ESTRATÉGIAS PARA AUTOMAÇÃO DE REPRESENTAÇOES DE DADOS EM AMBIENTES DIGITAIS

MSC. MARCIO AUGUSTO REOLON SCHMIDT
PROFA. DRA. LUCIENE STAMATO DELAZARI

Universidade Federal do Paraná - UFPR
Curso de Pós-Graduação em Ciências Geodésicas - CPGCG
marcio.schmidt@gmail.com , luciene@ufpr.br

RESUMO – Frente a disseminação de programas para representação de dados, uma linha de pesquisa tem tratado da geração automática de visualizações. Programas desse tipo objetivam evitar a necessidade do usuário de conhecer, projetar e especificar uma nova visualização para cada possível resultado de um programa de análise de dados. Originalmente desenvolvidos para a análise estatística, o uso cada vez mais intenso de elementos e técnicas da Cartografia, por parte desses programas de visualização automática, justifica um estudo comparativo para se buscar qual a relação desses com a Cartografia e como a combinação das qualidades de cada uma das técnicas pode melhorar o entendimento e a construção de representações. Por outro lado o estudo desses programas pode colaborar com a construção de sistemas que auxiliem o usuário com pouca experiência em Cartografia na criação de visualizações para exploração de dados assegurando o uso adequado da linguagem cartográfica.

ABSTRACT – Face the dissemination of programs for data representation, a research line has working on the automatic generation of data visualization. These softwares aim to avoid the necessity of the user to know, to project and to specify a new visualization for each possible result of a data analysis program. Originally developed to statistics analysis, the use each time more intense of elements and techniques of the Cartography justifies this comparative study to search which the relation of these techniques with the Cartography and as the combination of their qualities can improve the agreement and the construction of representations. On the other hand the study of these programs can contribute with the systems construction that assist the user with little experience in Cartography on map-creation for data exploration assuring an adequate use of the cartographic language.

1 INTRODUÇÃO

Esse artigo propõe uma análise das características de programas de visualização científica em relação a Cartografia. Entretanto, mesmo se tratando de protótipos resultantes de alguns anos de pesquisa, encontrar referências sobre os mesmos e suas aplicações é uma tarefa complexa. Em 1999, Domik (1999) fez uma seleção de nove estratégias desenvolvidas em diferentes instituições pelo mundo para aplicação de conceitos de ViSC. Dessas nove estratégias, somente as quatro primeiras foram abordadas nesse estudo por partirem de um mesmo paradigma. Esse paradigma é o da Geração e Teste sob enfoque da emulação de raciocínio conforme a inteligência artificial simbólica.

Segundo Domik (1999), Geração e Teste é um paradigma para sistemas de visualização automática cujo objetivo é eliminar a necessidade de especificar, projetar e desenvolver uma visualização para cada resposta de um programa. Além disso, os usuários podem focar sua atenção mais apropriadamente nas tarefas de determinar e descrever a informação a ser apresentada nessa visualização. Esse paradigma de Geração e Teste requer um conjunto de primitivas gráficas, regras para gerar a visualização a partir dessas primitivas gráficas, e critérios sob os quais são testadas as visualizações geradas. Por analogia, no domínio da linguagem natural, palavras formam as primitivas da língua portuguesa e o conhecimento do mundo real é o critério pelo qual se testa a sentença para se determinar a validade da mesma.

Domik (1999) ainda afirma que as regras para testar as alternativas geradas são derivadas da combinação de estudos experimentais e orientações baseadas nesses experimentos. As regras se referem a qual tipo de informação (nominal, ordinal ou quantitativa) pode ser expressa por diferentes primitivas, e a eficácia relativa de primitivas alternativas que poderiam ser usadas para o mesmo tipo de informação. Por exemplo, tamanho e saturação podem ser ambos usados para expressar dados quantitativos, mas tamanho é mais efetivo que saturação.

As estratégias selecionadas para análise são:

1. Mackinlay: Ferramenta de apresentação gráfica (APT);
2. Casner: (BOZ);

3. Roth e Mattis: Sistema para explicação gráfica automática (SAGE);

4. Senay e Ignatius: Assistente de visualização (VISTA).

Após a descrição das abordagens, são feitas análises e discussões, de modo a buscar qual a relação desses com a Cartografia e como a combinação das qualidades de cada uma das técnicas pode melhorar o entendimento e a construção de representações**.**

2 ABORDAGENS

a) Mackinlay, J. - APT: FERRAMENTA DE APRESENTAÇÃO.

