de cada imagem (retângulo) a ser classificada por QMAS: “*haar wavelets*” em dois níveis de resolução e valores médios de vermelho, verde e azul. As “*wavelets*” são ferramentas matemáticas para a representação do incremento de informações entre duas resoluções ou escalas sucessivas. Estas são comumente usadas em processamento de imagens digitais para a extração de características que representam a textura das imagens.

3. Resultados e Discussão

A seguir, são apresentados os resultados da aplicação das técnicas MAXVER e QMAS para classificação da imagem Geoye-1 para a cultura do café.

3.1 Classificação Supervisionada MAXVER

A Figura 2 apresenta a resposta da classificação supervisionada MAXVER e a Tabela 1 apresenta a matriz de erros.

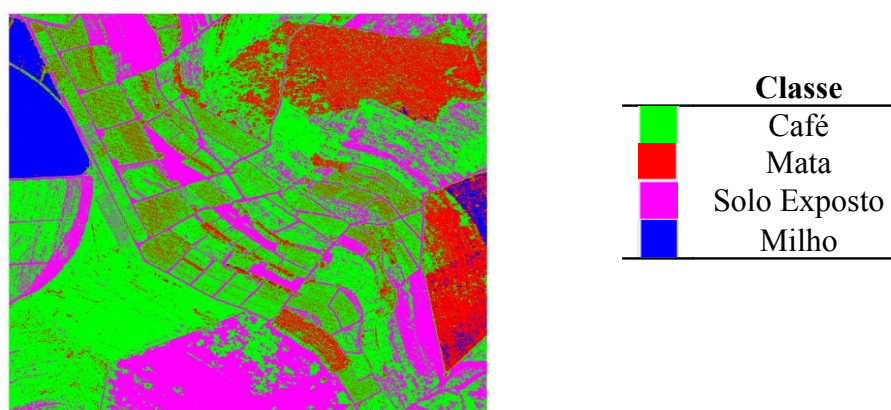


Figura 2- Imagem Classificada. Em verde: Café; em vermelho: Mata; em azul: Milho e em magenta: Solo Exposto.

Tabela 1 – Matriz de erros do classificador MAXVER

Classe	Café	Solo Exposto	Milho	Mata	Total
Não-Classificado	0,00	0,00	0,00	0,00	0,00
Café	78,57	0,00	0,00	30,10	16,62
Solo Exposto	0,04	99,99	0,00	0,00	16,05
Milho	0,00	0,00	99,98	0,00	52,82
Mata	21,38	0,01	0,02	69,90	14,51
Total	100,00	100,00	100,00	100,00	100,00

A classificação supervisionada MAXVER apresentou índice Kappa 0,87, com acurácia de 91% na classificação, sendo, portanto, considerada excelente segundo Fosenca (2000). Embora o índice Kappa e a acurácia tenham sido altos, a matriz de erros aponta a confusão clássica entre café e mata. Na análise visual, também é possível verificar essa confusão, conforme ilustrado na Figura 2. Pelo menos seis talhões de café apresentaram confusão com

mata. Pela matriz de erros, é possível verificar que, aproximadamente, 30% da área de café foi confundida com mata. A maior parte das confusões foi feitas no meio dos talhões de café (Figura 3a), o que dificulta o processo de pós-classificação manual. Em relação às áreas da classe Mata, também houve confusão com café (Figura 3b). Vilela et al (2000) classificou imagens Landsat com o algoritmo MAXVER e também encontrou confusão entre as classes café e mata para as composições das bandas 2, 3 e 4, e bandas 3, 4 e 5. Souza et al (2009), classificando café em Minas Gerais por meio de análise de textura, verificou que a precisão fica em torno de 83% entre as classes de café e mata, sendo que o erro de omissão da classe café foi de 18%, a da mata foi de 16%, e a de outros usos ficou em 1%. Os mesmos autores ainda destacam que a resposta espectral do café assemelha-se muito com a da mata nativa e, como os classificadores automáticos de imagens digitais convencionais utilizam apenas a resposta espectral da imagem, estes métodos não são indicados para classificar café.

