

1. **I**ntrodução

No Processamento Digital de Imagens, um dos processos mais importantes é a classificação de imagens (Mascarenhas e Velasco, 1989), capaz de extrair informações qualitativas e quantitativas das mesmas (Young, 1986). São várias as técnicas de classificação existentes. Em suma, tem-se os classificadores numéricos, a metodologia visual e os classificadores assistidos por sistemas especialistas.

Os métodos de classificação numéricos constituem-se em dois grandes grupos: os não-supervisionados e os supervisionados (Saebo et al, 1985). Os primeiros dividem as classes automaticamente e boa parte deles baseiam-se no princípio de Fisher (Johnson, 1982 e Fisher, 1936) da década de 30. Existem outros métodos baseados também em funções discriminantes matemáticas, abordagens estatísticas assintóticas ou modelos de pertinência nebulosa. Os supervisionados são mais precisos por levarem em consideração informações (verdade terrestre) fornecidas pelo usuário (Saebo et al, 1985). Os mesmos são ainda divididos em dois subgrupos: os clássicos - usam a informação contida no próprio elemento de imagem (“pixel”) para a classificação do mesmo - e os contextuais - atribuem uma das classes disponíveis de acordo com a análise do “pixel” e dos seus vizinhos (Moraes, 1992). Há ainda o conceito de pós-classificação ou “classificação em fases”, onde é empregada uma metodologia para aumentar a acurácia da classificação gerada por outra.

Outra forma de se classificar uma imagem é a classificação visual, que constitui uma metodologia de análise para quem não dispõe de nenhum recurso computacional. Trata-se da Classificação ou Interpretação Visual da Imagem (I.V.I.) (Rudorf et al., 1986). A imagem é fornecida pelo Instituto de Pesquisas Espaciais - INPE - apresentando características, tais como tamanho, cor, data, etc, que melhor se adequem à necessidade do usuário. Reconhece-se, então, o que existe na imagem, extraíndo dela informações de interesse. Os fotointerpretadores, especialistas que realizam a classificação visual, são classificados em três grupos: os que possuem conhecimento de campo, os que não possuem este conhecimento e os que não possuem este conhecimento, mas que possuem fontes de informação adicionais. Os primeiros têm maior facilidade para interpretar as imagens que os demais, pois os do terceiro grupo podem ter dificuldades em cruzar as informações de que

dispõem com a imagem, e os do segundo, dificilmente conseguirão fazê-lo sem o uso de chaves de interpretação - conhecimentos utilizados que são estruturados como se fossem formulações para se resolver um problema. É de grande utilidade que as chaves de interpretação sejam feitas de forma invariante, não mudando de um fotointerpretador para outro. Para isso, criam-se legendas que funcionam como critérios de separação das diversas classes, auxiliando uma leitura mais rápida.

As metodologias de classificação assistida por sistema especialista surgiram para tentar automatizar a classificação visual. Desse modo, o processo decisório pode ser suficientemente rico e envolver uma quantidade razoavelmente pequena de informação, podendo, então, o conhecimento ser armazenado em termos de regras, ser aplicável e poder ser tratado de forma simplificada (James, 1985). A desvantagem que essa metodologia possui é que o processo decisório não irá responder corretamente fora dos padrões para os quais foi delineado, ao contrário dos métodos exclusivamente numéricos.

2. **O**bjetivo

Neste projeto procura-se estudar alguns dos métodos de classificação numéricos, a metodologia da classificação visual e alguns métodos de classificação assistidos por sistema especialista. Dentre os métodos numéricos, objetiva-se estudar os classificadores da Máxima Verossimilhança, da Distância Mínima, do K-NN (ou K-vizinhos) e o Classificador pelo Método do Paralelepípedo. Dentre os classificadores assistidos por sistemas especialistas, serão estudados o sistema ICARE (Desachy, 1991) e o de Moraes (1998).

Além do aspecto teórico, a bolsista tem o objetivo de estudar o Sistema de Processamento de Imagens Khoros, sobretudo a “Classify Toolbox”, contida neste, e que implementa alguns dos classificadores mencionados acima. Através desta “toolbox” serão comparados os diversos métodos estudados.

O objetivo do trabalho foi plenamente atingido de acordo com o plano de trabalho e com o cronograma em anexo, sendo que foram classificadas poucas imagens pelo classificador do KNN, devido ao grande tempo levado pelo computador (processador Pentium 166 Mhz e 64Mbytes de memória RAM) para processar essas imagens (da ordem de dias) e devido à instabilidade nas redes de energia elétrica do CCEN, o que nos levou a uma análise menos profunda do mesmo.

Este trabalho possibilita a bolsista obter um bom grau de conhecimento na área de Classificação de Imagens, podendo ser aplicado em seus estudos e pesquisas futuros.

3. *M*etodologia

3.1 Sensoriamento Remoto

Ao conjunto de técnicas que permitem a exploração de um objeto sem contato físico dá-se o nome de *Sensoriamento Remoto* (Garcia, 1982), onde essa exploração é realizada através de sensores, tais como câmaras fotográficas, de televisão, de satélites, etc.

Os registros das primeiras fotos aéreas datam da segunda metade do século passado, quando eram tiradas a partir de balões (Garcia, 1982). Deste então, esse tipo de estudo foi sendo aperfeiçoado e desenvolvido para muitas áreas de aplicação, além das pesquisas com fins militares.

O Sensoriamento Remoto possui várias áreas de aplicação, sendo as principais meteorologia, geologia, militar, agricultura, indústria, ecologia e demografia. Dependendo dessas áreas, existem vários níveis de sofisticação dos sistemas de sensoriamento, com precisão e resolução determinados.

As imagens de radar não têm ainda seu potencial civil devidamente explorado. Já as imagens de satélites, possuem um potencial de utilização muito grande, e sejam, talvez, as mais exploradas atualmente, tornando-se, assim, nosso objeto de estudo.

3.2 Considerações Sobre as Imagens

O termo *Sensoriamento Remoto* se refere aos métodos que utilizam a energia eletromagnética para a detecção e medida de objetos, incluindo a luz, o calor e as ondas de rádio. A energia eletromagnética se propaga na forma de ondas harmônicas, com a velocidade da luz, sendo detectada através de sua interação com a matéria, o que produz mudança nas suas propriedades. Os sensores detectam e registram estas mudanças, e as imagens e dados resultantes são analisados com vistas a identificar as características da matéria que produziu tais modificações.

Os sensores se classificam em ativos - aqueles que emitem um sinal artificial e medem a energia refletida deste pelos objetos - e passivos - que medem as radiações

naturalmente refletidas ou emitidas pelos mesmos. No primeiro caso estão as câmaras fotográficas com ‘flash’ e os radares. No segundo, os sensores de satélites.

Os sensores não captam todo o espectro eletromagnético, mas registram algumas faixas de frequência, que correspondem a bandas do espectro. O olho humano, por sua vez, consegue captar uma faixa muito estreita deste espectro, cabendo aos sensores a detecção das demais faixas não visíveis, que são codificadas de modo a tornar possível a visualização de uma imagem. As imagens são, portanto, registros de cenas focalizadas por sensores formada pela radiação eletromagnética com frequências perfeitamente definidas do espectro (Garcia, 1982).

3.3 Características Gerais das Imagens

As principais características de uma imagem são:

- **Escala**

Escala é a razão entre a distância na imagem e a distância correspondente no terreno (Garcia, 1982).

As escalas das imagens são determinadas de acordo com o ângulo de visada do sistema de sensoriamento, altura de onde é obtida a imagem e de acordo com o fator de ampliação empregado na reprodução da imagem.

- **Contraste**

Contraste é definido pela razão entre as partes mais brilhantes e mais escuras de uma imagem. É definida a razão de contraste como o brilho máximo dividido pelo brilho mínimo da imagem. Esta também pode descrever a relação entre o brilho de um objeto e o brilho da circunvizinhança.

Imagens de baixo contraste dificultam muito o estudo da área, e independentemente da causa, seja o próprio contraste do terreno, distorções pela atmosfera ou insuficiência do sistema de sensoriamento, podem ter o contraste ampliado por métodos fotográficos e digitais (Garcia, 1982).

- Resolução

Resolução é a capacidade de distinção entre dois objetos muito próximos numa imagem. Assim em uma imagem de baixa resolução, esses objetos aparecem como um só.

É o poder de resolução do sistema de imageamento que determina a resolução das imagens produzidas pelo mesmo.

3.4 Elementos de Reconhecimento

Conhecimentos básicos de interpretação de imagens são essenciais no reconhecimento e na extração de dados e informações de uma imagem.

- Tonalidade e cor

A tonalidade é usada para cada sombra distinta entre o branco e o preto, no entanto os foto-intérpretes possuem o conceito mental para tons cinza e áreas numa imagem são caracterizadas como de tonalidade clara, média ou escura.

A tonalidade das fotografias aéreas é determinada pela sua habilidade em refletir a luz incidente, já nas imagens obtidas de sensores, a tonalidade é definida por outras propriedades físicas.

Com a utilização de filmes especiais capazes de registrar em camadas diferentes o azul, o vermelho e o amarelo, obtém-se o filme colorido. Este ajuda significadamente no processo foto-interpretativo, já que o olho humano distingue um número muito maior de cores do que tons diferentes de cinza (Garcia, 1982).

- Textura

A textura é formada pela repetição tonal de grupos de objetos por vezes muito pequenos de serem vistos individualmente. Ela se apresenta de muitas formas, podendo ser grosseira, fina, áspera, aveludada, etc (Garcia, 1982). É muito usada para caracterizar a vegetação e o solo.

- Padrão

O padrão é uma característica de objetos feitos pelo homem ou de ocorrências naturais. Uma textura é constituída quando os elementos de reconhecimento do padrão se tornam muito pequenos.

- Tamanho

O tamanho dos objetos varia com a escala da imagem e serve para diferenciar objetos de forma idêntica. Pela tamanho, os foto-intérpretes são capazes de retirar alguns objetos da imagem, de modo a facilitar o estudo.

- Forma

A forma é muito importante na identificação de objetos, sendo uma expressão topográfica ou de contornos (Garcia, 1982).

- Sombra

As sombras podem ser bem vistas em um imagem ou não. Elas são importantes para realçar a forma dos objetos e muitas vezes usadas para cálculo da altura destes. No entanto, em algumas imagens, constituem um empecilho na visualização de outros objetos.

- Localização

É importante o conhecimento da localização de certos objetos com relação ao relevo, ajudando a conhecer a região como um todo.

Com os conhecimentos acima, para a interpretação de uma imagem convém estudar uma associação de evidências ou elementos de reconhecimento, de modo que se identifiquem os elementos da imagem em estudo.

3.5 Satélites Landsat

Os *satélites Landsat* foram lançados desde 1972, com o objetivo de se obter uma ferramenta prática no estudo e no manejo dos recursos naturais da terra (Moraes, 1992).

3.5.1 Características

Os satélites Landsat têm três metros de altura, um metro e meio de diâmetro e quatro metros de largura, possuindo os painéis solares estendidos.

Sua órbita é quase circular, com altitude entre 880 e 940 quilômetros, sol-síncrona (assegurando condição permanente de iluminação, embora isto não assegure iluminação uniforme ao longo do curso), e quase polar, com o objetivo de manter a inclinação de aproximadamente 99° com o plano equatorial, para que seus sensores tomem uma imagem de uma mesma área com o mesmo ângulo de inclinação solar. Deste modo, os satélites possuem equipamentos que mantêm seu posicionamento, com parâmetros orbitais ajustados para que faixas adjacentes se superponham 10% de cada lado.

Os satélites passam pelo equador a cada 103 minutos, percorrendo a 14ª órbita em 24 horas, quando a próxima seqüência tem início. Assim sendo, os mesmos percorrerão a superfície da terra em 18 dias, sendo alguns satélites defasados em 9 dias.