Segundo Mackinglay (1986) esse sistema trabalha com apresentação de dados discretos bidimensionais com gráficos estatísticos como barra, linhas, torta. As primitivas gráficas incluem áreas (como círculos e barras), linhas, marcas (símbolos pontuais como ponto ou como x) e atributos visuais como cor, orientação e tamanho. As primitivas são organizadas em linguagens gráficas. Cada uma destas linguagens usa uma ou mais primitivas gráficas. A linguagem gráfica inclui: eixo horizontal, eixo vertical, gráfico de linha, gráfico de barras, gráfico de dispersão, cor, forma, tamanho, saturação, textura, orientação, árvore e rede. A gramática para geração de representações alternativas é definida pelas regras de composição, que combinam a linguagem gráfica dentro de gráficos complexos. As regras são: composição bidimensional e composição unidimensional.

Para Mackinglay (1986), a regra de composição bidimensional, nessa abordagem, pode compor sentenças gráficas com eixos horizontal e vertical idênticos e a regra de composições unidimensionais alinha duas sentenças gráficas com eixos horizontais ou verticais idênticos. Por exemplo, em dois gráficos de barras lado a lado, um simbolizando o PIB por país e outro a população por país, o eixo comum é país, enquanto que PIB e população são os eixos diferentes. Este autor descreve que no APT as apresentações gráficas são sentenças formadas a partir das línguas gráficas, e são similares a outras línguas formais que têm definições sintáticas e semânticas precisas. Esta análise conduz aos critérios de expressividade e de eficácia para avaliar projetos gráficos e uma álgebra da composição para gerar alternativas do projeto.

Os critérios da expressividade identificam as línguas gráficas que expressam a informação desejada (Mackinglay, 1986). Um conjunto de fatos é expressivo em uma linguagem se esse conjunto contém uma sentença que codifique todos os fatos do conjunto e não codifique dados adicionais incorretos. O autor desse programa ainda afirma que os critérios da eficiência identificam qual destas línguas gráficas, dada uma situação, é a mais eficaz em explorar as potencialidades do meio de saída e do sistema visual humano. O critério de eficiência pode ser baseado em diferentes fatores como acuracidade, rapidez ou mesmo custo, dependendo da interpretação do usuário. Para cada representação gerada, as primitivas são filtradas, com seus critérios de expressividade usados para gerar uma lista de possíveis representações (Domik, 1999). Os critérios de efetividade são usados para ordenar as possíveis representações de modo que a mais efetiva possa ser a primeira escolha. A álgebra de composição, segundo o autor, consiste em uma base contenha as linguagens gráficas (figura 1), cada uma de englobando uma das técnicas gráficas de Bertin para codificar a informação e os operadores da composição.

Na figura 1, abaixo, está uma demonstração do total de carros vendidos no ano de 1979 por países utilizando a geração automática no APT. A primitiva para quantidade de carros é a linha, nessa abordagem entendida como a primitiva barra. As sentenças gramaticais são os eixos vertical e horizontal relativos a modelos de automóveis e países respectivamente. E linguagem que faz a composição, ou seja, a relação entre as duas informações.

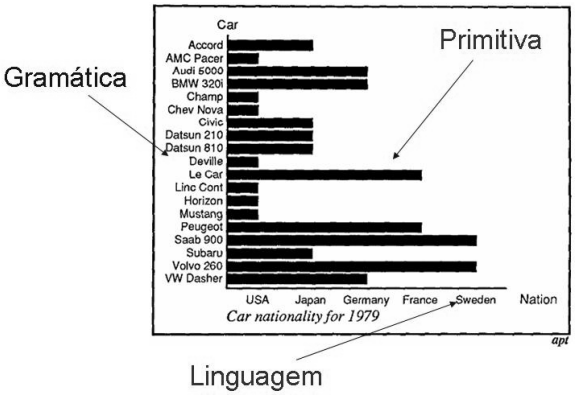

Figura 1: Exemplo de uso do programa APT. Fonte: Mackinglay, 1986, adaptado pelo autor.

b) Casner,S. – BOZ

Segundo Casner (1991), BOZ é uma ferramenta gráfica automatizada de desenho e apresentação que projeta gráficos baseados em análises da tarefa para a qual é desenvolvido. O objetivo desse programa é aperfeiçoar o processamento e representação da informação de modo a permitir aos usuários substituir inferências lógicas complexas por inferências perceptivas simples, e dinamizar a busca dos usuários para informação necessária.