Por outro lado, o algoritmo conseguiu detectar árvores no meio da plantação de café, nas áreas de café arborizado. As áreas de solo exposto foram bem classificadas, inclusive no meio da plantação de café. As áreas de milho também foram bem classificadas.

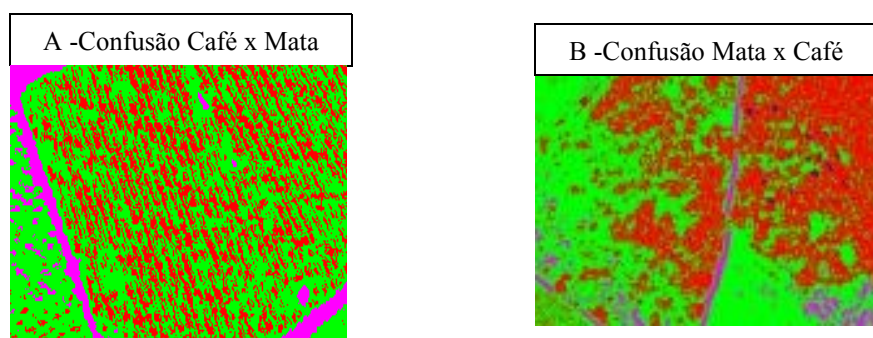


Figura 3 – Exemplo de erro do classificador MAXVER –Em A.: Confusão Café x Mata no meio do talhão de café; em B.: Confusão Mata x Café na área de Mata.

3.2 Classificação QMAS

Após a classificação pelo algoritmo QMAS, pode-se identificar que 22,9% da imagem foi classificada como Café, 50,4% foi classificada como Solo exposto, 26,5% foi classificada como Mata e 0,1% como Milho. A Figura 4 apresenta a imagem gerada pela classificação do QMAS.

Pode-se verificar que para a classe milho, o algoritmo não teve uma boa resposta, praticamente não encontrando pixels dessa classe. QMAS confundiu-se entre milho e mata, e milho com solo exposto, classificando as áreas de Milho como Mata e solo exposto. O inverso, porém não ocorreu, ou seja, nenhuma área de mata foi classificada como milho, e nenhuma área de solo exposto foi classificada como milho.

Para as demais classes escolhidas, o algoritmo gerou uma ótima classificação. Assim como o algoritmo MAXVER, o QMAS conseguiu identificar as classes de solo exposto entre os talhões de café, conforme exemplo na Figura 5a e 5b. De todos os talhões de café da imagem, apenas um talhão foi confundido com a classe Mata (quadrados em verde na Figura 5a), os demais foram classificados corretamente, sem a confusão no meio do talhão, característico do algoritmo MAXVER. Em relação à classificação da Mata, o algoritmo conseguiu classificar corretamente um fragmento florestal localizado no meio do café (Figura 6b), além de classificar algumas árvores no meio do cafezal (Figura 6a). No entanto, na área de mata, houve uma pequena confusão na classificação da mata, confundindo com alguns pixels de café (Figura 6c).

Na avaliação geral do algoritmo QMAS para classificação de café, houve um resultado bastante satisfatório, visto que praticamente todos os talhões foram classificados corretamente e houve pouca confusão em relação à classe Mata. Note que o mesmo não ocorre na classificação feita através do algoritmo MAXVER.

