

## Classificação de Dados LIDAR Utilizando Estatística Multivariada

Marcos Benedito Schimalski <sup>1</sup>  
Edson Aparecido Mitishita <sup>2</sup>  
Tamara Mokwa <sup>3</sup>

<sup>1</sup> Universidade do Contestado – Programa de Mestrado em Des. Regional  
Caixa Postal 1 - 89460-000 - Canoinhas - SC, Brasil  
mbs@cni.unc.br

<sup>2</sup> Programa de Pós-Graduação em Ciências Geodésicas  
Setor de Ciências da Terra – Universidade Federal do Paraná  
Caixa Postal 19001 - 81531-990 - Curitiba - PR, Brasil  
mitishita@ufpr.br

Universidade do Contestado – Curso de Engenharia Florestal  
Caixa Postal 1 - 89460-000 - Canoinhas - SC, Brasil  
tamaramokwa@yahoo.com.br

**Abstract.** The forestry sector employs information referenced to a geodetic system for different purposes, in order to guide its field operations that are driven primarily by the Brazilian Forest Law. The availability of updated three-dimensional cartographic databases is a key factor for the success of forest activities. In this scenario, the laser scanner on board aircraft has been used for the derivation of products related to morphology of mapped terrain such as the contour lines and digital terrain model (DTM). If there is a need for information on land cover, these are generated from digital images obtained by sensors onboard airborne platforms or space. The purpose of this research was to develop methods that allow the classification of land cover based on accurate information coming from the laser scanning and functions of recognition of multivariate statistics. Experiments were carried out to classify the laser data relating to a rural area located near the city of União da Vitória, in the state of Paraná. Vegetation height, density of points, average and standard deviation of the orthometric height and intensity of the return laser pulse were selected as cofactors of classification. For the automatic classification of classes for soil cover, the Fisher's Linear Discriminant Function and Quadratic Discriminant Function were employed. The accuracy of the process was verified by comparing the data obtained by the ranking functions and data obtained through manual photo interpretation of aerial imagery of the site.

**Palavras-chave:** LASER Scanner, Classification, Automation, Fisher's Linear Discriminant Function, Quadratic Discriminant Function, LASER Scanner, Classificação, Automação, Função Discriminante de Fisher e Função Discriminante Quadrática.

## 1. Introdução

As questões ambientais, principalmente aquelas relacionadas ao aquecimento global, têm ocupado lugar de destaque nos meios de comunicação, na sociedade e na esfera governamental. A pressão pela preservação e mesmo recuperação dos recursos naturais tem fortalecido a política e a legislação ambiental no Brasil. Neste contexto, o Código Florestal Brasileiro (Lei nº 4.771 de 15 de setembro de 1965) apresenta parâmetros norteadores às atividades no espaço físico, visando limitar as pressões sobre os recursos naturais além de subsidiar juridicamente as atividades de fiscalização e planejamento.

O Código Florestal Brasileiro define as áreas de preservação permanente (APP), cuja utilização é bastante restritiva. Neste caso, citam-se como exemplos, as formações naturais flúvio lacustres relacionadas à largura dos rios e áreas de encosta, nas quais a declividade seja superior a 45°.

O contexto apresentado evidencia a necessidade de informações tridimensionais sobre a superfície terrestre, que possibilitem verificar a observância das leis ambientais e também derivem informações auxiliares às atividades florestais em campo, otimizando desta forma recursos.

As técnicas disponíveis para a obtenção de informações tridimensionais dividem-se naquelas que obtêm observações diretamente do terreno e as técnicas indiretas. As técnicas de mensuração indireta apresentam vantagens, principalmente devido à facilidade da obtenção das informações, comparativamente às técnicas de mensuração em campo. A geração de pares estereoscópicos associada às técnicas automáticas de geração do MDT (Modelo Digital do Terreno) são as mais comumente empregadas para a extração da altimetria (BALTSAVIAS, 1999).

