

C.N. Francisco
Mestranda de Geoprocessamento da EPUSP
Xavier-da-Silva, J.
Prof. Titular do Dep. de Geografia
Universidade Federal do Rio de Janeiro
Laboratório de Geoprocessamento - Dep. de Geografia
Inst. de Geociências - CCMN
Ilha do Fundão - 21.941-590 - RJ

RESUMO

A entrada de dados em Sistemas de Informações Geográficas - SIG's demanda aproximadamente metade do tempo necessário para o desenvolvimento de um projeto que faça uso desta ferramenta. Os SIG's, em sua maioria, estão preparados para receber os dados via mesa digitalizadora, o que consome esforços de tempo e recursos. O scanner pode ser uma alternativa viável para entrada de mapas nestes sistemas, porém deve-se atentar para algumas limitações deste equipamento e do produto gerado em particular quanto às exigências específicas do SIG que se pretenda utilizar. O objetivo deste trabalho é analisar a viabilidade de scanners para entrada de dados em SIG's face a essas limitações.

ABSTRACT

The GIS' data input spends approximately 50% of the time needed to development of projects. In the most, GIS are prepared to get data by digitizing tablet, what demands much time and resources. The scanners may be an option to input maps in the system, but it's necessary to pay attention to some restrictions of this tool and the product obtained by its used. The goal of this paper is analysing the possibilities of using scanner to input data in GIS before its restrictions.

1. INTRODUÇÃO

Entrada de dados em Sistemas de Informações Geográficas constitui um processo de transformação dos dados capturados fisicamente em estruturas numéricas passíveis de armazenamento eficiente. Os dados a serem capturados podem ser espaciais e não-espaciais; os

primeiros referem-se aos mapas propriamente ditos e sua entrada baseia-se na digitalização da geometria das feições e conteúdo lógico ambientais de interesse. Em termos geométricos, estas feições podem ser entendidas como polígonos, linhas e pontos que compõem o mapa. Os dados não-espaciais constituem-se dos atributos sem territorialidade relacionados às feições espaciais

identificadas. São exemplos destes atributos não-espaciais o nome do proprietário de um terreno, informações sobre multas e taxas em vigor. Os dados não-espaciais podem ficar contidos numa estrutura de bancos de dados convencional acoplada à base geocodificada.

A entrada de dados constitui-se numa das questões mais importantes e complexas para montagem do SIG. Segundo Aronoff (1989), o custo inicial de elaboração de uma base de dados é da ordem de 5 a 10 vezes mais do que custo dos softwares e hardwares que compõem um SIG. Ao mesmo tempo, é a etapa que mais consome tempo num projeto que faça uso desta ferramenta.

Tradicionalmente, a entrada de dados vem sendo feita, na maior parte dos casos, por mesa digitalizadora. Num trabalho realizado por Teixeira (1990), em que foram analisados 64 SIG's através de dados contidos no "GIS SOURCE BOOK" de 1989, foi constatado que em 54 deles a mesa digitalizadora era utilizada para entrada de dados. Apesar do seu uso intensivo, esta ferramenta não vem satisfazendo de maneira plena os seus usuários, já que a entrada de dados via mesa consome muito tempo. Com este quadro, os scanners vem se tornando uma alternativa viável, principalmente, pela queda de preços dos scanners de mesa (Cartensen, 1991).

A alternativa pelo scanner se deve ao fato da captura da geometria ser bem rápida; os pré e pos-processamentos, porém, que necessitam ser executados podem consumir tempo excessivo (Cartensen, 1991 e Peuquet, 1984). O

objetivo deste trabalho é analisar a viabilidade de scanners para entrada de dados espacializados em SIG's, a partir da experiência que vem sendo adquirida junto ao Laboratório de Geoprocessamento do Departamento de Geografia da UFRJ. São enfatizadas as características deste equipamento, os processamentos necessários para rasterização dos mapas e as soluções encontradas face às características dos mapas rasterizados e as limitações dos softwares de digitalização e edição.