O programa analisa uma descrição lógica feita pelo usuário através de uma tarefa a ser executada (descrição através de operadores lógicos como SQL, por exemplo), e cria uma tarefa perceptiva equivalente. Então o programa projeta e renderiza um gráfico que codifica a estrutura dos dados. Essa codificação é interativa procurando atender a todas as inferências perceptivas criadas. Paralelo à criação do gráfico, é criado um procedimento que descreve como usar o gráfico. Quando todas as inferências forem atendidas, o programa é encerrado resultando uma imagem. Segundo Casner (1991), uma característica chave do programa BOZ é a possibilidade de se criar diferentes representações de uma mesma

informação preparada que atendam às exigências de tarefas diferentes.

Segundo Domik (1999), o BOZ usa os seguintes cinco componentes para realizar a análise baseada em tarefas:

• Componente de tarefa lógica de descrição de linguagem: provê um meio de descrever as atividades de processamento da informação que a representação deve atender. Esta linguagem, a entrada do BOZ, contém dois componentes básicos: uma descrição dos procedimentos lógicos (como funções de programação de computadores), e uma anotação para expressar fatos lógicos manipulados pelas funções lógicas;

• Componente de substituição de operador perceptivo: cada operador lógico (*Logical Operator Perceptive* - LOP) é classificado em uma classe equivalente de operadores perceptivos. Esta substituição se baseia em um catálogo de operadores perceptivos, e uma linguagem gráfica;

• Componente da estrutura de dados perceptiva: esta parte analisa a relação entre operadores pela representação de cada operador como um vetor definido de um dos conjuntos de dados. Um diagrama completo de todas as relações entre vetores revela como a informação deve ser agregada em objetos gráficos e dividida entre as representações. Há quatro tipos de relacionamentos vetoriais contemplados no BOZ: conjunção, paralelismo, ortogonalidade e disjunção;

• Componente de seleção de operador perceptivo: esse componente escolhe um único operador perceptivo de uma lista de possibilidades geradas pelo componente de substituição de operadores perceptivos. Há três importantes considerações: o programa BOZ procura escolher o operador perceptivo mais eficiente e acurado segundo o operador humano; o BOZ usa um ranqueamento baseado em Mackinlay (APT) para encontrar o melhor operador perceptivo; a seleção de operadores perceptivos deve estar associada com a linguagem gráfica para codificar os fatos lógicos manipulados pelo operador. E, se dois operadores puderem ser codificados no mesmo objeto gráfico e estes não são combináveis, então ambos devem ser desqualificados;

• Componente de renderização: esse último componente traduz fatos lógicos em gráficos. O gráfico gerado pelo componente de renderização suporta interações entre conjunto de dados e imagens gráficas. Além disso, é capaz de mudanças efetivas na representação gráfica através da manipulação de fatos internamente, o usuário pode mudar fatos lógicos armazenados pela manipulação de objetos gráficos na representação.

c) Roth,S. et al - SAGE – SISTEMA PARA EXPLICAÇÃO GRÁFICA AUTOMÁTICA

SAGE também é baseado no paradigma de Geração e Teste e, segundo seus criadores, Roth et al (1994a), este é um sistema automático de apresentação que contém muitas características de sistemas relacionados como APT e BOZ. As entradas são caracterizações dos dados a serem visualizados e da aplicação da representação para um usuário. As operações do projeto incluem selecionar as técnicas baseadas em critérios de expressividade e eficácia, compô-las e apresentá-las como gráficos apropriados aos dados e aos objetivos do usuário. O programa usa variações e sínteses de representações estáticas 2D encontradas em pacotes gráficos estatísticos (gráficos de barras e linhas, técnicas visuais usando forma, cor ou tamanho e tabelas). O tipo de informação com a qual o sistema é abastecido são as relações entre os dados no banco de dados relacional. Roth et al (1994b) afirmam que as especificações do usuário servem como diretrizes do projeto e direcionam o algoritmo de busca na seleção e composição de gráficos para criar um projeto.