Alternativa ao processo fotogramétrico, o ALS (*Airborne LASER Scanner* - LASER Scanner Aéreo) é uma inovação tecnológica bastante importante para o sensoriamento remoto, sendo uma tecnologia relativamente nova que possibilita a geração de coordenadas tridimensionais de uma nuvem de pontos sobre a superfície da Terra (DALMOLIN; SANTOS, 2003). A elevada automatização nos processos de coleta e processamento das informações, precisão, acurácia e alta densidade de pontos, além da rapidez, tornam os sistemas LIDAR inovadores.

Contudo, o processamento dos dados derivados pelo LIDAR carece de metodologias que possibilitem a extração de feições classificadas em antrópicas (produzidas pelo homem, citando como exemplos, edificações e rodovias) e naturais (formações vegetais, rios e relevo como exemplos). Para possibilitar esta classificação, alguns autores sugerem a integração das observações LIDAR com fontes de dados espectrais, possibilitando desta maneira a obtenção de informações planimétricas. Habib *et al.* (2004) destacam que dados fotogramétricos e de LIDAR têm qualidades específicas que tornam uma solução ou outra preferencial para certas aplicações. Uma desvantagem nas medições fotogramétricas é complementada por uma vantagem no sistema LIDAR, e vice-versa. As informações altimétricas obtidas pelo LIDAR tem qualidade muito superior quando comparadas com as informações planimétricas e o contrário pode ser verificado no processo fotogramétrico. Com isso, a integração entre informações espectrais e LIDAR apresenta benefícios interessantes.

Neste ponto deve ser destacado que o emprego de informações espectrais adicionais ao LIDAR, se traduz diretamente em um aumento de custos relacionados a obtenção e processamento destas informações. Destaca-se ainda que o processamento das informações espectrais não seja totalmente automatizado, o que demanda grande quantidade de tempo para a extração das informações planimétricas a partir desta fonte de dados.

A proposta apresentada neste trabalho foi o desenvolvimento de metodologias automatizadas capazes de extrair as classes temáticas relacionadas com a cobertura do solo a partir dos dados brutos LIDAR, sem a necessidade de uma fonte adicional de informações.

Como região de testes foi escolhida uma área de transição entre o ambiente urbano e o rural localizada nas proximidades de cidade de União da Vitória no estado do Paraná.

O reconhecimento de padrões foi empregado no desenvolvimento dos algoritmos de classificação, sendo escolhidas para os testes a Função Discriminante Linear de Fisher e a Função Discriminante Quadrática. Inicialmente, foi compilada a base cartográfica temática do local com o auxílio da fotointerpretação, restituição estereofotogramétrica e reambulação de campo. A partir desta base cartográfica foram selecionadas e então obtidas amostras das classes temáticas a partir dos dados LIDAR, tornando possível o ajuste das funções de discriminação. O processo termina com a classificação dos dados LIDAR e determinação dos índices de acerto para cada função empregada.

Como cofatores das funções de discriminação escolheram-se: altitude e intensidade média, desvio padrão da altitude e da intensidade, densidade de pontos e altura da cobertura vegetal. Para a obtenção deste último cofator foi necessário desenvolver um filtro para a separação dos pontos LIDAR localizados na superfície topográfica e acima desta.

## **2. Metodologia de Trabalho**

### **2.1 Fotointerpretação das fotografias aéreas verticais**

Para verificar a exatidão da metodologia proposta, os dados gerados a partir das observações fotogramétricas, associados com os obtidos na reambulação foram considerados, neste trabalho, como verdade terrestre.

Desta forma, para que a espacialização das observações fosse o mais próxima possível da realidade de campo foi necessário o treinamento na fotointerpretação da cobertura do solo. Este treinamento na fotointerpretação da vegetação foi feito com auxílio dos professores do Departamento de Ciências Florestais da Universidade Federal do Paraná.

Selecionaram-se dois pares estereoscópicos formados pelas fotografias 02, 03 e 04 da faixa 03, resultando nos modelos mo0203 e mo0304.