2. OS DOCUMENTOS CARTOGRÁFICOS E SUAS RELAÇÕES COM A ENTRADA POR SCANNER

A captura da geometria por scanners consiste na codificação das feições do documento para uma matriz de valores (x,y), onde cada pixel contém um valor médio de reflectância de uma pequena área do mapa original, gerando arquivos em formato raster, que se caracterizam por serem de grande tamanho. Isto pode limitar o uso desta ferramenta, caso não haja compactação eficiente, já que se necessita de equipamentos mais potentes para o armazenamento e processamento dos dados. Neste ponto é extremamente importante atentar para algumas características do processo de digitalização por scanner com vistas a gerar arquivos que sejam menores possíveis e que não contenham dados, muitas vezes, desnecessários para o usuário.

O primeiro ponto a ser considerado refere-se à resolução espacial, isto é, o tamanho do pixel da matriz gerada pela digitalização. A resolução é dada pela

DPI (pontos por polegada); quanto maior esta for, mais refinadas vão ser as feições. A DPI é geralmente definível pelo usuário e deve ser baseada na espessura das linhas no original e nos detalhes geométricos que devem ser preservados no produto final. Assim, para rasterização fiel de fotos aéreas, por exemplo, há necessidade de uma DPI mais elevada, já que o tamanho dos grãos de sais usados em fotografias varia de 0,1 a 1 μm , ou seja, para reproduzir esta resolução há necessidade de DPI acima de 1000. Já a rasterização de mapas demanda uma DPI menor, pois a menor espessura de linha encontrada num mapa está em torno de 0,1 mm, exigindo, assim, resolução maior que 250 DPI para reproduzir estas feições.

É importante para o usuário observar a relação entre DPI, tamanho do arquivo e produto final. Uma DPI elevada produz feições mais refinadas, porém não mais refinadas do que do mapa original, assim quando ultrapassa-se o refinamento do documento original, há geração de arquivos contendo dados que são supérfluos. Por sua vez, o acesso e processamento dos dados ficam demorados e, muitas vezes, inviáveis devido às restrições do equipamento. Deve-se também mencionar que uma maior DPI gera um maior número de pixels compondo a largura de cada linha, o que pode tornar a edição do mapa mais difícil e demorada.

Uma outra questão é a relação DPI e precisão dos mapas rasterizados. Cartensen (1991) testou a digitalização por scanner com três níveis de resolução - 75, 150 e 300 DPI - os resultados

deste teste mostram que não houve erro maior que 0,1 mm e que os níveis de erros não foram afetados pela resolução espacial nem estavam em função da distância entre os pontos no mapa original.

Assim, a escolha da DPI pelo usuário deve ser baseada no tamanho do arquivo a ser gerado e nos detalhes que se deseja obter no produto final. Os mapas no Brasil são classificados em três níveis quanto a exatidão (decreto lei nº 89.817/84), o nível mais elevado (nível A) exige uma exatidão de 0,3 mm na escala da carta, que representa uma resolução em torno de 75 DPI. Em consequência, pode-se afirmar que uma resolução mais grosseira que esta afeta a exatidão da carta digitalizada no nível A de exatidão.

O segundo ponto importante que se deve atentar quanto a digitalização por scanner é a resolução radiométrica, ou seja, o intervalo de valores que cada pixel pode discriminar. A resolução pode abranger, normalmente, 1 bit (preto e branco) a 24 bits ("true color"). Quanto maior a resolução radiométrica, maior o tamanho do arquivo gerado.

A escolha da resolução radiométrica deve se basear na qualidade do mapa original. Se o mapa contém apenas linhas (preto) e fundo (branco), 1 bit atende as necessidades. Porém, se o mapa contém diversos tons ou cores e se deseja manter esta diversidade no mapa rasterizado, então deve-se optar para digitalizá-lo em resoluções superiores a 1 bit. Deve-se, porém, observar que um mapa que foi rasterizado em mais de 1 bit também não identifica as feições e

seus atributos, ou seja, há necessidade de editá-los.