Roth et al (1994b) descrevem detalhadamente em seu artigo como o SAGE faz a escolha das representações baseadas nas características dos dados. Essas características são: ordenação, coordenadas versus quantidades, domínio da natureza do dado e dependência algébrica.

• Ordenação: A ordem dos elementos do conjunto de dados é um critério para escolha da técnica gráfica no SAGE. Uma técnica de ordenação pode ser exemplificada como quantitativa, ordinal ou nominal;

• Coordenadas versus Quantidades: O programa é capaz de reconhecer que elementos de um conjunto ordenado são coordenadas se cada elemento especificar um ponto ou localização temporalmente ou espacialmente. Em contrapartida, dados como peso e temperatura são classificados como quantidades;

• Domínio da Natureza do Dado: A caracterização do SAGE vai além da caracterização do APT por reconhecer que conjuntos podem pertencer a diferentes domínios de tempo, espaço, temperatura ou massa. Essa informação ajuda a preservar sutis convenções estilísticas, como por exemplo, uso do eixo horizontal para coordenadas e vertical para temperaturas. Para o SAGE três propriedades são definidas as quais descrevem como as relações são feitas de um conjunto de dados para outro. São elas a Cobertura Relacional que descreve se cada elemento de um conjunto pode ser mapeado com no mínimo um elemento de outro conjunto; a Cardinalidade, que expressa o número de elementos de um conjunto que possam mapear outro elemento em outro conjunto de dados; e a Unicidade que se refere à relação do mapa a um único valor para cada elemento do conjunto;

• Dependência Algébrica: Dependência algébrica entre elementos da base de dados sugere outra dimensão que pode afetar uma representação. Dependências (relações) podem ocorre entre atributos ou entre valores de um conjunto de dados. Por exemplo, um banco de dados organizacional pode conter três relações para custos: custos de material, custo de mão de obra e custo total.

O SAGE usa a técnica de segmentação da representação total em sub-representações para coordenar textos e visualizações por conversão de um esboço

preparado por um processador em uma lista de sub-representações. Isto indica ao sistema gráfico que a informação expressa em porções contíguas de texto deve ser considerada relacionada, e por isso mostrada em conjunto. Generalizando, esta característica provê um meio para expressar dois objetivos: a necessidade de expressar o máximo de informação possível e a necessidade de selecionar partes de conjuntos ou relações para visualização mais fácil e coesa. A figura 2, abaixo, apresenta uma representação da campanha de Napoleão de 1812. Nele se percebem a gramática como eixos representando latitude e longitude, a linha como primitiva gráfica e as variáveis visuais da cartografia temática tom de cor e tamanho usadas como linguagem para expressar variações de temperatura e quantidade de homens no exército da campanha.

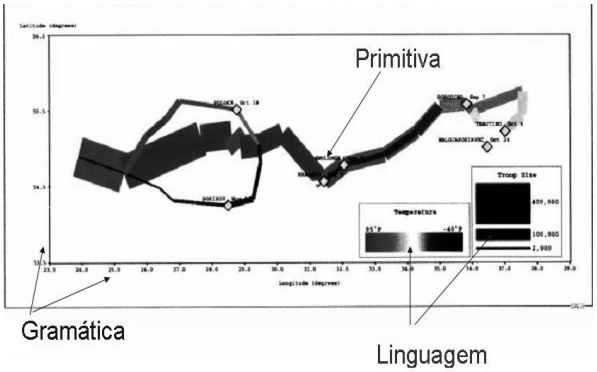

Figura 2: Exemplo de uso do SAGE – campanha de Napoleão, 1812. Fonte: Roth et al (1994b), adaptado pelo autor.

d) Senay, H., Ignatius, E. - VISTA: ASSISTENTE DE VISUALIZAÇÃO.