Os satélites Landsat possuem dois sensores:

- Imageador multiespectral (MSS)
- Conjunto de três câmaras Vidicon (RBV)

Além disso, possui dois gravadores de banda larga, que gravam dados de áreas fora da ação de estações terrestres rastreadas, um sistema de coleta de dados, que coleta dados transmitidos por bases de coleta de dados localizados em pontos estratégicos da superfície terrestre, e outros equipamentos complementares que transmitem os dados coletados pelos sensores e informações relativas ao funcionamento dos mesmos e do próprio satélite.

O MSS é o mais importante sensor de bordo, por ser o responsável pela imagem da superfície terrestre simultaneamente em quatro bandas do espectro (bandas 3, 4, 5 e 7).

Esse equipamento possui um espelho que imageia a superfície do terreno em 33 milissegundos de oeste para leste, com uma frequência de 13,62 hertz, movendo-se para o sul registrando linhas não exatamente perpendiculares à órbita. Seis linhas de dados de

comprimento de 185,2 quilômetros são imageadas de cada vez por banda. 24 a 26 detetores são usados para registrar essas seis linhas em quatro bandas (Moraes, 1992) - figura 1.

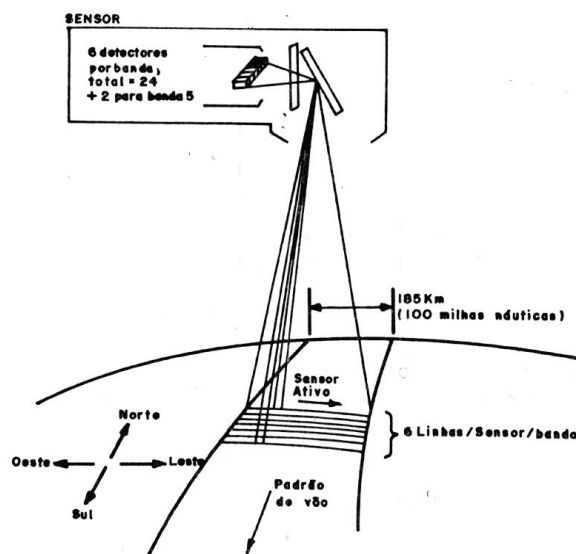


Figura 1: Obtenção da imagem pelo Landsat

3.5.2 Formação das Imagens do MSS

Quando o espelho imageador visa o terreno, cada detetor produz uma voltagem amostrada a cada 9,95 milissegundos. Logo ao longo de uma linha de 185,2 km são tomadas aproximadamente 3300 amostras. Assim a visão instantânea de campo (VIC) de 79 x 79 metros, move-se de 56 metros entre cada amostragem, resultando em um elemento de imagem, “pixel”, de área 56x79 m². Essa redução de área ocorre devido ao recobrimento, ou seja, uma imagem apresenta dados de outra imagem, e para evitar distorções, esses dados são arrançados como se tivessem sido tomados de uma área de 56x79 metros (Moraes, 1992) - figura 2.

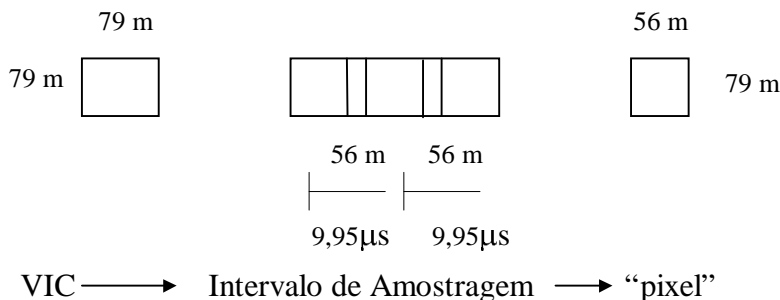


Figura 2: Obtenção de um “pixel”.

Os objetos são detectados em um terreno dependendo da relação entre o tamanho do objeto e seu brilho, em relação ao brilho da vizinhança. Se um objeto for menor que a VIC, ele só será detectado se o seu valor de brilho ou índice de luminosidade se sobrepor ao da sua circunvizinhança. Um objeto de tamanho igual ao do VIC, além de precisar ter um índice de luminosidade superior ao da circunvizinhança, a VIC precisa sobrepor-se exatamente sobre ele, o que é difícil de acontecer. Também se um objeto menor que a VIC for colocado na área de recobrimento de dois “pixels”, ele pode influenciar a luminosidade dos dois, e também é possível que dois objetos possam ser detectados em uma banda espectral e não serem detectados em outra banda.

Além das possibilidades apontadas acima, pode ocorrer o que se chama ‘confusão espectral’, onde um objeto que tem uma curva espectral definida, tem esta curva alterada na presença de outro objeto. A figura 3 abaixo ilustra melhor este fato.

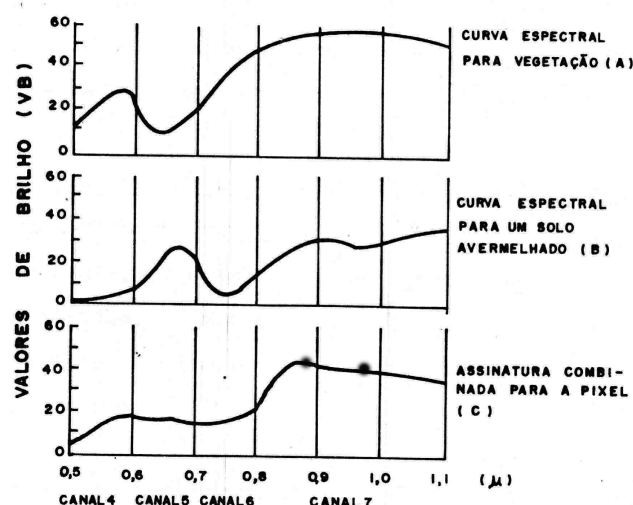


Figura 3: Confusão espectral.

3.5.3 Características das Imagens Obtidas do Satélite

Os dados fornecidos pelo MSS, RBV, e as demais informações do sistema, após ser transmitidas para a terra, são recebidas pela estação de recepção do INPE (Instituto de Pesquisas Espaciais) em Cuiabá, MT. Depois de gravados são processados pela estação de processamento, sendo convertidos em imagens fotográficas e fitas compatíveis com computador.

A imagem Lansat típica na escala 1:1000000 não é um retângulo porque durante o tempo em que é imageado, a Terra girou, devido à sua rotação; além disso, as linhas de latitude e longitude fazem um certo ângulo com o topo e a base da imagem, por causa dos 90° de inclinação da órbita do satélite.

As imagens podem apresentar-se em diferentes escalas e podem ser coloridas ou não. Em geral a resolução varia de 200 a 250 metros no terreno, podendo ser registradas feições menores como estradas e canais. As escalas utilizadas variam de 1:1000000 a 1:250000, embora se possam utilizar ampliações maiores.

3.5.4 Efeitos Atmosféricos que Alteram a Qualidade das Imagens

A radiação solar se propaga livremente pelo espaço, sofrendo dispersão e absorção ao encontrar a atmosfera terrestre.

A atmosfera dispersa alguns comprimentos de onda, acrescentando luminosidade à paisagem, sendo os elementos que mais influenciam o vapor de água e gases dispersos. Já a absorção atmosférica afeta principalmente a faixa infravermelha do espectro, retirando luminosidade da paisagem, tornando as imagens mais difíceis de serem corrigidas (Moraes, 1992). Um problema comum ocorre na presença de nuvens e poluição afetando a qualidade da luz refletida, mudando a assinatura espectral.

3.6 Processamento Digital de Imagens (PDI) em Sensoriamento Remoto

Codificar uma imagem significa representá-la como uma matriz de pequenos quadrados (“pixel”), contendo, cada um, um valor de sombreamento proporcional a reflectância da imagem na respectiva posição, o qual é atribuído uma cor ou níveis de cinza na geração da imagem.

Na figura 4 encontra-se um fluxograma que fornece uma visão global do PDI, sendo descrito adiante cada um de seus pontos.

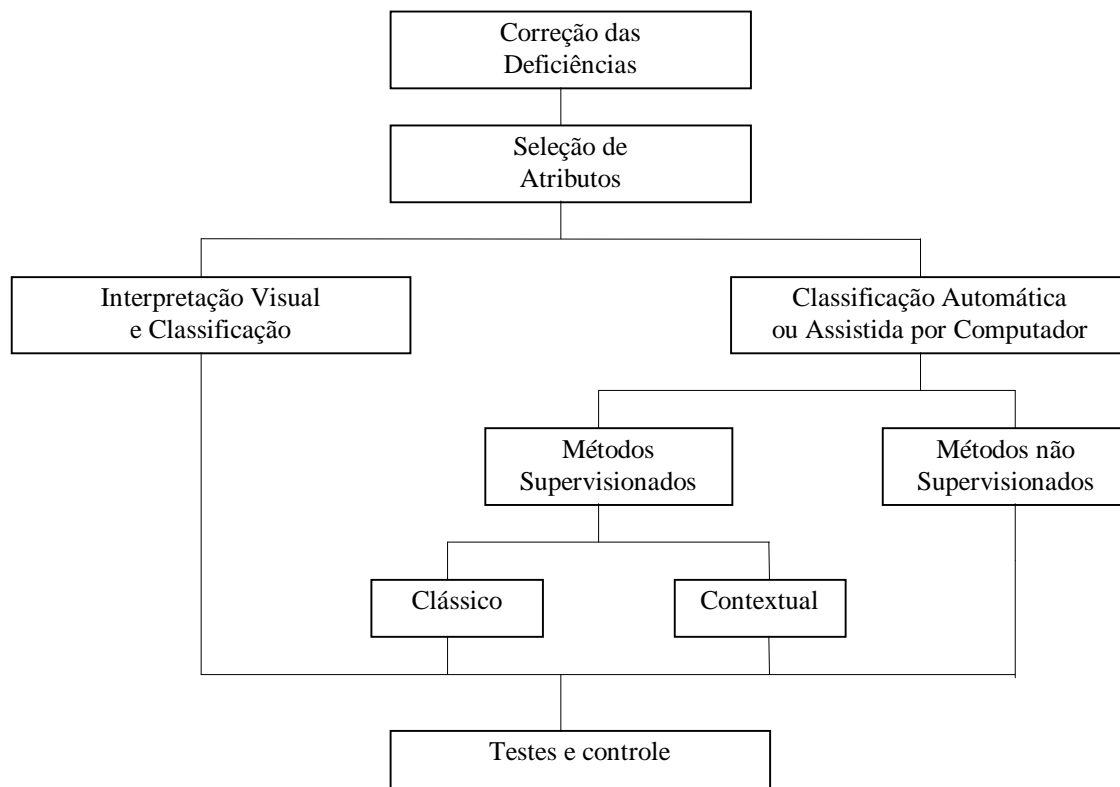


Figura 4: Fluxograma do PDI em Sensoriamento Remoto

3.6.1 Recursos do PDI

As imagens passam primeiramente por uma correção de deficiências, onde elas podem ser modificadas ou restauradas, através dos seguintes recursos:

3.6.1a Restauração e eliminação do ruído

Em imagens cuja resolução é alta ou em condições adversas, como por exemplo movimento relativo objeto-sensor e turbulência, são verificadas distorções, borramentos e outras imperfeições.

A restauração destes problemas na imagem se dá com o uso de técnicas obtidas de modelos matemáticos, estatísticos ou físicos. Uma das técnicas mais usadas para remoção do ruído aleatório consiste na filtragem da imagem, que pode ser feita por 'software' ou 'hardware', e que faz com que o ruído presente na imagem seja removido (Moraes, 1992).

3.6.1b Realce de contraste

Este recurso é utilizado para facilitar a visão humana na distinção dos níveis de sombreamento muito próximos numa imagem. Pode-se modificar as tonalidades de sombreamento com o uso de filtros lineares, não lineares e técnicas de equalização de histograma (Moraes, 1992). Também usa-se muito o recurso da falsa cor, que consiste em atribuir para cada nível de sombreamento uma cor, independente da sua cor original, aumentando, assim, a capacidade de distinção de pontos, pela vista humana, numa imagem.