Um outro ponto é quanto ao formato do arquivo gerado pelo software de rasterização, este deve ser compatível com o SIG e o programa de edição a serem utilizados. O formato TIFF - Tagged Image File Format - é um dos gerados comumente por estes softwares e tem uma aceitação de 25% dos SIGs, segundo "GIS SOURCE BOOK" de 1993 (Brandalize, 1993).

O tamanho dos arquivos também pode ser diminuído a partir da compressão, chegando a uma taxa de mais de 90% de acordo com arquivo.

Por fim, deve-se mencionar que os scanners não são dotados de nenhum nível de inteligência capaz de identificar as feições dos mapas, desta forma o scanner digitaliza todas as informações contidas no mapa, diferentemente da mesa digitalizadora, onde apenas as feições selecionadas pelo operador são digitalizadas. Em função desta limitação, há necessidade de se executar pré-processamentos, ou seja, preparar o mapa original a ser rasterizado. Os pós-processamentos, ou seja, edição dos mapas após a digitalização, também é uma etapa necessária. O tempo e os recursos dispendidos nestas fases são menores caso o documento original apresente-se em bom estado.

3. ETAPAS NA ENTRADA DE DADOS POR SCANNERS

Neste ítem do trabalho, pretende-se discorrer sobre dois métodos de entrada de dados por scanner que visam otimizar este processo conforme esquema abaixo:

Método 1

Pré-processamento ### Rasterização
Edição

Método 2

Rasterização ### Tratamento de
imagem ### Edição

A escolha por um destes métodos dependerá do estado do mapa original e também dos recursos de softwares disponíveis para a edição dos mapas rasterizados.

O método 1 é indicado para os originais disponíveis que não se apresentem em bom estado; há necessidade de se executar pré-processamento com vistas a eliminar as imperfeições dos originais. Também é indicado para mapas que contenham muitos símbolos e cores que dificultam a classificação das feições no processamento digital e na edição dos mapas rasterizados ou, ainda, quando se deseja extrair apenas algumas informações entre as específicas existentes no mapa. Nestas duas últimas situações, o objetivo do pré-processamento é a separação das informações para mapas diferentes, por exemplo, extrair a hidrografia da carta topográfica.

O método 2 restringe-se aos mapas originais em bom estado de conservação e, também, àqueles onde as cores e os símbolos não apresentam obstáculos a sua edição.

3.1. Pré-processamento

Os objetivos desta fase, prevista apenas para o método 1 são a separação das informações em mapas separados e a eliminação das imperfeições do original.

Ela consiste em passar as feições a limpo, contendo o mapa a passar pelo scanner apenas as informações desejadas. Deve-se atentar para a elaboração de um mapa o mais limpo possível, com as feições bem destacadas e as espessuras das linhas próximas ao original. Também pode-se redesenhar os mapas sem separar as informações que se desejam em vários cartogramas, sendo que cada classe de informação é colocada em uma cor. Este processo é importante para garantir a superposição perfeita de informações que estão espacialmente associadas. Por exemplo, curvas de nível e drenagem, as primeiras recebem uma cor e os rios outra, sendo estes separados por métodos classificatórios contidos em programas de processamento digital de imagens.

A qualidade do documento original reduz o esforço da edição a posterior. Segundo Peuquet (1984), geralmente é menos custoso e mais rápido preparar o original a ser rasterizado, mesmo que signifique redesenhá-lo. É também sugerido que as instituições responsáveis pela produção de mapas mantenham o original sem símbolos, de modo que possam ser utilizados na digitalização por scanner.

3.2. Rasterização

Nesta etapa, prevista em posições diferentes nos dois métodos, deve-se atentar para a DPI e a resolução radiométrica a serem definidas pelo usuário. A opção por 1 bit pode ser utilizada para mapas que contenham apenas uma classe de informação ou, ainda, para aqueles que passaram por uma etapa de pré-processamento e compõem-se de feições apenas com uma cor e o fundo

branco. A vantagem desta resolução é a geração de arquivos de menor tamanho e a de linhas com o mesmo valor (0 ou 1), o que facilita a edição posterior.