O programa VISTA é um sistema baseado em conhecimento sobre percepção gráfica que ajuda cientistas a desenvolver técnicas de visualização. Esse programa estende a metodologia de ferramenta automática da apresentação (APT) de Mackinglay (1986). A função preliminar do VISTA é gerar uma representação eficaz para um conjunto de dados, permitindo aos usuários modificar interativamente a representação.

Na criação da visualização, as várias características dos dados ditam o tipo de gráfico que expressará melhor os mesmos. Assim, saber as características dos dados relevantes à visualização é a primeira etapa na escolha de uma técnica eficaz de representação dos dados usando esse programa. Os autores desse programa entendem que se pode descrever técnicas de visualização em termos de marcas simples tais como pontos, linhas, áreas, e volumes. Pode-se também combinar diversos gráficos primitivos para visualizar uma série de dados multidimensional. Como em outras estratégias, o VISTA se utiliza de uma analogia: a linguagem humana e aplica regras para combinar símbolos e técnicas de visualização. Segundo Senay e Ignatius (1994), a arquitetura do VISTA é dividida em três principais unidades.

A primeira é a unidade de dados, na qual o usuário pode selecionar um conjunto de operadores de manipulação interativa de dados. Existem alguns menus dinâmicos a partir dos quais o usuário pode selecionar subconjuntos de dados do banco de dados por meio da seleção de variáveis de um ou mais conjuntos de dados. A ordem de seleção define um ordenamento padrão de importância, podendo ser redefinido posteriormente pelo usuário.

A segunda parte é a unidade de desenho, na qual o sistema gera uma técnica de visualização composta em três passos: decomposição do conjunto de informações, seleção da técnica de representação para cada subconjunto criado e composição da visualização. Uma vez que o VISTA tenha gerado o usuário pode modificá-la na unidade de desenho. Uma vez que todas as técnicas de visualização correspondentes às partições de dados estejam combinadas, a unidade de desenho apresenta uma descrição da técnica de composição da visualização ao usuário para possíveis modificações. Quando o usuário termina o processo de desenho, a unidade de desenho produz uma visualização personalizada que codifica todo o subconjunto selecionado no estágio de manipulação.

A terceira parte é a unidade de renderização. Esta unidade cria uma imagem de acordo com a descrição da técnica através de algoritmos de renderização apropriados. Para isso, a unidade usa regras de composição para unificar as primitivas criadas por cada algoritmo de renderização. Por último, a unidade de renderização mostra a imagem resultante desse processo na tela do usuário. A unidade de renderização também prove facilidades interativas para manipulação de imagens, por exemplo, rotação, translação, zoom e outros.

Para o entendimento do programa VISTA é necessário entender o que Senay & Ignatius (1994) colocam como vocabulário de visualização, técnicas primitivas de visualização e regras de composição. O vocabulário de visualização identifica os grupos de técnicas de visualização científica de dados. Em visualização de dados, um símbolo é o bloco mais primitivo que pode codificar alguma informação. Um símbolo pode ser simples ou composto. Há quatro tipos de primitivas gráficas para símbolos simples: pontos, linhas, áreas, e volumes. Um símbolo composto é uma coleção de primitivas gráficas que formam uma única unidade de perceptiva, por exemplo, linhas de contorno, setas de fluxo, e partículas.

Os autores entendem que símbolos têm propriedades posicionais, temporais, e propriedades visuais. Uma codificação posicional mostra a variação de coordenadas dos símbolos. Uma codificação temporal mostra como o dado se comporta no tempo. Uma codificação visual identifica qualquer variação das marcas que o olho pode perceber além de posição, como tamanho, saturação, tom ou orientação. Senay & Ignatius (1994) afirmam que técnicas primitivas de visualização codificam uma variável em uma das três categorias:

posicional, temporal, ou visual, dependendo da propriedade do símbolo primário manipulado. Técnicas posicionais podem ser uni/, bi/ ou tridimensionais. A única técnica temporal é animação. Técnicas visuais correspondem ao conjunto de variáveis visuais dos símbolos, chamada pelos autores de propriedades de retina. Na figura 3, abaixo, são mostradas as técnicas primitivas do VISTA. Algumas destas técnicas são composições de outras, porém os autores entendem que estas são técnicas de visualização primitivas.