Tomando por referência a classificação vegetal proposta por Veloso, Filho e Lima (1991) e com o auxílio do estereoscópio de espelhos identificaram-se doze classes temáticas referentes à cobertura do solo na área dos experimentos. Estas classes temáticas podem ser agrupadas em:

- Região da Floresta Ombrófila Mista:
  - Estágio Inicial da Sucessão Vegetal herbáceo-arbustiva – capoeirinha (CS);
  - Estágio Inicial da Sucessão Vegetal arbórea – bracatinga (BR);
  - Estágio Inicial da Sucessão Vegetal arbórea – capoeira (IN);
  - Estágio Médio da Sucessão Vegetal arbórea – capoeirão (ME);
  - Estágio Avançado da Sucessão Vegetal – floresta (AV) e
  - Floresta de galeria (CI).
- Áreas de Formações Pioneiras com influência Flúvio-lacustre – várzea (VA).
- Áreas de antropismos:
  - Agropecuária (AP);
  - Reflorestamento (RE);
  - Edificações (ED);
  - Solo exposto (SO) e
  - Corpos d'água (AG).

### **2.2 Reambulação e obtenção do apoio fotogramétrico**

Para a obtenção da base cartográfica digital contendo os elementos foto interpretados realizou-se a restituição fotogramétrica estereoscópica, com a utilização de um restituidor analógico com captura dos dados da representação vetorial em ambiente digital. O estágio inicial nesta atividade consiste na obtenção de modelos estereoscópicos fotogramétricos

orientados em relação ao espaço objeto. Para o estabelecimento desta relação é necessária a obtenção das coordenadas tridimensionais de pontos foto-identificáveis (pontos de apoio), referidos a um referencial terrestre.

A finalidade primordial, mas não a única, dos pontos de apoio é a materialização do referencial com que se deseja trabalhar. A fim de materializar um referencial tridimensional sete parâmetros terão de ser determinados, sendo que estes parâmetros definem a posição da origem do sistema, a orientação de seus três eixos e a escala (ANDRADE, 2003). Foram obtidos pontos próximos ao de Gruber que possibilitam a orientação absoluta e relativa do modelo. Para a determinação das coordenadas foram empregados receptores de sinais GPS (*Global Positioning System*) de uma frequência e com estação de referência localizada próxima a área. O método de ocupação foi o Estático, com tempo de permanência de 30 minutos por ponto de controle.

### 2.3 Referenciamento cartográfico da fotointerpretação

Conforme mencionado, para a representação cartográfica da foto-interpretção, em meio digital, foram empregadas técnicas de restituição fotogramétrica. A restituição planialtimétrica dos dois modelos (0203 e 0304) foi conduzida por profissional experiente em atividades desta natureza. A coleta das informações foi realizada em um restituidor marca Wild, modelo B8S, adaptado para a coleta das informações diretamente no computador (restituidor semi-analítico).

As etapas de trabalho de restituição foram:

- orientação interior;
- orientação relativa;
- orientação absoluta;
- Restituição planimétrica: nesta etapa o fotogrametrista utilizando o restituidor semi-analítico procedeu a coleta das observações, nos modelos fotogramétricos, que possibilitou a delimitação da cobertura vegetal. O resultado desta etapa foi o mapa temático referente às classes de cobertura vegetal que foi empregado para avaliar os resultados obtidos pelas funções discriminantes;

Para a realização da restituição plani-altimétrica da área de estudo foram necessárias 45 horas de trabalho.

### 2.3 Armazenamento das observações LIDAR

Para o armazenamento das observações brutas LIDAR, optou-se por adoção de uma hiper-matriz armazenada no programa Matlab 2008b. Cada célula da hiper-matriz apresentou resolução geométrica de 10 metros. A Figura 1 apresenta uma célula da hiper-matriz com os respectivos pontos.