3.6.1c Incorporação de informações

Pode-se incorporar informações adicionais ao “pixel” sobre a topografia do terreno, o que é feito através dos Modelos Digitais de Terreno (Moraes, 1992). Tem aplicação para construção de imagens em perspectiva, estudos de luminosidade, projetos e estudo e previsão de catástrofes.

Após a correção das deficiências, geralmente é necessária uma compressão dos dados, ou uma seleção de atributos, para que se possa estudar a imagem a níveis aceitáveis pelo usuário e pelo equipamento, o que significa diminuição da dimensão do vetor de atributos, ou seja, diminuir o espaço de todas as medidas feitas pelos sensores do satélite.

São utilizados dois conjuntos de técnicas estatísticas com esta finalidade. Um deles é constituído de transformações no espaço dos atributos, e outro usa medidas de distância.

No primeiro caso, reduz-se a dimensão preservando ao máximo a representação de padrões e/ou a discriminação entre as classes de atributos. Consiste de duas técnicas: Técnica dos Componentes Principais (Critério do Erro Médio Quadrático) e Critério da Entropia (Moraes, 1992).

No segundo caso, no qual usa-se medidas de distância, reduz-se a dimensão minimizando a probabilidade de erro. São dois os critérios mais utilizados que fornecem, realmente, limites inferiores e superiores para tal probabilidade: Critério da Divergência e Distância J-M (Jeffries-Matusita) (Moraes, 1992).

Depois da seleção de atributos, a imagem está pronta para ser classificada. Existem dois métodos de se classificar uma imagem: - a Classificação Visual, em que a imagem é interpretada através de especialistas - e a Classificação Automática ou Assistida por

Computador ou, ainda, Classificação Numérica. Os métodos de classificação numéricos podem ser não-supervisionados ou supervisionados, este último se subdividindo em outros dois métodos: Métodos Clássicos e Métodos Contextuais.

Os Métodos Não-Supervisionados dividem as classes automaticamente, de acordo com os parâmetros estabelecidos na sua implementação. Geralmente baseiam-se no Princípio de Fisher, mas também podem ter base em funções matemáticas, abordagens estatísticas assintóticas e modelos de pertinência nebulosa (Moraes, 1992).

Os Métodos Supervisionados, levam em conta informações adicionais fornecidas pelo usuário e requerem conhecimentos prévios das classes, e por isso são mais precisos que os métodos de classificação não-supervisionados. Têm base na teoria de Bayes (Moraes, 1992).

O Método Supervisionado Clássico classifica cada elemento da imagem levando em consideração as informações contidas nele. Já o Método Contextual, além de levar em conta as informações dos “pixels”, considera também as informações da vizinhança. Isto é feito introduzindo-se um fator de correção contextual (Moraes, 1992).

Os métodos de classificação supervisionados necessitam de estimativas para os parâmetros das funções discriminantes. Geralmente são necessários conhecimentos prévios, denominados verdades terrestres, que consiste em informações retiradas de áreas de treinamento na própria imagem, fotografias aéreas, mapas, ou investigação direta no campo, e são na realidade “pixels” sobre os quais se conhecem as verdadeiras classes. Usa-se, então, estes “pixels” para treinar o classificador, de modo que este reconheça “pixels” de estrutura semelhante.

A estimativa dos parâmetros das classes e da densidade de probabilidade para o vetor de atributos requerem tipos diferentes da verdade terrestre. No caso de classificadores contextuais, as funções discriminantes necessitam da verdade terrestre dos pontos e de seus vizinhos.

Finalmente, após a classificação da imagem de interesse, a mesma é submetida a testes e controle, que observam os “pixels” classificados, comparando-os com a verdade terrestre. Seja qualquer método de classificação, existirá sempre uma margem de erro, ou seja, aparecerão sempre dúvidas em relação à classificação obtida.

Seria ideal a verificação do local visado na imagem. Como isto nem sempre é viável, procura-se comparar imagens obtidas de um mesmo local em diferentes épocas e em diferentes bandas do espectro.

É sempre recomendável o melhor aproveitamento e utilização de recursos e a maior sofisticação possível, quando se faz um processamento de uma imagem de grande interesse, garantindo, com isso, a minimização dos erros cometidos.

3.7 Análise Classificatória

Um problema geralmente enfrentado na classificação de imagens obtidas através de sensoriamento remoto, é a necessidade de decidir a qual dos possíveis grupos um elemento da imagem pertence. Na maioria das vezes os processos de classificação existentes se baseiam em observações. No entanto, o acesso à computação tem facilitado os processos de classificação, servindo, pelo menos, de guia e ajuda para muitos profissionais da área. Esses programas de computação são geralmente implementados com base na teoria da probabilidade, e se constituem basicamente em análises estatísticas.

Como já ressaltado, uma análise classificatória é dirigida aos problemas de assinalar um objeto para um dos possíveis grupos de objetos existentes, com base em observações feitas do objeto, apesar de nada se ter dito sobre o conhecimento dos diferentes grupos existentes.

No caso de classificação de imagens, objeto de nosso estudo, temos um problema de análise classificatória, em que procuramos classificar um pixel em uma das classes que constituem a imagem.

3.7.1 Revisão de Algumas Noções de Probabilidade

A probabilidade de um evento “A” ocorrer será denotada por $P(A)$; também denotamos por $P(A/B)$ a probabilidade condicional, que representa a probabilidade de ocorrência do evento “A”, dado a ocorrência do evento “B”. A probabilidade condicional pode ser entendida como um modo de se utilizar um conhecimento da situação, de modo a aperfeiçoar a estimativa da probabilidade da ocorrência do evento. Desse modo, novos

conhecimento contribuem para melhorar a probabilidade inicial, onde as novas probabilidades condicionais são designadas probabilidades posteriores.

É importante salientar que a presença de informações adicionais podem aumentar o valor da probabilidade inicial, diminuí-la ou não alterá-la.

Dois eventos A e B são independentes se :

$$P(A) = P(A/B)$$

Isto significa que a ocorrência de um dos eventos não afeta a ocorrência do outro.

3.7.2 Regras de Classificação

Deseja-se uma regra de classificação que minimize o erro total de classificação, o qual deve ser entendido como a probabilidade de que a regra considerada desclassifique um objeto (James,1985).

3.7.2a Regra de Bayes

Classifica-se um objeto pela regra de Bayes, o qual assinala um objeto para o grupo que possuir a maior probabilidade condicional. Formalmente:

Se existem G grupos, então, um objeto pertence ao grupo i quando:

$$P(G_i/X) > P(G_j/X) \quad \forall \quad i \neq j$$

Se existir mais do que um grupo com a maior probabilidade condicional, o objeto será classificado em um desses grupos.

A regra de Bayes, é então, a regra ótima que se procurava, não sendo difícil provar que é a regra que minimiza o ETC (James,1985).

No entanto, na prática, a probabilidade $P(G_i/X)$ é difícil de ser encontrada por estimativas simples, enquanto que a probabilidade $P(X/G_i)$ pode ser estimada tomando uma amostra de objetos do grupo i . Esta última, representa a probabilidade de se tomar a variável X, dado o objeto ser do grupo i.

Desse modo, relaciona-se essas duas probabilidades, pelo Teorema de Bayes:

$$P(G_i/X) = \frac{P(X/G_i).P(G_i)}{\sum_{\forall i} P(X/G_i).P(G_i)}$$

em que $P(G_i)$ é a probabilidade do objeto ser do grupo i ou seja, a proporção do grupo i , na população.

Assim, a regra de Bayes fica:

Classifica-se o objeto para o grupo i se :

$$\frac{P(X/G_i).P(G_i)}{\sum_{\forall k} P(X/G_k).P(G_k)} > \frac{P(X/G_j).P(G_j)}{\sum_{\forall k} P(X/G_k).P(G_k)} \quad \forall \quad i \neq j$$

$$\Rightarrow P(X/G_i).P(G_i) > P(X/G_j).P(G_j) \quad \forall \quad i \neq j$$

que é a regra prática final de Bayes.

3.7.2b Utilização da Regra de Bayes

Apesar da regra de Bayes ser a regra ótima para resolver o problema de classificação, é muitas vezes inviável, por necessitar de grande volume de dados coletados, para estimar a probabilidade $P(X/G_i)$ (James, 1985).

Grande parte da teoria de classificação é interessada em achar formas práticas desta regra que é apropriada em condições especiais ou encontrar aproximações simples da regra que tem os menores erros aceitáveis.

3.7.2c Divisões do Espaço Amostral

É geralmente interessante visualizar como uma regra de classificação trabalha, através de diagramas de dispersão.

Uma regra de classificação divide o diagrama ou melhor, o espaço amostral em um número de regiões, que correspondem aos diversos grupos existentes.

A simplicidade e fácil utilização da regra de classificação está diretamente relacionada com a simplicidade da divisão do espaço amostral, ou vice-versa. Geralmente a divisão do espaço amostral fornece uma idéia das propriedades dos grupos colocados juntos, de algum modo, sugerindo que a regra de classificação divide o espaço amostral, de maneira simples (James,1985).

Quando mais do que duas propriedades do grupo estão envolvidas (ou seja, quando temos mais do que duas variáveis aleatórias), o espaço amostral é em três dimensões ou mais, tornando difícil a visualização do espaço envolvido.

3.7.2.4 Classificação Prática - O Caso Normal

A regra de Bayes, no caso em que as propriedades ou variáveis usadas tenham distribuição normal, resulta numa regra de classificação simples de ser aplicada na prática.

3.7.2.4.1 A Normal Multivariada

A distribuição normal descreve a probabilidade simples, que é denotada em termos de dois parâmetros: sua média (μ) e seu desvio (σ).

A função densidade normal corresponde a equação (James,1985):

$$f(X) = \left(\frac{1}{\sqrt{2\pi}\sigma} \right) \exp \left[- (X - \mu)^2 / 2\sigma^2 \right]$$

Desse modo a média controla a localização e seu desvio, a largura da distribuição.

Quando mais do que uma variável é envolvida, então a distribuição normal multivariada deve ser usada, sendo uma generalização da distribuição normal.

Assim, em vez de uma média simples controlando a localização da distribuição, teremos um “vetor média”, formado pelas médias das variáveis envolvidas. O desvio equivalente na distribuição multivariada é representado por uma matriz de covariâncias.

No caso da normal multivariada, para duas variáveis X_1 e X_2 (ou seja, uma normal bivariada), o vetor média e a matriz de covariâncias são representados por:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

onde: μ_1 e μ_2 são as médias respectivas de X_1 e X_2 . Os elementos da diagonal da matriz de covariâncias, σ_{11} e σ_{22} são simplesmente os desvios de X_1 e X_2 , respectivamente. Os elementos σ_{12} e σ_{21} são na realidade iguais e representam a “associação” entre as duas variáveis.

Na figura 5, apresenta-se uma distribuição circular no caso em que $\sigma_{11} = \sigma_{22}$ e $\sigma_{12} = \sigma_{21} = 0$:

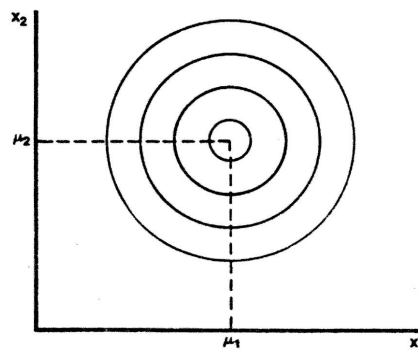


Figura 5: A normal bivariada com $\sigma_{11} = \sigma_{22}$ e $\sigma_{12} = \sigma_{21} = 0$

À medida que σ_{12} aumenta (ou seja, aumenta a associação entre as variáveis), a distribuição se torna - figura 6:

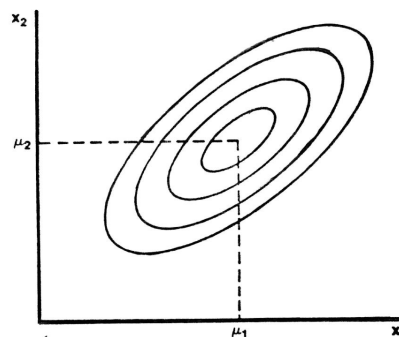


Figura 6: A normal bivariada com $\sigma_{12} > 0$

onde as variáveis variam juntas.