Para mapas coloridos, que posteriormente passarão pela fase de processamento digital de imagens, com vistas a identificação das classes do mapa (método 2), a resolução adequada é a de 24 bits, já que permite extrair as classes com maior facilidade, pois utiliza o valor do pixel no espaço conhecido como RGB (vermelho, verde e azul), onde cada canal representa uma cor. Deve se ressaltar que resolução de 24 bits produz arquivos muito grandes, na ordem de 24 vezes maior do que um arquivo de 1 bit sem compressão.

3.3. Tratamento de imagem

Quando o mapa for digitalizado em mais de 1 bit, pode-se utilizar métodos de classificação de imagens para identificar as feições e alocar os atributos. Porém a classificação destas imagens não é um processo simples, já que diferentes classes, com suas respectivas cores, podem possuir respostas espectrais semelhantes, principalmente, quando o fundo do mapa for composto por alguma cor e/ou padronagem, como é o caso das curvas de nível em áreas urbanas nas cartas topográficas do IBGE. A classificação é facilitada quanto maior for a resolução radiométrica, já que o número de combinações de cores e padrões de impressão é grande neste caso.

A principal vantagem desta etapa é o reconhecimento das classes do mapa de modo mais inteligente para o sistema,

economizando esforços na edição dos mapas.

3.4. Edição

A edição consiste no estabelecimento das ligações entre a geometria das feições capturadas e seus atributos lógicos, espaciais e não-espaciais. Existe uma carência de operações de edição de mapas rasterizados em alguns SIG's, sendo necessário o uso de pacotes gráficos externos a estes sistemas. Em muitos softwares disponíveis atualmente, é necessário o uso de pacotes gráficos externos aos SIG's para execução das tarefas de edição. Estes pacotes, muitas vezes, não atendem às necessidades dos produtos cartográficos, principalmente, aquelas relacionadas a precisão do produto final. Essa carência existe pelo fato destes SIG's terem desenvolvido suas entradas de dados via mesa digitalizadora e, ao mesmo tempo, permitirem a importação de arquivos com formato raster, porém sem incorporarem opções eficientes para edição dos arquivos importados. Abaixo estão relacionados procedimentos importantes na etapa de edição de dados rasterizados.

- Eliminação de ruídos - consiste na eliminação das imperfeições do raster relacionadas a pequenas sujeiras existentes no documento original. Estes ruídos podem ser eliminados por filtros ou apagando diretamente no monitor.

- Junção de mapas - esta operação é fundamental, já que os scanners de mesa com preços mais acessíveis são aqueles que comportam mapas de menor tamanho. Assim, para mapas de maiores há a necessidade de rasterizar os mapas em

partes e, depois, juntá-los em algum programa. Esta junção pode ser feita no pacote gráfico ou no próprio SIG.

- Rotação de imagens - esta função é necessária quando o mapa foi digitalizado inclinado em relação ao sistema de coordenadas, havendo necessidade, assim, de rotacioná-lo para colocá-lo na posição adequada. Estes softwares devem permitir uma rotação da ordem de minutos

- Restituição de feições - é usual que ocorram falhas nas linhas nos mapas rasterizados, assim uma outra função é completar estas linhas. O objetivo aqui é corrigir a deometria das feições as feições, sendo permissíveis os exageros típicos das representações cartográficas.

- Alocação de atributos - um mapa em estrutura raster, cada pixel contém um valor que refere-se ao atributo da entidade cartografada. Quando a imagem é digitalizada em 1 bit, o valor contido no pixel é 0 ou 1; se o mapa contém mais de um atributo, é necessário que se converta o mapa para uma estrutura matricial de maior número de bits por pixel, de forma que este possa receber um intervalo maior de valores e, assim, cada valor relacionar-se a um atributo. A alocação de atributos, quando se usa pacotes gráficos externos ao SIG, consiste, para o usuário, na pintura de cada polígono do mapa com uma cor, que possui um respectivo valor, este é associado a uma classe; assim, por exemplo, num mapa de uso do solo, os polígonos que têm como atributo o uso urbano recebem a mesma cor e, conseqüentemente, possuem o mesmo código

identificador. Já com estes mapas no interior do SIG, os valores são identificados e a legenda é elaborada.