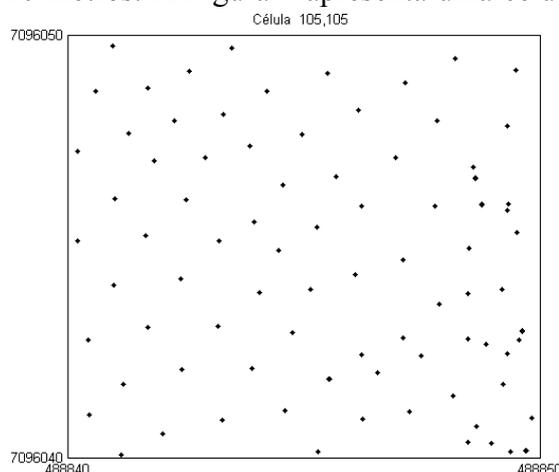


Figura 1 – Exemplo de uma célula da hiper-matriz de dados

Foram realizados vários testes visando a determinação das características que melhor predissessem a cobertura vegetal. Após estes testes com os dados foram calculados para cada célula os seguintes co-fatores:

- h: altitude ortométrica média;
- $\sigma_h$ : desvio padrão das altitudes ortométricas;
- I: intensidade média;
- $\sigma_I$ : desvio padrão das intensidades;
- $\Delta h$ : altura média da cobertura vegetal;
- De: densidade de pontos.

A obtenção das amostras representativas das classes relativas a cobertura vegetal foi realizada a partir da análise da base cartográfica obtida pelo processo fotogramétrico. A distribuição das amostras na área de estudo é mostrada na Figura 2.

As amostras foram então empregadas para o ajuste das Funções de Discriminação. A Função Discriminante Linear de Fisher para o reconhecimento de várias populações é:

$$Y = \underline{c}'\underline{X} = (\underline{\mu}_1 - \underline{\mu}_2)' \underline{\Sigma}^{-1} \underline{X}$$

Maiores detalhes sobre a Função Discriminante de Fisher poderão ser obtidos em Johnson (1998).

A segunda função de discriminação empregada foi a Quadrática:

$$D_i = \frac{1}{2} \ln |S_i| - \frac{1}{2} \underline{X}' S_i^{-1} \underline{X} + \bar{X}_i' S_i^{-1} \underline{X} - \frac{1}{2} \bar{X}_i' S_i^{-1} \bar{X}_i + \ln (P_i)$$

Paixão (2006) em sua dissertação de mestrado apresenta um detalhamento maior sobre esta função.

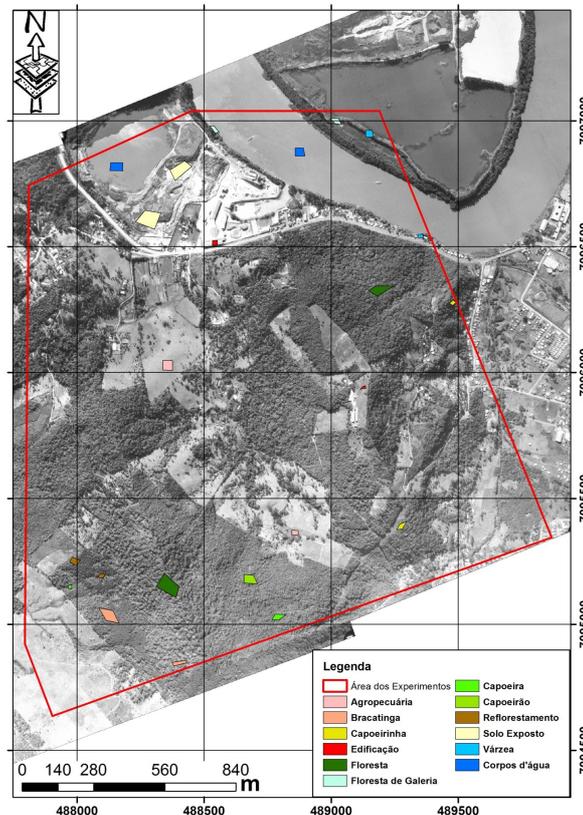


Figura 2 – Distribuição das amostras na área de estudo

Após o ajuste das funções discriminantes procedeu-se a classificação automatizada da hiper-matriz. A etapa final do trabalho foi a comparação dos resultados obtidos com a

classificação proposta com a obtida no processo fotogramétrico. A partir disto foi possível obter a matriz de confusão para cada classificador e o acerto relativo a cada processo.