Se σ_{12} é negativo, a tendência das variáveis é variar em sentidos opostos, apresentando distribuição da forma mostrada na figura 7.

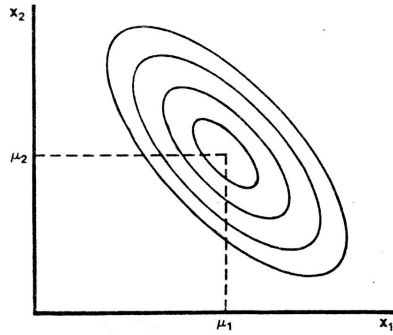


Figura 7: A normal bivariada com $\sigma_{12} < 0$

Dessa forma, a matriz de covariâncias é responsável pela forma da distribuição, onde σ_{11} e σ_{22} afetam a distribuição na direção de X_1 e X_2 e σ_{12} afeta a distribuição entre as duas direções.

Os diagramas apresentados acima representam a distribuição normal bivariada, que por sua vez fornece a probabilidade de que algum valor de X_1 e X_2 ocorra. As linhas desenhadas conectam os pontos que têm iguais probabilidades de ocorrência, resultando em uma série de elipses concêntricas, centradas no vetor média.

Ao se estudar a distribuição normal de mais de duas variáveis, encontra-se o problema de visualização desta distribuição em mais de duas dimensões.

Particularmente, três variáveis geram superfícies de iguais probabilidades em três dimensões, onde as elipses tomam forma de elipsóides. Neste caso temos:

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} \quad \Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ & \sigma_{22} & \sigma_{23} \\ & & \sigma_{33} \end{bmatrix}$$

Do mesmo modo que no caso anterior, os elementos $\sigma_{11}, \sigma_{22}, \sigma_{33}$ afetam a distribuição na direção de X_1, X_2, X_3 , respectivamente e $\sigma_{12}, \sigma_{13}, \sigma_{23}$ afetam a orientação do elipsóide.

A equação da distribuição normal multivariada na forma matricial é (James,1985):

$$f(X) = \left(\frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \right) \exp \left[-\frac{1}{2} (X - \mu)' \Sigma^{-1} (X - \mu) \right]$$

onde X é o vetor coluna das variáveis aleatórias.

Um diagrama de dispersão dos dados oriundos de estruturas normais bivariadas deve ser de forma elíptica, com densidade de pontos aumentando em torno da média. Do mesmo modo, um diagrama de dispersão de dados em três dimensões deve conter grupos elipsóides, e assim por diante. No entanto, a visualização de um diagrama de dispersão qualquer não prova que os dados são obtidos através de uma distribuição particular (James,1985).

3.7.2.4.2 A Forma Normal da Regra de Bayes

Se dentro de cada grupo as variáveis que satisfazem a propriedade vetorial X, possuem distribuição normal multivariada, a seguinte forma é conhecida (James,1985):

$$P(X/G_i) = \frac{1}{(2\pi)^{n/2} |\Sigma_i|^{1/2}} \exp \left[-\frac{1}{2} (X - \mu_i)' \Sigma_i^{-1} (X - \mu_i) \right]$$

Desse modo, estimar $P(X/G_i)$ significa estimar dois parâmetros para cada grupo, o vetor média (μ_i) e a matriz de covariâncias (Σ_i).

Se existem n propriedades em X, então existem n grupos média no vetor média e $n(n+1)/2$ elementos na matriz de covariâncias, somando um número grande de quantidades a serem estimadas para cada grupo, mas que é consideravelmente pequeno em relação ao número de estimativas requeridas numa aplicação direta da regra de Bayes (James,1985).

Substituindo a forma normal $P(X/G_i)$ na regra de Bayes anteriormente estabelecida, tem-se:

Classifica-se o objeto para o grupo i , se:

$$\begin{aligned} & \frac{P(G_i)}{(2\pi)^{n/2} |\sum_i|^{1/2}} \exp\left[-\frac{1}{2}(X - \mu_i)' \sum_i^{-1} (X - \mu_i)\right] > \\ & > \frac{P(G_j)}{(2\pi)^{n/2} |\sum_j|^{1/2}} \exp\left[-\frac{1}{2}(X - \mu_j)' \sum_j^{-1} (X - \mu_j)\right] \quad \forall i \neq j \end{aligned}$$

Tomando o logaritmo natural de ambos os lados, tem-se:

$$\begin{aligned} & -\frac{1}{2}n \ln(2\pi) - \frac{1}{2} \ln|\sum_i| - \frac{1}{2}(X - \mu_i)' \sum_i^{-1} (X - \mu_i) + \ln(P(G_i)) > \\ & > -\frac{1}{2}n \ln(2\pi) - \frac{1}{2} \ln|\sum_j| - \frac{1}{2}(X - \mu_j)' \sum_j^{-1} (X - \mu_j) + \ln(p(G_j)) \quad \forall i \neq j \end{aligned}$$

e ainda cancelando os termos em comum em ambos os lados, tem-se:

$$\begin{aligned} & -\ln|\sum_i| - (X - \mu_i)' \sum_i^{-1} (X - \mu_i) + \ln(P(G_i)) > \\ & > -\ln|\sum_j| - (X - \mu_j)' \sum_j^{-1} (X - \mu_j) + \ln(P(G_j)) \quad \forall i \neq j \end{aligned}$$

multiplicando ambos os lados por (-1) , tem-se:

$$\begin{aligned} & \ln|\sum_i| + (X - \mu_i)' \sum_i^{-1} (X - \mu_i) - \ln(P(G_i)) < \\ & < \ln|\sum_j| + (X - \mu_j)' \sum_j^{-1} (X - \mu_j) - \ln(P(G_j)) \quad \forall i \neq j \end{aligned}$$

As quantidades em ambos os termos são designadas marcas discriminantes, produzindo resultado final em número para cada grupo, dependendo de X .

Definimos a função discriminante por:

$$d_i(X) = \ln|\sum_i| + (X - \mu_i)' \sum_i^{-1} (X - \mu_i)$$

Assim sendo, a regra de Bayes fica:

Classifica-se o objeto para o grupo i se, e somente se:

$$d_i(X) - \ln(P(G_i)) < d_j(X) - \ln(P(G_j)) \quad \forall i \neq j$$

quando X tem distribuição normal multivariada para cada grupo.

3.7.2.4.3 Divisão do Espaço Amostral:

Para ilustrar as divisões do espaço amostral, toma-se o caso simples de dividi-lo em dois grupos, cuja variável aleatória assume os valores x_1 e x_2 .

Assim, uma região do espaço amostral que pertence ao grupo 1 é caracterizada por valores de x_1 e x_2 em que:

$$d_1(x_1, x_2) - \ln(P(G_1)) < d_2(x_1, x_2) - \ln(P(G_2))$$

A curva que divide o grupo 1 do grupo 2 pode ser obtida fazendo-se:

$$d_1(x_1, x_2) - \ln(P(G_1)) = d_2(x_1, x_2) - \ln(P(G_2))$$

onde esta forma corresponde a uma curva quadrática, de acordo com a álgebra matricial e a geometria (James, 1985). Esta curva representa os contornos de iguais probabilidades dos dois grupos.

Pode-se pensar em linhas divisórias entre os grupos em outras dimensões, onde a curva quadrática se torna uma superfície quadrática.

É essa natureza quadrática das superfícies ou curvas, que separam os grupos, de forma simples, que resultam na forma normal da Regra de Bayes, tornando-se conhecida como função discriminante quadrática.

3.7.2.4.4 Iguais Covariâncias

Se as relações entre as variáveis são independentes dos grupos, então estes possuem a mesma matriz de covariâncias e a função discriminante se reduz a (James,1985):

$$d_i(X) = -\mu_i' \Sigma^{-1} X + \frac{1}{2} \mu_i' \Sigma^{-1} \mu_i$$

Definimos a nova função $f_i(X) = -d_i(X)$, de modo que a função acima seja multiplicada por (-1). Assim:

Classifica-se o objeto para o grupo i se, e somente se:

$$f_i(X) + \ln(P(G_i)) > f_j(X) + \ln(P(G_j)) \quad \forall \quad i \neq j$$

onde:

$$f_i(X) = \mu_i' \Sigma^{-1} X - \frac{1}{2} \mu_i' \Sigma^{-1} \mu_i$$

em que o segundo termo da função acima é constante para cada grupo e representada por C_{oi} . Também a multiplicação de matrizes envolvida no primeiro termo resulta em um vetor C_i , e a função acima escreve-se:

$$f_i(X) = C_i' X + C_{oi}$$

ou

$$f_i(X) = \sum_{K=1}^n C_{Ki} X_K + C_{oi}$$

onde C_{Ki} são os elementos do vetor C_i .

No caso de duas variáveis ($n=2$), teremos:

$$f_i(X) = C_{1i} x_1 + C_{2i} x_2 + C_{oi}$$

Verificamos que esta função tem a forma de uma regressão (James,1985).

3.7.2.4.5 Discriminante Linear

Do mesmo modo a linha que divide o espaço amostral no caso simples (2 grupos/ 2 variáveis) é dada por valores de x_1 e x_2 que satisfazem :

$$f_1 (x_1, x_2) = f_2 (x_1, x_2)$$

assumindo $P (G_1) = P (G_2)$.

Esta divisão representa uma linha reta. No geral, $f_i (X)$ separa grupos por superfícies ou hiperplanos. A forma linear de $f_i (X)$ e a linha reta, ou plano que divide o espaço amostral fizeram com que $f_i (X)$ fosse designada por função discriminante linear.

No caso específico encontrado em muitos problemas de classificação, em que dois grupos apenas estão envolvidos, onde os grupos assumem covariâncias iguais, tem-se:

Classifica-se o objeto no grupo 1 se, e somente se:

$$f_1 (X) + \ln (P(G_1)) > f_2 (X) + \ln (P(G_2))$$

ou

$$f_1 (X) - f_2 (X) > \ln (P(G_2)) - \ln (P(G_1))$$

onde a diferença entre as funções discriminantes é escrita como uma nova função:

$$w (X) = f_1 (X) - f_2 (X)$$

3.8 Classificação Supervisionada

A classificação supervisionada é o procedimento mais utilizado na análise de dados em imagens de sensoriamento remoto.

O objetivo desta análise é classificarem-se “pixels” na imagem, representando um grupo particular dentre os demais grupos ou classes existentes.

Existem vários métodos de classificação supervisionados, como se vê adiante, mas todos geralmente seguem os seguintes passos:

- Decidir as classes em que a imagem deve ser segmentada, como por exemplo, água, regiões urbanas, etc.
- Escolher pixels representativos destas classes, ditos dados de treinamento, que podem ser estabelecidos através de visitas ao local, mapas, fotografias aéreas ou até mesmo através de outra imagem já interpretada. Geralmente os pixels de treinamento vão falhar nas regiões perto das bordas, chamadas rampas de treinamento (Richards, 1986).
- Usar os dados de treinamento para estimar os parâmetros a serem usados em um classificador particular. Esses parâmetros serão as propriedades do modelo probabilístico usado ou as equações que definem as partições no espaço multiespectral. O conjunto de parâmetros para uma dada classe é denominada assinatura desta classe.
- Usar o classificador treinado para classificar cada pixel da imagem em uma das classes.
- Produzir tabelas ou mapas temáticos que sumariem os resultados da classificação.

A seguir serão discutidos alguns métodos de classificação supervisionados.