Nos mapas que foram digitalizados em mais de 1 bit e onde foi bem sucedida a classificação com vistas a identificação das feições existentes, ou seja, foi possível separar as várias classes do mapa pela cor através de métodos de classificação de imagens, não há necessidade de se executar a fase de pintura interativa das feições, já que os atributos já estão identificados; talvez haja necessidade de alguma correção gráfica, como, por exemplo, pintar alguns pixels que não foram bem identificados pelo scanner.

- Georeferenciamento - esta etapa consiste na atribuição de coordenadas geográficas de um sistema de projeção cartográfica à base de dados coligidos. Este procedimento é executado no SIG, através da definição de pontos reconhecíveis.

Conforme mencionado, algumas das etapas descritas acima podem ser executadas pelo próprio SIG, principalmente, se estes rodarem em plataformas e softwares mais sofisticados. Caso os SIGs não possuam estas funções, pode-se fazer uso de pacotes gráficos comerciais, porém estes, em geral, não satisfazem plenamente os requisitos para processar produtos cartográficos, onde a precisão é fundamental.

Outra dificuldade presente relaciona-se às feições lineares como rios, estradas etc. Nestas há a necessidade de se afinar as linhas, função ausente, com

frequência, nos pacotes gráficos disponíveis. Quando se digitaliza uma imagem por scanner, a espessura da linha contém vários pixels; este número está relacionado com a DPI definida. Nas feições poligonais, isto não é um grande problema já que definem áreas, podendo haver, entretanto, superestimativas no cálculo destas áreas. Para as lineares, por outro lado, deve-se afinar as linhas, para que cada uma contenha apenas 1 pixel. Se no pacote gráfico utilizado não estiverem disponíveis procedimentos eficientes de afinamento de linhas, tal tarefa pode tornar-se muito demorada e mesmo penosa.

Quando se digitaliza um mapa em scanner com objetivo de convertê-lo ao formato vetorial, o software de vetorização deve ser verificado quanto a existência de procedimentos destinados a correções e melhoria automática das feições rasterizadas. Caso estes procedimentos existam, a edição pode ser efetuada nestes softwares.

É comum, o uso da mesa digitalizadora para a entrada de dados com posterior conversão do arquivo para formato raster. Isto ocorre, principalmente, no caso de estudos ambientais, onde o formato raster apresenta vantagens sobre o vetorial em algumas funções, por exemplo, as associadas a superposição de mapas. No entanto, é frequente verificar SIG's que analisam basicamente dados em formato raster associarem a entrada de dados ao formato vetorial. Na prática, isto significa entrar dados por mesa digitalizadora para serem utilizados sob a forma raster. Não seria mais

interessante que estes sistemas oferecessem uma entrada de dados eficiente diretamente por scanners?

4. EXPERIÊNCIA SAGA-UFRJ

O SAGA - Sistema de Análise Geo-Ambiental - desenvolvido pelo Laboratório de Geoprocessamento do Departamento de Geografia da UFRJ, é um sistema cuja entrada de dados é feita com uso de scanners; para isso foram desenvolvidos programas de entrada por scanners e edição dos mapas rasterizados.

O módulo de entrada de dados do SAGA é um software destinado ao processamento de mapas rasterizados, contendo as funções e a precisão necessárias para processar produtos cartográficos, sendo baseado em scanners de baixo custo. A captura da geometria do mapa é feita por programa próprio, também podendo ser aceitos mapas digitalizados por outros softwares em formato TIFF.

A entrada de dados do SAGA possui as seguintes operações: conversão de resolução, refinamento de linhas, junção de mapas e georeferenciamento. Estas operações preparam o mapa para o reconhecimento das feições e alocação dos respectivos atributos.

A identificação das feições é feita pelo programa TRAÇAVET. O procedimento consiste na perseguição da linha que compõe uma feição de modo interativo com o usuário. Esta perseguição possui vários níveis de automação e estes níveis estão relacionados com a capacidade de superar as dúvidas da

perseguição do cursor. Assim, o nível 1 persegue a linha que for contínua com largura de um pixel, já o nível 4, e último, prossegue a perseguição mesmo com falhas de até 2 pixels e decide sobre caminhos, em certos casos, quando apresenta-se mais de uma opção. Esta forma de identificação das feições é extremamente interessante, já que permite ao usuário corrigir as possíveis falhas existentes no mapa rasterizado de modo interativo com a máquina.