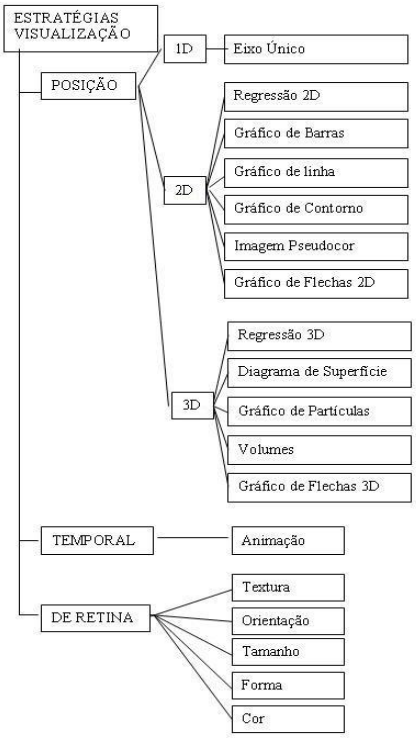

Figura 3: Técnicas primitivas de visualização.fonte: Domik, 1999

Um importante ponto na estratégia do VISTA são as regras de composição. Segundo Senay & Ignatius (1994), essas regras são necessárias para mostrar dados multidimensionais e definem as condições sob as quais o VISTA pode combinar pares de técnicas de visualização. Analogamente, estas regras são a gramática que o programa usa. Existem cinco regras de composição com as quais se pode descrever uma grande quantidade de técnicas compostas de visualização que são utilizadas pelo programa:

- Composição de símbolos: combina símbolos das técnicas de visualização componentes, através da comparação dos símbolos usados em um conjunto de dados com um conjunto compatível de símbolos no outro conjunto;
- Composição por superposição: combina símbolos das técnicas de visualização componentes sobrepondo uma a outra;
- Composição por união: combina símbolos gráficos pela união de representações;
- Composição por transparência: combina um par de técnicas de visualizações pela manipulação de opacidade dos símbolos pertencentes às mesmas;
- Composição por intersecção: combina um par de técnicas de visualizações computando sua intersecção e então sobrepondo a interseção em uma das técnicas componentes.

3 ANÁLISES E DISCUSSÕES

Os programas analisados foram desenvolvidos para a aplicação de visualização científica, porém muitas das técnicas e recursos utilizadas nesses sistemas são provenientes da Cartografia. Mesmo se tratando de protótipos resultantes de alguns anos de pesquisa, o acesso às referências sobre os mesmos e suas aplicações é bastante restrito. Destaca-se que são programas inovadores no sentindo de dotar um sistema de autonomia na criação de qualquer tipo de representação gráfica cabendo ao usuário apenas decidir se a visualização criada atende ou não à sua necessidade. Entretanto, apesar de empregarem elementos usados na Cartografia para a criação de representações de dados há um afastamento conceitual dessa, de modo que não existe clareza na aplicação de variáveis e regras. Cada autor de sistema ViSC tem suas próprias concepções sobre como definir primitivas gráficas, regras e critérios para a escolha de visualizações.

Os programas analisados, conforme apresenta o Quadro 1, apesar de terem sido desenvolvidos para visualização científica, trazem muitos elementos de representação cartográfica. Um aspecto ligado ao uso do mesmo paradigma é que todos utilizam como critérios para seleção de visualizações, os critérios de eficiência e eficácia, definidos originalmente por Mackinglay (1986).

Quadro 1: quadro comparativo entre as principais estratégias acadêmicas de visualização científica.