3.8.1 O Classificador do Paralelepípedo

O Classificador do Paralelepípedo é um dos mais simples métodos de classificação supervisionada, em que amostras são utilizadas para extrair histogramas dos componentes espectrais individuais, a partir dos quais seus valores menores e maiores são determinados e usados para descrever um paralelepípedo ou uma caixa multidimensional, representando uma determinada classe.

Assumimos, então, que pixels que pertencem à mesma classe têm seus valores de atributos dentro de um paralelepípedo definido pelas amostras desta classe. Desse modo, um pixel X é classificado para a classe W se todos os valores do vetor X estão compreendidos entre as bordas definidas para a classe W . Na figura 8, temos um exemplo de aplicação deste método, no espaço de atributos com duas dimensões.

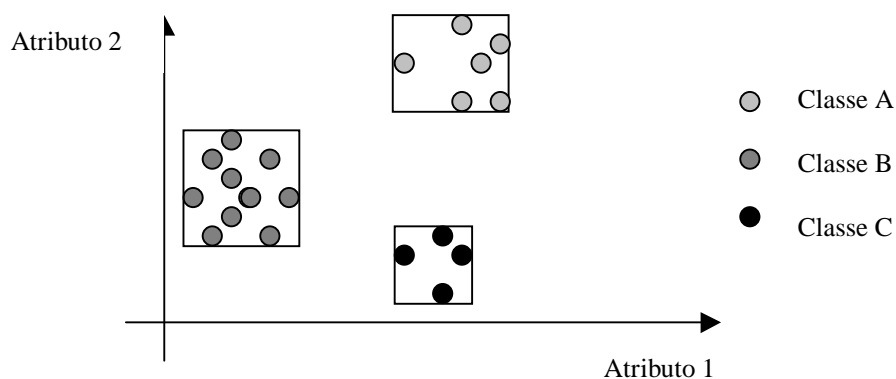


Figura 8: Exemplo de Classificação pelo método do Paralelepípedo

A classificação é feita checando-se se o pixel está dentro ou fora dos paralelepípedos para cada atributo e juntando os resultados verdadeiros/falsos. Se alguma checagem dentro/fora é falsa, o pixel não será classificado, ou seja, será rejeitado.

Uma das desvantagens deste método são as regiões de inseparabilidade dos paralelepípedos. Pixels contidos nestas regiões não serão classificados. Desse modo, este classificador pode ser usado quando sabemos que não haverá classes superpostas no espaço de atributos multidimensional. Outra limitação é que as probabilidades iniciais das classes não são consideradas. Quando existe uma superposição de classes para um atributo, como mostra a figura 9, para as classes A e C do atributo 1, outros atributos são usados para delimitar as classes. No entanto, se a superposição das classes persistir para todos os atributos, o “pixel” que estiver nesta superposição não poderá ser classificado, como mostra a figura 10.

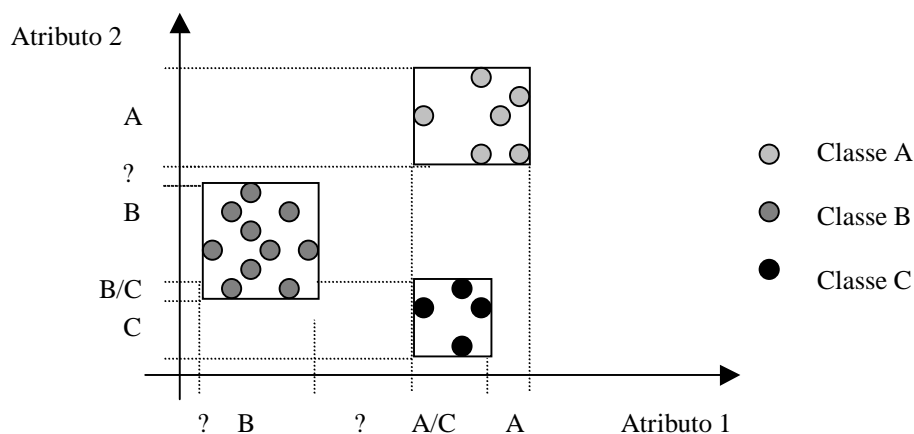


Figura 9: Classificação do Paralelepípedo com superposição de classes para 1 atributo.

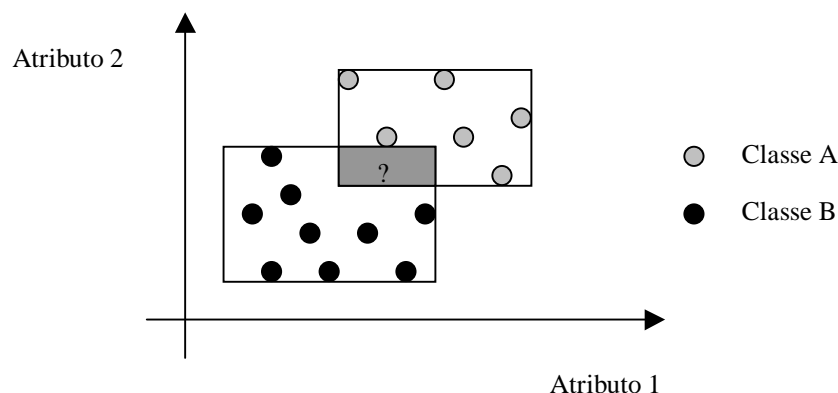


Figura 10: Classificação do Paralelepípedo com superposição de classes em todos os atributos.

Outro problema que este classificador apresenta é a determinação das bordas para as classes, pois se uma determinada classe possui valores de atributos dispersos, as bordas maiores e menores estarão bastante separadas, aumentando a chance de termos regiões superpostas para dada classe e atributo. Pode-se, no entanto, estabelecer limiares, de modo que apenas valores superiores a estes serão considerados na formação dos paralelepípedos.

As vantagens deste classificador são sua simplicidade e velocidade, já que a classificação é feita por comparação dos valores dos pixels com os valores estabelecidos pelos paralelepípedos (Santos, 1997).

3.8.2 O Classificador da Máxima Verossimilhança

Este método é o mais comum método de classificação supervisionado usado com dados de imagens de sensoriamento remoto.

O classificador da Máxima Verossimilhança utiliza informação sobre a distribuição das amostras para cada classe, de modo que determinado “pixel” será classificado de acordo com o valor médio das amostras e de acordo com a distribuição dos “pixels” de treinamento em torno da média de cada classe – figura 11.

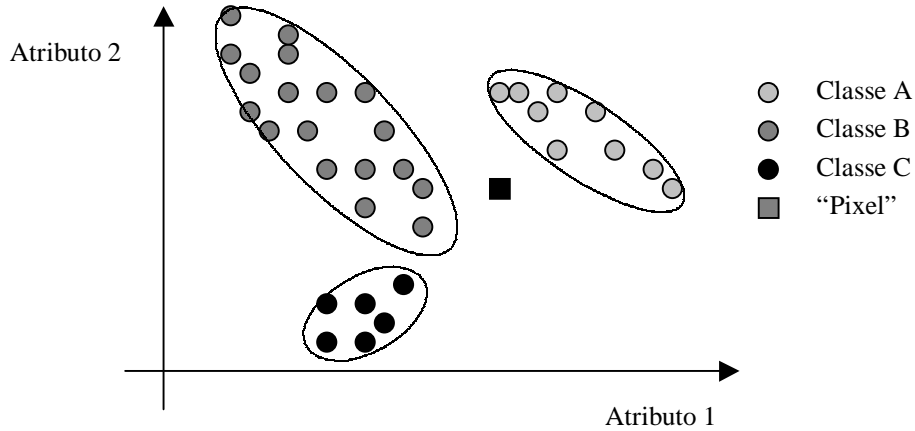


Figura 11: Classificação de um “pixel” pelo método da Máxima Verossimilhança.

Este classificador consiste em se tomar a regra de classificação de Bayes, com uma modificação, onde é estabelecida a seguinte função discriminante:

$$g_i(X) = \ln \{P(X/w_i) P(w_i)\} = \ln P(X/w_i) + \ln P(w_i)$$

onde o vetor X representa valores de brilho do pixel e os grupos agora são denominados de classes w_i da imagem.

Tem-se assim a seguinte regra de classificação:

$$X \in w_i \text{ se ,e somente se, } g_i(X) > g_j(X) \quad \forall i \neq j$$

que é a regra de decisão da máxima verossimilhança. No caso da normal multivariada esta regra de classificação corresponde à regra de Bayes discutida na parte anterior.

Um tópico embutido neste processo, e não discutido anteriormente é o uso de limiares. Os limiares são usados no caso de termos classes espectrais superpostas, de modo a evitar que sejam feitas classificações errôneas. Desse modo, amostras que têm probabilidades para todas as classes inferiores ao limiar, não são classificadas.

Na prática, os limiares são aplicados às funções discriminantes e não às distribuições de probabilidades. Dessa forma a nossa regra de decisão fica:

$$X \in w_i \text{ se ,e somente se, } g_i(X) > g_j(X) \text{ e } g_i(X) > T_i$$

onde w_i se refere à classe i , $g_i(X)$ são as funções discriminantes e T_i é o limiar estabelecido para a classe espectral w_i . De maneira a estimar T_i consideremos os seguintes fatores:

Temos uma classificação aceitável se,

$$\ln P(w_i) - \frac{1}{2} \ln |\Sigma_i| - \frac{1}{2} (X - \mu_i)' \Sigma_i^{-1} (X - \mu_i) > T_i$$

ou

$$(X - \mu_i)' \Sigma_i^{-1} (X - \mu_i) < -2T_i - \ln |\Sigma_i| + 2 \ln P(w_i)$$

O lado esquerdo desta equação tem distribuição χ^2 com N graus de liberdade, se X tem distribuição normal e onde N representa a dimensão do espaço multiespectral. Assim, pode-se consultar uma tabela de tal distribuição para determinar o valor desta expressão (Richards, 1986).

3.8.2.1 Número de “pixels” de treinamento para cada classe

As amostras de treinamento para cada classe devem ser suficientes para que se possa estimar os elementos do vetor média e da matriz de covariâncias, que devem ser determinados. Para um espaço multiespectral de dimensão N , são necessárias no mínimo $N + 1$ amostras. Segundo Swain e Davis (1978), $10 N$ amostras para cada classe espectral devem ser obtidos no mínimo, sendo $100 N$ uma quantidade ótima de amostras. Apesar disto, é importante ter em mente que devemos utilizar o maior número de pixels possíveis para treinamento (Richards, 1986).

3.8.3 O Classificador de Mahalanobis

Tomando a função discriminante para o classificador da Máxima Verossimilhança no caso particular de ter-se iguais probabilidades iniciais e iguais covariâncias, define-se a distância de Mahalanobis como:

$$d(X, \mu_i)^2 = (X - \mu_i)' \Sigma_i^{-1} (X - \mu_i)$$

Classifica-se o “pixel” para a classe i se, e somente se:

$$d(X, \mu_i)^2 < d(X, \mu_j)^2, \forall i \neq j$$

A vantagem deste classificador sobre o da Máxima Verossimilhança é a velocidade com que a classificação é realizada (Richards, 1986).

3.8.4 O Classificador da Distância Mínima

Quando o número de amostras para uma classe é limitado e insuficiente para a aplicação da regra de classificação anterior, temos de recorrer a um outro classificador que não use informações das covariâncias e dependa apenas das posições média das classes espectrais, já que para um dado número de amostras estes parâmetros podem ser melhor estimados que os outros.

O classificador da distância mínima utiliza dados de treinamento que são usados apenas para determinar as médias das classes e baseia-se na classificação do pixel na classe de média mais próxima, como mostra a figura 12. Vê-se que os “pixels” 1 e 2 de classes desconhecidas foram assinalados às classes de média (indicada com o símbolo “+”) mais próximas.