Os atributos são alocados ao mesmo tempo que as feições são identificadas, as quais recebem um código, o que possibilita a implementação de ligações com bancos de dados convencionais. A partir do conjunto de feições identificadas, podem ser gerados diversos mapas, por exemplo, as estradas podem compor o mapa básico ou, ainda, serem apresentadas no mapa de uso do solo. Isto é possível, já que na edição são criados arquivos isolados contendo as classes de feições, podendo-se, assim, com elas compor e gerar vários mapas.

Por fim, deve-se mencionar os arquivos de pequeno tamanho que são gerados por este sistema, com uma taxa de compressão acima de 50%. Este aspecto é bastante importante, já que é comum acreditar-se que o tamanho dos arquivos raster é excessivo e limitante para sua adoção como estrutura de armazenamento.

5. CONCLUSÕES

Sendo a digitalização por mesa digitalizadora um processo tedioso,

sujeito a erros humanos e, com a queda de preços dos scanners de mesa, a captura da geometria das feições por scanners vem se tornando uma boa alternativa. Deve-se lembrar, no entanto, que o tempo demandado pelos processamentos prévios e posteriores à digitalização, quando dependentes de softwares externos aos SIG's, pode tornar desvantajoso o uso de scanners, principalmente, pela escassez de programas específicos de edição de mapas em formato raster para SIG's.

O tamanho dos arquivos gerados pelo scanner, quando não compactados convenientemente, pode limitar o uso destes, devido ao tempo consumido para o processamento dos dados.

A superação destas limitações conduz a viabilidade do uso de scanner na digitalização de mapas para SIG's de modo mais efetivo. A entrada de dados por scanner do SAGA-UFRJ, baseada em equipamentos de baixo custo, diferentemente de outras alternativas que existem no mercado, apresenta-se com vantagens que merecem atenção.

6. BIBLIOGRAFIA

- Aronoff, S. Geographic Information Systems: A Management Perspective. WDL Publications. 1989.
- Braga Filho, J.R.; Xavier-da-Silva, J.; Oliveira, O. M.; Pinheiro, N. F. Uma Entrada de Dados para SGI's. Anais da IV Conferência Latinoamericana sobre SIG. 123-134, 1993.
- Brandalize, A.A. Formatos de Arquivos: Chega de Quebrar a Cabeça. Revista Fator GIS, 2(2): 7-9, 1993.
- Cartensen, L.W.; Campbell, J. B. Desktop Scanning for Cartographic Digitization and Spatial Analysis. Photogrammetric Engineering and Remote Sensing. 57(11): 1437-1446, 1991.
- Graça, L.M.A. O Uso de Scanners para digitalização de Cartas Topográficas e para Implantação de um Sistema de Geo-informações Urbanas. Anais do Simpósio de Brasileiro de Geoprocessamento. 219-224, 1990.
- Petrie, G. Digital Mapping Technology: Procedures and Applications. In: Engineering Surveying Technology. Blackie and Son Wiley & Son Inc. 328-390, 1990.
- Peuquet, D.J; Boyle, A.R. Interactions between the Cartographic Document and Digitizing Process. In: Basic Readings in GIS. SPAD Systems. 35-43, 1984.
- Teixeira, A.L.A.; Gerardi, L.H.O.; Ferreira, M.C. Sistema de Informações Geográficas: Revisão e Comentários. Apostila do IV Curso de Sensoriamento Remoto - Interpretação de Imagens de Satélite. UNESP, Rio Claro, 1990.
- Xavier-da-Silva, J.; Saito, C.H.; Braga Filho, J.R.; Oliveira, O; Pinheiro, N.F. Um Banco de Dados Ambientais para Amazônia. Revista Brasileira de Geografia, IBGE, 53(3):91-124,1991.