### 3. Resultados das Classificações

A partir das amostras ajustou-se as 6 equações de discriminação, cujos coeficientes são mostrados na Tabela 1.

TABELA 1 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE REGULAR DE 10m EMPREGANDO TODAS AS VARIÁVEIS

Função	h	I	$\Delta h$	$\sigma H$	$\sigma I$	De	Constante
$\hat{Y}_1$	0,09277	0,00115	0,04537	-0,11933	0,00996	0,87532	-78,35670
$\hat{Y}_2$	-0,01576	0,01137	0,29000	0,22009	0,02284	-1,21332	11,65110
$\hat{Y}_3$	-0,01865	0,02845	-0,06903	0,44061	0,10804	2,27860	6,06356
$\hat{Y}_4$	0,00179	0,04524	-0,03173	0,19485	-0,08249	-0,40956	-3,19649
$\hat{Y}_5$	0,00000	0,00115	-0,05719	-0,91753	0,16001	-1,29507	-0,18246
$\hat{Y}_6$	0,00572	-0,01427	-0,08636	0,78176	-0,01095	-1,07782	-2,21513

Com ajuste das funções de discriminação calcularam-se as coordenadas dos centróides relativos a cada população (classe temática). Com o ajuste das funções discriminantes e o cálculo dos centróides das classes procedeu-se a classificação da grade regular de 10 m. A classificação obtida foi confrontada com a base cartográfica obtida pela restituição fotogramétrica e então se obteve a matriz de confusão que é apresentada na Tabela 2.

TABELA 2 - MATRIZ DE CONFUSÃO RELATIVA A CLASSIFICAÇÃO - FISHER

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma Parcial
AP	0,834	0,000	0,023	0,000	0,053	0,000	0,000	0,000	0,000	0,002	0,000	0,088	1,000
BR	0,000	0,855	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,017	0,128	0,000	1,000
CS	0,020	0,008	0,701	0,000	0,026	0,006	0,003	0,014	0,055	0,148	0,009	0,011	1,000
CI	0,003	0,000	0,000	0,703	0,000	0,000	0,004	0,028	0,015	0,247	0,000	0,000	1,000
ED	0,353	0,000	0,002	0,000	0,306	0,005	0,005	0,006	0,000	0,000	0,000	0,324	1,000
AG	0,006	0,000	0,002	0,006	0,001	0,625	0,310	0,000	0,000	0,033	0,000	0,018	1,000
VA	0,004	0,000	0,000	0,014	0,000	0,343	0,638	0,000	0,000	0,000	0,000	0,001	1,000
AV	0,000	0,032	0,000	0,001	0,006	0,000	0,001	0,768	0,120	0,068	0,003	0,000	1,000
ME	0,000	0,009	0,000	0,001	0,002	0,001	0,003	0,093	0,792	0,072	0,026	0,000	1,000
IN	0,005	0,004	0,069	0,075	0,000	0,000	0,010	0,034	0,045	0,754	0,003	0,001	1,000
RE	0,000	0,098	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,902	0,000	1,000
SO	0,107	0,000	0,000	0,003	0,015	0,002	0,005	0,000	0,000	0,001	0,000	0,867	1,000

A diagonal da matriz apresentada na Tabela 2 verifica-se o índice de acerto entre as classes apresentadas na seção 2.1.

Visando comparar os resultados obtidos com a Função Discriminante Linear de Fisher, o sexto experimento empregou a Função Discriminante Quadrática com o emprego de todas as variáveis na classificação da grade de 10 metros.

O ajuste das funções de classificação ocorreu de maneira similar: entrada dos valores dos cofatores e cálculo dos coeficientes das 12 funções.