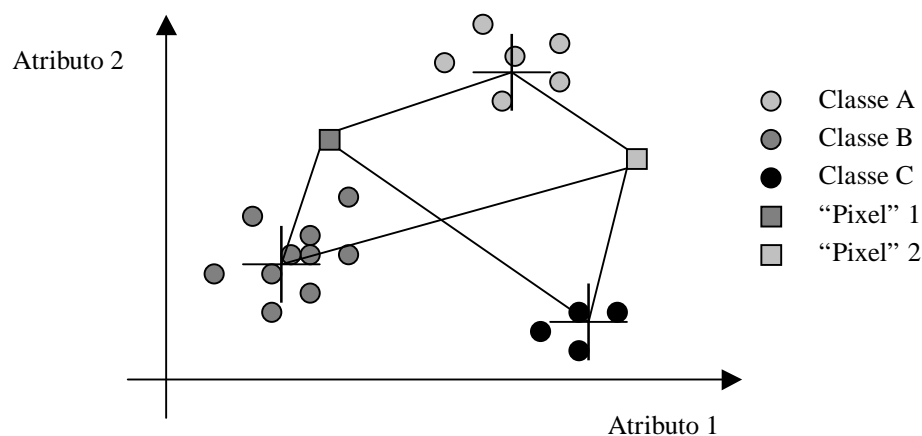


Figura 12: Classificação pelo método da Distância Mínima

Apesar de ser uma técnica mais rápida do que a técnica da classificação da máxima verossimilhança, não é tão flexível quanto esta, sendo portanto menos eficiente, porém fornecendo, ainda, resultados aceitáveis (Richards, 1986).

3.8.4.1 A Função discriminante

A função discriminante para o classificador da Distância Mínima é a Distância Euclidiana, definida como segue:

Se μ_i , $i = 1, \dots, M$ são as médias das M classes determinadas dos dados de treinamento e X é a posição do pixel a ser classificado, a Distância Euclidiana é definida na forma vetorial como:

$$d(X, \mu_i)^2 = (X - \mu_i)^t (X - \mu_i) = (X - \mu_i) (X - \mu_i)^t, \quad i = 1, \dots, M.$$

$$\Rightarrow d(x, \mu_i)^2 = X \cdot X - 2\mu_i \cdot X + \mu_i \mu_i$$

desse modo: $X \in w_i$ se $d(x, \mu_i)^2 < d(x, \mu_j)^2 \quad \forall i \neq j$

se definirmos $g_i(x) = 2\mu_i \cdot x - \mu_i \cdot \mu_i$ então:

$$X \in w_i \text{ se } g_i(X) > g_j(X) \quad \forall i \neq j$$

3.8.4.2 Degeneração do Classificador da Máxima Verossimilhança para o da Distância Mínima

A diferença básica entre estes dois classificadores é o que último classifica o pixel de acordo com sua distância à média mais próxima, sem levar em conta a direção dessa média, que é determinada das informações contidas na matriz de covariâncias, que são consideradas no classificador da máxima verossimilhança. Assim uma performance superior é esperada por este último, em geral.

Existem casos, entretanto, que o uso deste classificador não apresenta vantagem. Isto pode ocorrer na prática, quando a classe de covariâncias é dominada por ruídos (Richards, 1986).

Considerando a matriz de covariâncias de todas as classes serem diagonais e iguais e as variâncias de cada componente serem idênticas, então:

$$\sum_i = \sigma^2 I \quad \forall_i$$

Assim a função discriminante para o classificador da máxima verossimilhança fica:

$$g_i(X) = \frac{1}{2} \ln \sigma^{2n} - \frac{1}{2} \sigma^2 (X - \mu_i)^t (X - \mu_i) + \ln P(w_i)$$

onde o termo σ^{2n} é comum para todas as classes, podendo ser ignorado assim como $X \cdot X$ que resulta do produto $(X - \mu_i)^t (X - \mu_i)$. Assim:

$$g_i(X) = \frac{1}{2\sigma^2} (2\mu_i X - \mu_i \cdot \mu_i) + \ln P(w_i)$$

Se as probabilidades iniciais são iguais, então:

$$g_i(X) = 2\mu_i X - \mu_i \cdot \mu_i$$

que é a função discriminante para o classificador da distância mínima. Assim os dois classificadores são equivalentes para distribuições espectrais de classes simétricas e idênticas.

3.8.4.3 Superfícies de decisão

A superfície que separa as classes i e j é dada por:

$$\begin{aligned} g_i(X) - g_j(X) &= 0 \Rightarrow \\ 2(\mu_i - \mu_j) \cdot X - (\mu_i \mu_i - \mu_j \mu_j) &= 0 \end{aligned}$$

que define uma superfície linear ou hiperplano em mais de três dimensões. Desse modo as superfícies são menos flexíveis do que as superfícies do outro classificador, que são quadráticas (Richards, 1986).

Limiares também podem ser aplicados ao classificador da distância mínima, onde a distância limiar é especificada de acordo com o número de desvios de uma classe média.

3.8.5 O Classificador do K-NN

O classificador do KNN (K-Nearest Neighbors Classifier) é um método de classificação em que “pixels” são assinalados em classes com base no número de “pixels” vizinhos que pertencem às mesmas.

Desse modo, um algoritmo que implementa este método geralmente seleciona os K “pixels” com classes conhecidas que estão perto do “pixel” X sendo investigado e assinala X para a classe que apresenta mais “pixels” nesta vizinhança,

Na figura 13 têm-se um exemplo da aplicação deste método no espaço de atributos bidimensional. Existem três classes e os “pixels” desconhecidos 1 e 2, que serão classificados, tomando $K=7$ (ou seja, tomando os 7 vizinhos mais próximos), para as classes B e A, respectivamente.

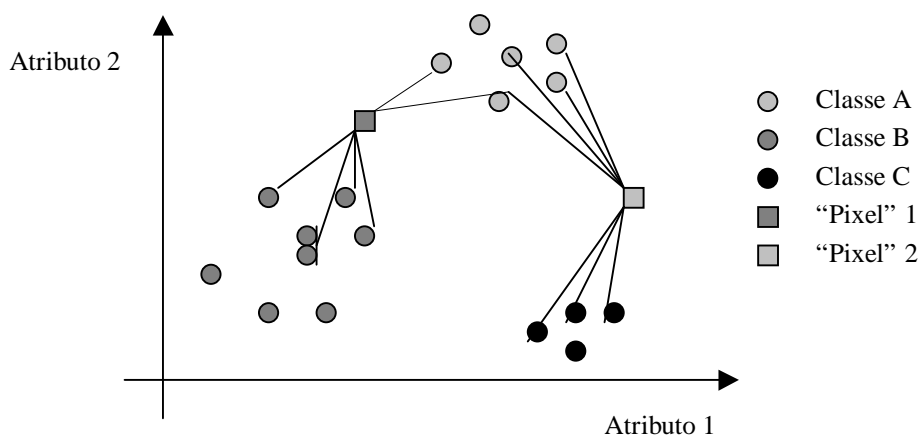


Figura 13: Exemplo de Classificação pelo Método do KNN.

Uma das desvantagens deste classificador é que para cada ponto que será classificado, será necessário encontrar seus K vizinhos mais próximos, ou seja, calcular a distância do “pixel” a ser classificado a todos os “pixels” vizinhos de classes conhecidas, determinar os “pixels” que estão às K menores distâncias e contar as quantidades dos mesmos para cada classe. Isso consome muito tempo e o processo de classificação é lento (Santos, 1997).

Existe uma variação deste método que consiste em selecionar “pixels” que estão dentro de uma hiperesfera de raio R centrada no “pixel” desconhecido e contar a quantidade de “pixels” para cada classe que está dentro desta hiperesfera. O raio R é escolhido empiricamente e os resultados podem variar muito dependendo destes raios (Santos, 1997).

3.9 Classificação Visual

Geralmente uma imagem é adquirida previamente classificada, porém sem uma finalidade específica, fornecida pelo INPE e sendo uma grande foto, colorida ou não, em papel, de determinada época, mono ou multiespectral, de acordo com o interesse do usuário. A interpretação visual é muitas vezes atrapalhada quando na imagem adquirida foi usado um processo de classificação com tendência a uma melhor qualidade na área específica de interesse.

Possuindo a imagem, deve-se reconhecer o que existe nela, extraindo o que interessa. Se faz necessário conhecimentos prévios, como por exemplo da época adequada para obtenção da imagem, tipo de produto, bandas a serem escolhidas, escalas, etc.

Os fotointerpretadores são de três tipos, os que possuem conhecimentos de campo, os que não o possuem, e os que possuem apenas fontes de informações adicionais. Dentre eles, os primeiros interpretam de forma mais fácil que os outros, já que os últimos podem encontrar dificuldades ao interagir as informações de que dispõem com a imagem, e os segundos usam para tal chaves de interpretação (conhecimentos estruturados como formulações para se resolver um problema). Procura-se fazer com que as chaves de interpretação não mudem de um fotointerpretador para outro, criando, para isso, legendas que ajudarão na separação das diversas classes (Moraes, 1992).

Interpretada a imagem, extrai-se as informações procuradas, através de vários métodos, que atendam às necessidades de cada estudo específico.

3.10 Classificação por Sistema Especialista

As metodologias de classificação assistida por sistema especialista surgiram para tentar automatizar a classificação visual. Desse modo, o processo decisório pode ser suficientemente rico e envolver uma quantidade razoavelmente pequena de informação, podendo, então, o conhecimento ser armazenado em termos de regras, ser aplicável e poder ser tratado de forma simplificada (James, 1985). A desvantagem que essa metodologia possui é que o processo decisório não irá responder corretamente fora dos padrões para os quais foi delineado, ao contrário dos métodos exclusivamente numéricos.

3.10.1 Conjuntos Nebulosos

3.10.1a Definições

Um conjunto nebuloso é definido matematicamente da seguinte forma (Zadeh, 1965):

Seja X um espaço de pontos, com um elemento genérico de X denotado por x . Um conjunto nebuloso A em X é caracterizado por uma função de pertinência $\mu_A(x)$ que associa a cada ponto em X um número real no intervalo $[0,1]$, onde $\mu_A(x)$ representa para x o seu grau de pertinência em A . Quando $\exists x \in X$, tal que $\mu_A(x) = 1$, então A é dito ser normalizado.

Um conjunto nebuloso pode, dessa forma, ser visto como a representação de um conjunto na Teoria dos Conjuntos Clássica, do qual só se tem um conhecimento imperfeito.

Na Teoria dos Conjuntos Clássica, cada conjunto A de um universo X pode ser definido através de uma função característica $\mu_A: X \rightarrow \{0,1\}$, tal que:

$$\mu_A(x) = \begin{cases} 1, & \text{se e somente se } x \in A \\ 0, & \text{caso contrário} \end{cases}$$

Pode-se ver que a pertinência de um elemento a um conjunto nebuloso não é mais dada por um valor em $\{0,1\}$, como na Teoria dos Conjuntos Clássica, mas por um valor

qualquer no intervalo $[0,1]$. Isto significa que a pertinência de um elemento a um conjunto nebuloso pode ser parcial.

O conjunto ordinário dos elementos x pertencentes a um conjunto nebuloso A , com valores de pertinência de pelo menos $\alpha \in [0,1]$ é definido pelo conjunto de nível α , A_α de A , tal que:

$$A_\alpha = \{x \in X / \mu_A(x) \geq \alpha\}$$

A função de pertinência do conjunto nebuloso A pode ser expressa pelos cortes de nível α , de acordo com a expressão:

$$\mu_A(x) = \sup \min\{\alpha, \mu_{A_\alpha}(x)\}, \alpha \in (0,1]$$

onde

$$\mu_{A_\alpha}(x) = \begin{cases} 1, & \text{se e somente se } x \in A_\alpha \\ 0, & \text{caso contrário} \end{cases}$$

Exemplo: Seja um universo X , um conjunto nebuloso A pode ser denotado por $A = \sum \mu_A(x)/x$. Dado o seguinte universo X e o conjunto nebuloso A :

$$X = \{1, 2, 3, 4, 5, 6, 7, 8\}$$

$$A = 0.1/2 + 0.2/3 + 0.5/4 + 0.6/5 + 0.8/6 + 1.0/7$$

onde, segundo a notação, o elemento 2 possui um grau de pertinência 0.1, o elemento 3 possui um grau 0.2, e assim por diante.