ESTRATÉGIAS ACADÊMICAS DE VISUALIZAÇÃO CIENTÍFICA				
	Mackinlay, J. **APT** (1986)	Casner,S. **BOZ** (1991)	Roth et al, **SAGE** (1994)	Senay, H., Ignatius, E. **VISTA** (1994)
Tipos de gráficos trabalhados	Gráficos estatísticos (barra, linhas, torta) e regressão.	Gráficos estatísticos (barra, linhas, torta) e derivados.	Gráficos estatísticos (barra, linhas, torta), regressão e derivados.	
As primitivas gráficas	Áreas (círculos e barras), linhas, símbolos pontuais e atributos visuais como cor, orientação e tamanho.		Mesmo de APT e BOZ acrescido de tabelas.	Pontos, linhas, áreas, volumes e símbolos complexos.
Linguagem gráfica	Eixo horizontal, eixo vertical, gráfico de linha, gráfico de barras, gráfico de dispersão, cor, forma, tamanho, saturação, textura, orientação, árvore e redes.		Análise das características dos dados.	Técnicas de visualização posicional, temporal e de retina.
Regras de composição	Composição bidimensional e composição unidimensional.	Escolha de operador perceptivo (variáveis de Bertin e tipos de gráficos estatísticos).		Composição de símbolos por superposição, união, transparência, ou intersecção.
Critérios	Expressividade e eficiência.			
Entrada de dados	Banco de dados.	Descrição lógica como SQL.	Banco de dados relacional.	Descrição lógica como SQL.
Enfoque	Escolher a visualização mais eficiente.	Escolher o operador perceptivo mais eficiente.	Objetivos do usuário na visualização dos dados.	Baseada em percepção visual.
Paradigma	Geração e Teste.			

FONTE: pelo autor.

Outro aspecto é que em todas as abordagens a estrutura dos programas adota a emulação do raciocínio através de uma analogia com a estrutura de linguagem humana. Além disso, em todos os programas analisados o que os pesquisadores definem como primitivas gráficas diferem grandemente do conceito de primitivas gráficas em Cartografia. Por isso os principais pontos desses programas são discutidos segundo a ótica da Cartografia:

a) Primitivas gráficas e linguagem gráfica:

Em Cartografia, para representar um fenômeno geográfico são utilizadas as primitivas gráficas ponto, linha, área e volume e as variáveis visuais que aliadas ao nível de medida formam o trino para representação temática. Esses elementos são provenientes de resultados de pesquisas em visualização e comunicação cartográfica.

Na descrição feita pelos autores de sistemas ViSC, existem grandes diferenças entre as noções de primitiva gráfica adotada nesses programas e na Cartografia . Na abordagem desses programas, primitivas gráficas são entendidas como elementos básicos constituintes de qualquer tipo de representação gráfica. Ao se trabalhar com gráficos estatísticos como gráfico de barras, linhas e gráfico tipo torta, por exemplo, as primitivas gráficas são barra, linha e círculo, respectivamente. Tabelas e símbolos pontuais, como formas geométricas simples ou compostas, são consideradas primitivas gráficas nos programas analisados. O que esses autores definem como símbolos complexos, na Cartografia são entendidos como símbolos pictóricos. Símbolos pictóricos trazem além da informação posicional, uma imagem associada ao símbolo que permite a associação direta ao fenômeno representado. Por exemplo, duas picaretas cruzadas indicam uma mina, o desenho de talheres indica restaurante.

As definições de primitivas gráficas como aparecem nesses programas são confusas. Como exemplo, barra e linha são definidas por Mackinglay (APT, 1986) e Casner (BOZ, 1991) como duas primitivas gráficas distintas, enquanto que na Cartografia a linha é considerada a primitiva gráfica e, neste caso, com uso da variável visual tamanho. Ou seja, a espessura da barra traz a idéia de uma quantidade associada à primitiva gráfica linha. A principal implicação da confusão entre primitivas gráficas e variáveis visuais é a impossibilidade de uma clara definição de como representar um elemento ou uma classe do fenômeno que se está representando. Notadamente, podem surgir representações ambíguas ou mesmo incompreensíveis que não transmitem a informação da maneira mais adequada ou, ainda, mascarando discrepâncias presentes nos dados.

b) A analogia de linguagens:

Apesar de poucas variações entre os autores, todos usam a emulação da inteligência segundo a estrutura da língua humana escrita. Os programas consideram os elementos componentes como elementos gramaticais direcionando o uso dos símbolos na representação dos dados. Esses elementos provêm de gráficos estatísticos, algumas variáveis visuais da Cartografia e tabelas. Com esses elementos os símbolos são construídos sendo considerados análogos às palavras e definidos segundo regras da linguagem gráfica de cada programa. Os autores diferem sobre quais símbolos são mais apropriados para representações bem como quais linguagens gráficas.