O primeiro passo foi realizar o ajuste das funções quadráticas. As funções quadráticas dependem de coeficientes de segundo grau, primeiro grau e das constantes. A metodologia adotada para o ajuste das funções foi a mesma utilizada por Paixão (2006).

A partir das funções de discriminação foi implementada a classificação da hiper-matriz e então gerada a respectiva matriz de confusão. Esta matriz é mostrada na Tabela 3.

TABELA 3 - MATRIZ DE CONFUSÃO RELATIVA A CLASSIFICAÇÃO - QUADRÁTICA

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma
AP	0,773	0,000	0,051	0,029	0,026	0,000	0,000	0,023	0,032	0,030	0,000	0,037	1,000
BR	0,000	0,824	0,000	0,000	0,000	0,000	0,000	0,006	0,001	0,003	0,166	0,000	1,000
CS	0,024	0,000	0,702	0,023	0,008	0,001	0,002	0,052	0,034	0,149	0,000	0,005	1,000
CI	0,001	0,000	0,005	0,616	0,000	0,000	0,000	0,058	0,044	0,267	0,009	0,000	1,000
ED	0,310	0,000	0,008	0,003	0,363	0,000	0,000	0,020	0,006	0,000	0,000	0,290	1,000
AG	0,000	0,000	0,007	0,006	0,030	0,578	0,342	0,009	0,021	0,000	0,000	0,007	1,000
VA	0,011	0,000	0,000	0,017	0,042	0,302	0,620	0,004	0,000	0,000	0,000	0,004	1,000
AV	0,004	0,000	0,038	0,016	0,000	0,000	0,000	0,757	0,079	0,095	0,010	0,000	1,000
ME	0,001	0,003	0,033	0,015	0,004	0,000	0,000	0,127	0,805	0,007	0,000	0,005	1,000
IN	0,015	0,004	0,064	0,159	0,010	0,000	0,000	0,062	0,074	0,609	0,001	0,002	1,000
RE	0,000	0,175	0,008	0,000	0,000	0,000	0,000	0,011	0,003	0,001	0,802	0,000	1,000
SO	0,181	0,000	0,000	0,005	0,002	0,003	0,000	0,000	0,002	0,000	0,001	0,806	1,000

Os resultados apresentados nas Tabelas 2 e 3 possibilitaram a criação do gráfico mostrado na Figura 3.

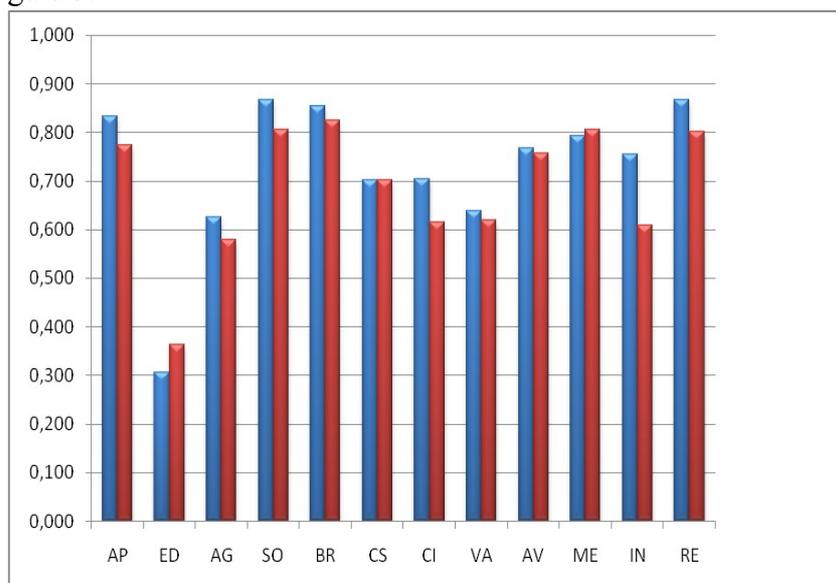


Figura 3 – Classificação por Fisher (azul) e Quadrática (vermelho)