Os conjuntos de nível α de A são dados por:

$$A_{0.1} = \{2, 3, 4, 5, 6, 7\}$$

$$A_{0.2} = \{3, 4, 5, 6, 7\}$$

$$A_{0.5} = \{4, 5, 6, 7\}$$

$$A_{0.6} = \{5, 6, 7\}$$

$$A_{0.8} = \{6, 7\}$$

$$A_{1.0} = \{7\}$$

3.10.1b Operações com Conjuntos Nebulosos

As operações sobre os conjuntos nebulosos são similares àsquelas da Teoria de Conjuntos Clássica, porém são generalizadas para os Conjuntos Nebulosos.

- Conjunto Vazio

Um conjunto nebuloso é vazio se e somente se a sua função de pertinência é identicamente zero sobre X , $\forall x \in X$.

- Igualdade de conjuntos

Dois conjuntos nebulosos A e B são iguais se e somente se $\mu_A(x) = \mu_B(x)$, para todo x em X .

- Complemento

O complemento de um conjunto nebuloso A em X , denotado por $\neg A$ é definido por (Moraes, 1998):

$$\mu_{\neg A}(x) = n(\mu_A(x)), \forall x \text{ em } X$$

onde $n:[0,1] \rightarrow [0,1]$ é um operador de negação com as seguintes propriedades:

$$n(0) = 1 \text{ e } n(1) = 0$$

$$n(a) \leq n(b) \text{ se } a > b$$

$$n(n(a)) = a, \forall x \in [0,1]$$

Uma negação é estrita quando é contínua e também satisfaz a condição:

$$n(a) < n(b) \text{ se } a > b$$

A operação $n(a) = 1 - a$ satisfaz todas as quatro condições acima, sendo considerada a principal operação de negação.

- Interseção

A interseção de dois conjuntos nebulosos A e B é um conjunto nebuloso C, cuja função de pertinência é dada por:

$$\mu_C(x) = T\{\mu_A(x), \mu_B(x)\}, \forall x \in X$$

onde $\mu_A(x)$ e $\mu_B(x)$ são as funções de pertinência dos conjuntos A e B e o operador T: $[0,1]^2 \rightarrow [0,1]$, chamado de T-norma, é uma função associativa, comutativa, monotonicamente crescente e que satisfaz a condição de $\forall x \in X, T(x,1) = T(1,x) = x$ (Moraes, 1998).

- União

A união de dois conjuntos nebulosos A e B, com funções de pertinência $\mu_A(x)$ e $\mu_B(x)$ é um conjunto nebuloso C, que tem a seguinte função de pertinência:

$$\mu_C(x) = \perp\{\mu_A(x), \mu_B(x)\}, \forall x \in X$$

onde o operador $\perp: [0,1]^2 \rightarrow [0,1]$, denominado T-conorma, é uma função associativa, comutativa, monotonicamente crescente e que satisfaz a condição de $\forall x \in X, \perp(x,0) = \perp(0,x) = x$ (Moraes, 1998).

Neste trabalho, as T-normas e as T-conormas utilizadas são a T-norma $\min(a,b)$ e a T-conorma $\max(a,b)$. Quando se utiliza essas duas funções, vê-se que $A \cap \neg A \neq \emptyset$ e $A \cup \neg A \neq X$, o que seria um absurdo na Teoria dos Conjuntos Clássica.

Exemplos: Seja o conjunto universo X e dois conjuntos A e B em X, abaixo:

$$X = \{1,2,3,4,5,6,7\};$$

$$A = 0.8/3+1/5+0.6/6;$$

$$B = 0.7/3+1/4+0.5/6$$

A união e interseção dos conjuntos A e B podem ser dados por:

$$A \cup B = \max(\mu_A, \mu_B)$$

$$A \cap B = \min(\mu_A, \mu_B)$$

e a negação por $\neg A = 1 - \mu_A$. Desse modo, temos:

$$A \cup B = 0.8/3+1/4+1/5+0.6/6;$$

$$A \cap B = 0.7/3+0.5/6;$$

$$\neg A = 1/1+1/2+0.2/3+1/4+0.4/6+1/7$$

- Implicação

A implicação entre dois conjuntos nebulosos A e B, com funções de pertinência $\mu_A(x)$ e $\mu_B(y)$ resulta em um conjunto nebuloso $C = A \Rightarrow B$, cuja função de pertinência é dada por:

$$\mu_C(x,y) = \nabla \{ \mu_A(x), \mu_B(y) \}, \forall x \in X, \forall y \in Y$$

onde $\nabla: [0,1]^2 \rightarrow [0,1]$ é um operador de implicação que tem as seguintes propriedades, para $\forall a, a', b, b', c \in [0,1]$ (Moraes, 1998):

- Se $b \leq b'$ então $I(a,b) \leq I(a,b')$
- $I(0,b) = 1$
- $I(1,b) = b$

As implicações “puras” também obedecem às propriedades:

- Se $a \leq a'$ então $I(a,b) \geq I(a',b)$
- $I(a,I(b,c)) = I(b, I(a,c))$

Existem vários operadores de implicação definidos na literatura, sendo as T-normas os modeladores mais utilizados na prática (Moraes, 1998) para modelar a implicação, apesar de não pertencerem à classe das implicações “puras”.

3.10.2 Sistemas Especialistas

Sistemas especialistas são sistemas de classificação que utilizam conhecimento de especialistas para auxiliar na tarefa de classificação, que consiste em decidir entre um conjunto de classes conhecidas qual é aquela que um determinado conjunto de informações pertence.

Estes sistemas trabalham sobre o conhecimento de um dado domínio, de modo a poder tratar de forma semelhante a um ser humano um conjunto de informações. Tal conhecimento deve ser representado ou codificado, para que possa ser processado por computador. Desse modo, o conhecimento é, em geral, caracterizado por um conjunto de regras e fatos, onde os primeiros representam a tradução do conhecimento do especialista e os últimos são dados adquiridos anteriormente ou durante o processamento dos fatos.

Em sistemas especialistas é importante que novos conhecimentos introduzidos não entrem em conflito com o conhecimento já armazenado, para que seja mantida a consistência da base de conhecimento.

Uma linguagem lógica de primeira ordem – $L(P,F,C,V)$ é determinada pela especificação dos seguintes conjuntos (Bittencourt, 1996):

- um conjunto P de símbolos de predicado
- um conjunto F de símbolos de função
- um conjunto C de símbolos de constante
- um conjunto V de símbolos de variável

Estes conjuntos formam o alfabeto da linguagem lógica, que em associação aos operadores lógicos (\neg é “não”, \wedge é “e”, \vee é “ou”, \Rightarrow é “implica”, \exists é “existe” e \forall é “qualquer”) possibilitam a formulação de conhecimentos, constituindo a base dos sistemas especialistas.

As regras são representações de fórmulas lógicas compostas por conjunções e disjunções, podendo levar a mais de uma conclusão.

Uma regra simples em processamento de imagens seria, por exemplo:

SE nível de cinza do “pixel” está no intervalo [0,13] na banda 4
ENTÃO classe é Rio

que é a representação da implicação:

SE <condição>
ENTÃO <conclusão>

As condições e conclusões de uma regra são usualmente representadas por uma lista da forma:

(<objeto>, <atributo>, <valor>)

Neste trabalho, os objetos são os tipos sobre os quais o conhecimento atua; os atributos são características presentes na imagem que podem ser referenciadas para a utilização do conhecimento e os valores são os valores numéricos ou sintáticos que os atributos podem assumir.

Desse modo, no exemplo anterior o objeto é um segmento de uma imagem (“pixel”) e o atributo é o intervalo dos níveis de cinza na banda 4, com valores entre [0,13]. Outro exemplo é dado pela lista (“pixel”-i, classe, “floresta”), em que classe é o atributo do “pixel”-i que tem como valor “floresta”.

Do ponto de vista lógico, a tripla (<objeto>, <atributo>, <valor>) pode ser codificada através de um predicado de igualdade, com (atributo(objeto) = valor), onde atributo é uma função, objeto é uma constante ou variável e valor é uma constante.

Com o objetivo de se deduzir novos fatos, utiliza-se dois mecanismos de inferência que combinam fatos e regras.

Na inferência “para-frente”, a partir da verificação das premissas se concluem fatos que podem validar premissas de outras regras. Na inferência “para-trás”, parte-se das conclusões possíveis e procura-se quais seriam as premissas que poderiam levar a elas e no final dessa sequência para-trás, de posse de todo o “caminho” percorrido, ele é validado.

Em casos quando os atributos não podem ser medidos com exatidão, ou quando o conhecimento não é válido em qualquer situação, podem ser introduzidos coeficientes de certeza para codificar essa incerteza na conclusão da regra. Essa informação incerta pode ser tratada através de probabilidades (Rich e Knight, 1993; James, 1985), crença e plausibilidade (Senne, 1988), possibilidades (Kruse et al., 1991; Dubois e Prade, 1996) ou “ad-hoc” (Desachy, 1991).

Sistemas especialistas que envolvem o uso de conjuntos nebulosos são tratados pela “lógica nebulosa”, na qual lança-se mão de uma forma de raciocínio denominado raciocínio aproximado (Moraes, 1998).

Sejam A_i e A'_i , $i = 1, \dots, n$, subconjuntos nebulosos em X_i , B um subconjunto nebuloso em Y com as funções de pertinência conhecidas, R uma regra e F um fato, onde:

R : SE $\langle x_1 = A_1 \rangle$ E $\langle x_2 = A_2 \rangle$ OU $\langle x_3 = A_3 \rangle$ ENTÃO $\langle y = B \rangle$.

F : $\langle x_1 = A'_1 \rangle$ E $\langle x_2 = A'_2 \rangle$ E $\langle x_3 = A'_3 \rangle$.

O processo geral de inferência é baseado nos passos seguintes (Sandri, 1997):

- Cria-se uma relação nebulosa A em $X_1 \times \dots \times X_n$ entre os A_i e uma relação nebulosa A'_i em $X_1 \times \dots \times X_n$ entre os A'_i .
- Cria-se uma relação nebulosa R entre as premissas dada pela relação nebulosa A e pela conclusão dada pelo conjunto nebuloso B , usando uma função de implicação.
- Faz-se uma composição de R com A' e projeta-se o resultado sobre Y , obtendo-se o subconjunto nebuloso B' .

Quando os A'_i representam fatos precisos, cada A'_i é tal que $\exists x_i^* \in X_i, \mu_{A'_i}(x_i^*) = 1$ e o processo é simplificado (Moraes, 1998).

Seja a base de conhecimento composta por n regras R_j do tipo:

$$R_j: SE \langle x_{ij} = A_{ij} \rangle E \dots E \langle x_{nj} = A_{nj} \rangle \\ ENTÃO \langle y = B \rangle$$

O processo é dividido em quatro fases:

- Seja o valor de realização de cada x_i dado por $x_i^* \in X_i$, isto é x_i^* representa o fato A'_i . Determina-se o grau de compatibilidade de x_i^* , $1 \leq i \leq n$, para cada A_{ij} da regra R_j , $1 \leq j \leq m$:

$$h_{ij} = \mu_{A_{ij}}(x_i^*);$$

- Determina-se o grau de compatibilidade global da premissa de cada regra R_j , $1 \leq j \leq m$, com uma T-norma T :

$$h_j = T(h_{1j}, \dots, h_{nj});$$

Quando a regra envolve conectivos OU, este passo envolverá também o uso de uma T-conorma.