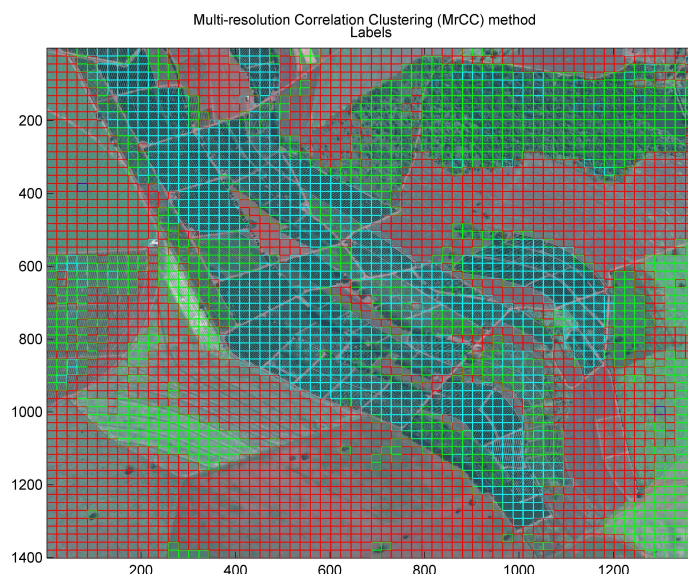


Figura 4- Imagem Geoeye-1 classificada pelo algoritmo QMAS. Em azul: áreas classificadas como café; em verde: áreas classificadas como Mata; em vermelho: áreas classificadas como solo exposto; e, em preto: áreas classificadas como milho.

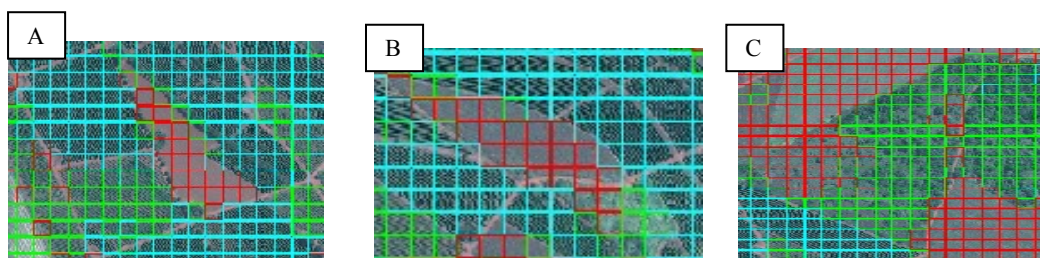


Figura 5 – Detalhes da classificação do algoritmo QMAS. Em A e B – detalhe da classificação correta de solo exposto entre os talhões de café; em C: detalhe da classificação correta de Mata que o MAXVER confundiu como sendo café.

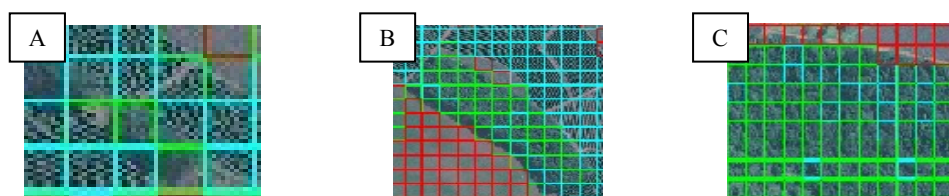


Figura 6 – Detalhes da classificação do algoritmo QMAS. Em A: Detalhe da classificação de árvores no meio do cafezal; em B; Detalhe da Classificação de um fragmento florestal e em C.: Detalhe da confusão de pixel de café na classe de mata.

4. Conclusão

A partir da avaliação dos resultados pode-se concluir que:

- A classificação tradicional por Máxima Verossimilhança, tradicional na área de sensoriamento remoto embora tenha tido um índice Kappa geral ótimo, apresentou confusão considerável entre as classes Café e Mata;
- As demais classes foram bem classificadas pelo método MAXVER;
- O algoritmo QMAS se mostrou satisfatório na classificação de áreas de café;
- As confusões entre as classes Café e Mata foram poucas e apenas um talhão não foi classificado pelo QMAS. Não houve confusão entre essas classes no meio dos talhões de café;
- O algoritmo QMAS não foi bom classificador para áreas de milho, muito provavelmente devido às características extraídas das imagens que não consideraram peculiaridades desse tipo de cultura;
- QMAS é uma alternativa bastante viável para classificação de café.