A análise dos resultados mostra que a melhor classificação foi obtida com as classes agropecuária (AP), solo exposto (SO), bracatinga (BR) e Reflorestamento com índices de acerto acima de 80%. O pior desempenho foi verificado na classificação das edificações com índices próximos de 30%. Neste caso, a resolução das células da hiper-matriz está diretamente relacionada, pois a partir da restituição verificou-se que das 209 edificações existentes na área de estudo, 133 apresentam área do telhado inferior a 70 m<sup>2</sup>, ou seja, menor que a área de uma célula que é de 100 m<sup>2</sup>.

O acerto médio da classificação empregando a função quadrática foi de 68,8%, 4,1% menor que os 72,9% obtidos com função de Fisher. Pode-se afirmar que as mesmas causas de confusão detectadas experimento empregando a função de Fisher afetaram o desempenho da classificação com a Função Discriminante Quadrática. Percebe-se, no entanto, que o modelo

de separação de classes não linear é mais apropriado para as classes do grupo antrópico percebendo-se o oposto para as formações naturais. Outro aspecto negativo que deve ser considerado é o esforço computacional necessário para obter a classificação empregando o discriminante quadrático.

#### 4. Conclusões

Os resultados obtidos possibilitam elaborar as seguintes conclusões:

- Os experimentos para a obtenção das classes temáticas foram realizados com êxito, comprovando que a metodologia proposta é capaz de obter a classificação dos dados LASER de acordo com a cobertura do solo detectada pela reambulação de campo e restituição fotogramétrica.

- O *laser scanner* é uma tecnologia emergente e seu emprego na obtenção de informações altimétricas está bastante consolidado. Carecem ainda, metodologias que propiciem o melhor aproveitamento da tecnologia, justamente na classificação da cobertura do solo das áreas imageadas. O desenvolvimento das metodologias de classificação apresentadas, ajudam a suprir esta lacuna conseguindo sucesso significativo na separação das classes temáticas relativas a cobertura do solo, destacando-se que foram utilizados somente os dados LASER no processo. O sucesso da metodologia é perceptível, pois obteve-se um índice de acerto médio na separação das classes temáticas referentes a cobertura do solo de até 72,9%.

- A escolha das variáveis (cofatores) constituiu-se em atividade primordial para as atividades de classificação das formações vegetais a partir das células da grade regular. A utilização de todos os cofatores analisados: altitude média, desvio padrão da altitude, intensidade média, desvio padrão da intensidade, altura média da cobertura vegetal e densidade de pontos, é a melhor configuração para a separação das classes temáticas.

#### Referências Bibliográfica

ANDRADE, J. B. **Fotogrametria**. Curitiba: SBEE, 2003, 2ª ed.

BALTSAVIAS, E. P. (a). A comparison between photogrammetry and LASER scanning. **Journal of Photogrammetry and Remote Sensing**. Vol. 54, p. 83-94, 1999.

DALMOLIN, Q.; SANTOS, D. R. DOS. **Sistema LASER scanning: conceitos e princípios de funcionamento**. Curitiba: Editora da UFPR, 2003.

JOHNSON, D. E. **Applied multivariate methods for data analysts**. U.S.A.: Duxbury Press, 1998.

HABIB, A. F.; GHANMA, M.; MORGAN, M.; MITISHITA, E. *Integration of LASER and Photogrammetric Data for Calibration Purposes*. In: XX ISPRS Congress. Istanbul, 2004. **Anais**.

PAIXÃO, L. A. **Avaliação da Qualidade do Óleo Isolante em Transformadores com o Emprego da Função Discriminante Quadrática**. Dissertação. Curso de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2006.

VELOSO, H. P.; FILHO, A. L. R. R.; LIMA, J. C. A.. **Classificação da vegetação brasileira, adaptada a um sistema universal**. Rio de Janeiro: IBGE, 1991.