- Determina-se o conjunto nebuloso B'_j , referente a cada regra R_j usando um operador de implicação ∇ :

$$\mu_{B'_j}(y) = \nabla (h_j, \mu_{B_j}(y)); \forall y \in Y;$$

- Determina-se o conjunto nebuloso B' , usando um operador \diamond para agregar os B'_j :

$$\mu_{B'}(y) = \diamond (\mu_{B'_1}(y), \dots, \mu_{B'_m}(y)); \forall y \in Y.$$

O operador que implementa \diamond é geralmente escolhido de acordo com a função da implicação ∇ usada (Moraes, 1998).

3.10.3 A função distância

A função distância é uma função que associa a cada “pixel” de certo nível de cinza de uma imagem de entrada, sua distância ao mais próximo “pixel” de valor zero.

3.10.4 A limiarização

A limiarização de uma imagem $f: E \rightarrow \{0, \dots, 255\}$ relativo ao intervalo $[a, b]$ é o operador dado por (Moraes, 1998):

$$\text{lim iarização}_{[a,b]}(f)(x) = \begin{cases} 255, & \text{se } f(x) \in [a, b] \\ 0, & \text{caso contrário} \end{cases}$$

A limiarização serve para delimitar distâncias específicas fornecidas pela função distância ou para isolar intervalos fechados de valores em uma imagem em níveis de cinza.

3.10.5 Sistemas Especialistas para Classificação de Imagens

Existem várias implementações práticas de classificadores na literatura, dentre os quais podemos citar:

- O Sistema ICARE (Desachy, 1991) que procura gerar um mapa para uma região, através de um sistema especialista nebuloso, que age sobre uma pré-classificação estatística, mapas e classificações anteriores da região de interesse.
- O trabalho de Haverkamp et al, (1994), que combina métodos heurísticos e numéricos para a classificação de blocos de gelo no mar, através de imagens radarmétricas.

- O trabalho de Mascarella (1995) que é uma versão do Sistema ICARE onde o conhecimento é armazenado em uma rede neural nebulosa, podendo ser tratado em linguagem natural.
- O trabalho de Wendling et al (1995), que interpreta imagens de satélite por conhecimento acerca das classes procuradas, sendo similar ao ICARE.
- O trabalho de Mathieu-Marni et al (1996), que utiliza um sistema especialista que gerencia uma biblioteca de programas para computação de misturas de classes e classificação de imagens Spot e Landsat.
- O trabalho de Roux e Desachy (1997) que se utiliza de um sistema especialista dado por uma rede neural nebulosa e um classificador nebuloso, sendo a fusão das informações dessas duas fontes, usada para a classificação.
- O trabalho de Queiroz et al (1997), que utiliza um pequeno sistema especialista nebuloso para classificação de axônios.

Será dada adiante uma breve descrição do Sistema ICARE, devido ao sistema especialista utilizado neste trabalho ter se inspirado no mesmo.

3.10.5a O Sistema de Interpretação de Imagens ICARE

O Sistema ICARE (“Image Cartographie Expert”) é um classificador automático de imagens, proposto por Desachy, que simula a interpretação visual de imagens e tem o objetivo de gerar novas cartas topográficas de uma certa região. O sistema possui quatorze classes.

O algoritmo de classificação é dividido em três fases. Na primeira, denominada de pré-classificação, o sistema obtém os valores numéricos para todas as classes possíveis de um “pixel”, levando em consideração a radiometria e a textura, e retém apenas os três maiores valores resultantes da função de classificação. Na segunda fase, denominada fase 1 da classificação, utiliza-se dados de Modelos Numéricos de Terreno, com auxílio de um sistema especialista. Na terceira fase, denominada fase 2 da classificação, o sistema especialista utiliza-se de informações sobre Cartas Topográficas anteriores, com o objetivo de aprimorar a classificação. Nestas duas últimas fases, o sistema também mantém, para

cada “pixel”, três valores resultantes da função de classificação e somente ao final de todo o processo decide pela classe que produz o maior valor desta função.

Nas fases 1 e 2 de classificação, são necessários dados referentes a área em estudo. Esta fonte de dados pode, inclusive, ser uma classificação gerada anteriormente. Na figura 14 encontra-se um diagrama esquemático do Sistema ICARE.

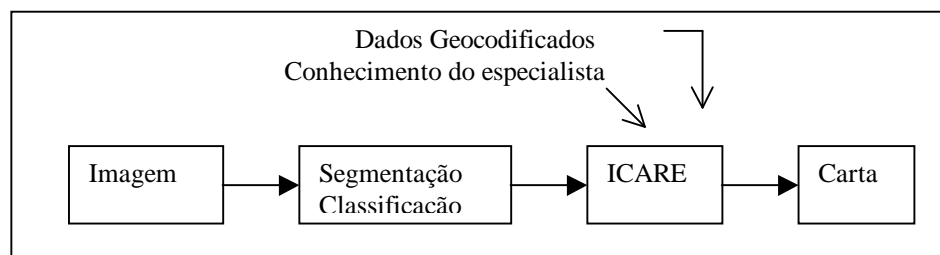


Figura 14: Diagrama geral de um sistema de interpretação de imagens.

O Sistema ICARE é capaz de, por exemplo, determinar se um “pixel”, estando próximo a um estrada, favorece alguma classificação em especial. Tal determinação é feita através da teoria dos conjuntos nebulosos aplicada a uma função distância.

No Sistema ICARE são utilizadas duas bases de dados: a Base de Conhecimento e a Base de Dados. Na primeira existem informações sobre a vegetação característica das regiões por tipo ou utilização do solo, clima e topografia. Na Segunda encontram-se modelos numéricos de cartas climáticas, de solos e topográficas.

As duas últimas etapas do processo de classificação utilizado fazem uso das bases de Conhecimento e de Dados, com o objetivo de aperfeiçoar uma classificação numérica inicial. O sistema especialista manipula o conhecimento armazenado na Base de Conhecimento que em certas ocasiões depende de dados que estão armazenados na Base de Dados.

3.10.5b Classificação pelo Sistema Especialista de Moraes (1998)

O sistema especialista de Moraes (1998) compreende uma série de regras, expostas adiante, que foram implementadas no sistema Khoros, e que foi utilizado na classificação da imagem objeto de nosso estudo.

3.10.5.1 Regras implementadas

As regras do sistema especialista utilizado implementadas no Khoros são as seguintes:

- SE níveis de cinza são escuros (0; 23) na banda 4 E são muito escuros (0;17) na banda 5, ENTÃO classe é Rio;
- SE porção de terra na direção leste, ENTÃO são terras à direita do Rio;
- SE porção de terra na direção oeste, ENTÃO são terras à esquerda do Rio;
- SE são terras à direita do Rio E estão próximas do Rio, ENTÃO é a Margem do Rio;
- SE são terras à esquerda do Rio E estão próximas do Rio, ENTÃO é a Margem do Rio;
- SE níveis de cinza são meio escuros (27;70) na banda 4 E níveis de cinza meio escuros ou muito claros (59;255) na banda 5 E níveis de cinza meio escuros (12 e 60) na banda 7 E estão nas terras à direita ou a esquerda do Rio, ENTÃO são Áreas Antropizadas;
- Floresta Aluvial à direita do Rio é uma área alongada que aparece em verde na composição RGB e que está a nordeste da Cidade de Aveiro. São as áreas na imagem RGB nas terras à esquerda do Rio.
- SE são áreas dentro do Rio e uma área à direita do Rio e outra à esquerda do Rio, E está à direita do Rio muito próximo da Área Antropizada e muito perto da Área Antropizada à esquerda do Rio, E são níveis de cinza muito escuros na banda 5 à esquerda do Rio, ENTÃO são Áreas de Contato.
- SE são níveis de cinza claros na banda 3 e na banda 7 que estão na margem direita do Rio e estão dentro da Área Antropizada, ENTÃO é a Cidade de Aveiro.
- A drenagem densa são os níveis de cinza escuros nas bandas 4 e 5, longe do Rio e à direita do mesmo. SE são níveis de cinza escuros nas bandas 4 e 5, longe do Rio, E estão muito próximos das áreas de drenagem densa, ENTÃO é a Floresta Ombrófila Densa Submontana.
- SE não for nenhuma das classes (Rio, Floresta Ombrófila Densa Submontana, Floresta Secundária, Floresta Aluvial, Área Antropizada, Área de Contato e Cidade de Aveiro), ENTÃO são Áreas de Mistura das classes Floresta Secundária e Área Antropizada.

Os operadores que delimitam níveis de cinza escuros e claros são obtidos por limiarização e as regras que implementam funções do tipo “perto”, “longe”, são obtidas pela aplicação da função distância.

3.11 O “Software” Khoros

O Khoros é um ambiente de aplicação e desenvolvimento de “softwares”, destinado ao processamento, análise, exploração e visualização de dados multidimensionais. Este sistema possui uma variedade de programas para o processamento de imagens, com diversas aplicações, permitindo ser utilizado como um conjunto de bibliotecas programadas visualmente em um ambiente específico, denominado Cantata. Tais bibliotecas são denominadas “toolboxes”, sendo algumas disponíveis na Internet.

Neste trabalho utilizamos o Cantata como nosso ambiente de trabalho, onde montamos “workspaces”, para a classificação de uma imagem, pelos classificadores supervisionados, que estão contidos na “toolbox Classify”. Na figura 15, encontra-se uma “workspace” criada para a classificação de uma imagem pelo método do paralelepípedo.

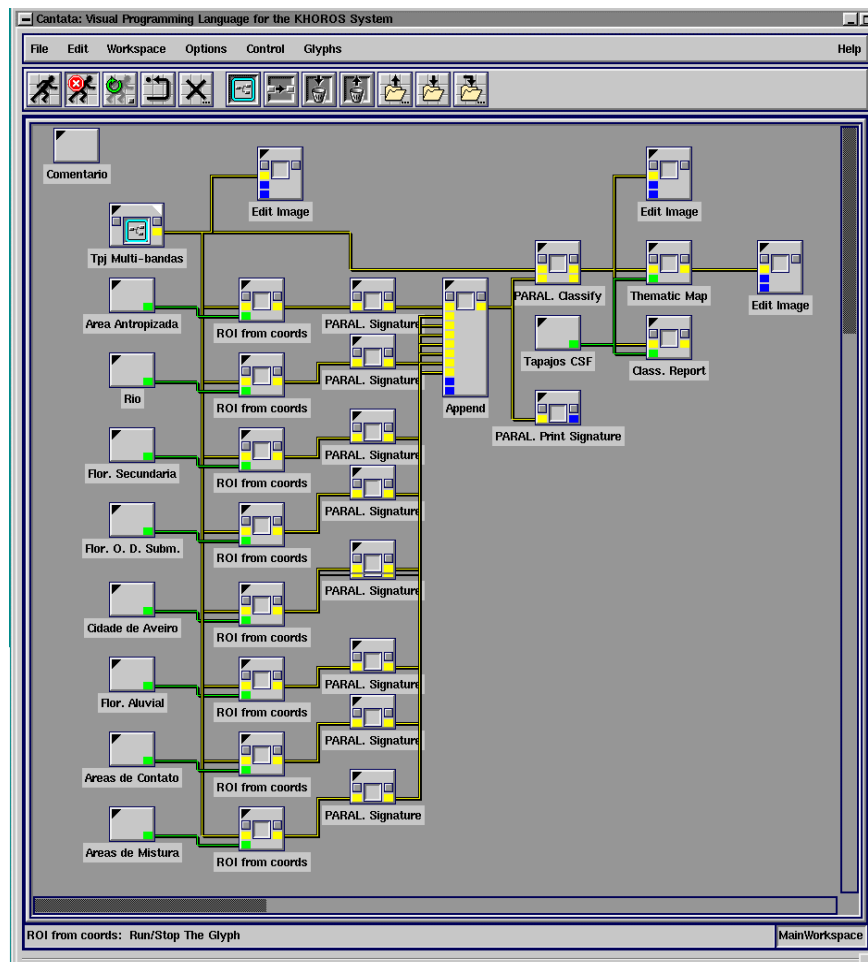


Figura 15: “Workspace” montada no subsistema Cantata do Khoros.