Agradecimentos: Os autores agradecem à FAPESP, Cnpq, CAPES, Microsoft Research e Embrapa Café pelo apoio financeiro e à EPAMIG pela área de estudo

4. Referências

ANDRADE, L.N.; VIEIRA, T.G.C.; LACERDA, W.L.; ALVES, H.M.R.; VOLPATO, M.M.L.; BRAGA,R.C.A.;SOUZA, V.C.O. Redes Neurais na classificação de áreas cafeeiras da região de Guaxupé. In.: In.: Simpósio de Pesquisa dos Cafés do Brasil, 6, 2009, Vitória. **Anais...** Minas Gerais: EPAMIG, 2009. On line.

BERNARDES, T.; ALVES, H.M.R.; VIEIRA, T.G.C. Classificação automática de imagens de satélite no mapeamento da região cafeeira de Patrocínio, MG. In.: Simpósio de Pesquisa dos Cafés do Brasil, 5,2007,SP. **Anais...**Minas Gerais:EPAMIG,2005. On-line.

CORDEIRO, R. L. F., Guo, Fan, Haverkamp, Donna S., Horne, James H., Hughes, Ellen K., Kim, Gunhee, Traina, Agma Juci Machado, Traina Jr., Caetano, Faloutsos, Christos. QMAS: Querying, Mining And Summarization of Multi-modal Databases In: IEEE International Conference on Data Mining - ICDM, 2010, Sydney, Australia (to appear).

Gong, P.; Howarth, P.J. An assessment of some factors influencing multispectral land-cover classification. **Photogrametric Engineering and Remote Sensing**. Bethesda, v. 56, n. 5, p. 597-603, 1990.

SHIBA,M.H. SANTOS, R.L.; QUINTANILHA, J.A.; KIM, H.Y. Classificação de imagens de sensoriamento remoto pela aprendizagem por árvore de decisão: uma avaliação de desempenho. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 12, 2005. Goiânia. **Anais...** São José dos Campos: INPE, 2005. Artigos, p4319-4326. CD-ROM, On line. ISBN 85-17-00018-8.Disponível em: <http://www.lps.usp.br/~hae/sbsr2005.pdf>

SOUZA, V.C.O.; VIEIRA, T.G.C.;ALVES,H.M.R.; VOLPATO, M.M.L. Análise e classificação textural de áreas de mata e café na região de Machado – MG. In: Simpósio Brasileiro de Sensoriamento Remoto (SBSR), 14, 2009. Natal. **Anais...** São José dos Campos: INPE, 2009. Artigos, p7925-7932. CD-ROM, On line. ISBN 85-17-00018-8.Disponível em: <http://marte.dpi.inpe.br/col/dpi.inpe.br/sbsr@80/2008/11.12.17.23/doc/7925-7932.pdf>. Acesso em 01 de nov de 2010.

MOREIRA, M. A., ADAMI, M. e RUDORFF, B. F. T. Análise espectral e temporal da cultura do café em imagens Landsat. **Pesquisa Agropecuária Brasileira**, v.39, n.3, p.223-231, 2004.

FONSECA, L.M. **Processamento digital de imagens**. São Jose dos Campos: Instituto Nacional de Pesquisas Espaciais, 2000. 105p.

JENSEN, J.R. **Introductory Digital Image Processing**. Prentice Hall, E. Cliffs, New Jersey, 285p. 1986.

VILELA,M.F.; SOARES,V.P.;RIBEIRO,J.C.;BRITES,R.S. Avaliação de técnicas de realce e classificação digital na elaboração de um mapa de uso da terra mediante imagem TM/Lanadsat-5. *Revista Árvore*, Viçosa-MG, v.24, n.2, p.161-173, 2000.