Para a composição das representações todo o espaço de distribuição do fenômeno é avaliado e regras de composição, semelhantes a regras gramaticais, fazem a escolha do tipo de representação final. O conjunto de todas as regras definidas em um desses programas é chamado de linguagem gráfica. As regras de composição diferem de autor para autor. Mackinglay (1986) usa regras para definir gráficos estatísticos 2D em eixos cartesianos. Seu programa (APT) foi o precursor dos programas de visualização científica e trabalha basicamente com gráficos estatísticos cartesianos uni/ e bi/ dimensionais definindo as suas variáveis ao longo desses eixos.

Casner (BOZ, 1991) e Roth et al (SAGE, 1994), confundem algumas variáveis visuais definidas por Bertin para linguagem cartográfica e alguns gráficos estatísticos no estabelecimento de suas regras de composição. Casner introduz o conceito de operadores perceptivos (POP). Esses operadores têm função análoga aquela dos operadores lógicos quando, por exemplo, realizam consultam em banco de dados. Os POP, segundo Casner (1991), caracterizam as atividades de processamento da informação executadas dentro do contexto de uma apresentação gráfica e seu desempenho depende da finalidade da representação gráfica que se está criando. Por exemplo, julgar a distância entre dois objetos em um gráfico e encontrar um objeto que tem uma cor particular são exemplos de operadores perceptivos. Esses operadores são organizados em função da linguagem gráfica disponível pelo programa. A linguagem gráfica para posição de um objeto na representação usada pelo BOZ inclui regras de composição como posição horizontal e vertical, altura, largura, área, conectividade, forma, tabelas, e outros. Ainda segundo Casner (1991) para linguagem posição horizontal, como exemplo, é admitido um conjunto de operadores perceptivos capazes de determinar: a posição horizontal de um único objeto gráfico, comparar posições de dois ou mais objetos horizontais (distância entre objetos), e encontrar o ponto médio entre duas posições horizontais. Esses operadores são: determinar posição horizontal, procurar objetos na posição horizontal, procurar posição horizontal de qualquer objeto, verificar objeto na posição horizontal, coincidência horizontal, à esquerda de, à direita de, determinar distância horizontal. Cada linguagem definida dentro do programa tem um conjunto próprio de operadores. Esses POP são similares no SAGE (Roth et al, 1994). Esse programa utiliza regras parecidas para definir suas regras de composição.

Senay e Ignatius (1994), com o VISTA, usam parâmetros da teoria de conjuntos para definição das regras de composição. Esses autores entendem que os dados estão sempre distribuídos no domínio da natureza do mesmo, ou seja, os dados sempre representam coordenadas, quantidades, números inteiros, e assim por diante. Esse programa difere dos demais pelas regras de composição definidas como composição de símbolos, por superposição, por união, por transparência e por intersecção. A regra que merece destaque é a composição por transparência, na qual dois temas são representados conjuntamente havendo entre eles uma diferença de opacidade. Dessa forma ambos podem ser vistos na mesma representação mas tem suas características alteradas, como tom, luminosidade, entre outras.

Comparativamente, a Cartografia apresenta uma melhor e mais clara definição dessas regras constituintes de representação gráfica. O que os autores tratam de maneira muito ampla como regras de composição, na Cartografia são definidas como relações topológicas. Nos programas de ViSC, cada autor dos programas analisados tem sua própria concepção de regras e como elas devem funcionar. Essas relações têm uma amplitude mais restrita na Cartografia , pois em função de toda experiência dessa arte, as relações topológicas são tanto ou mais acuradas e aplicáveis que as regras de composição, além de serem bem definidas. As regras topológicas são bastante aplicadas em programas de SIG e relativamente de fácil utilização pelo usuário. Questões como interceptação, união, continência ou mesmo relações entre coordenadas, são exemplos de regras topológicas simples e que substituem as complicadas regras de composição.

Essa falta de uniformidade entre as regras de composição no tratamento do dado, às vezes não deixa claro como os programas tratam os dados ou porque a regra é usada. Semelhante ao que pode ocorrer a um sistema especialista, um programa composto de muitas regras ou regras confusas gera resultados dúbios ou não condizentes com a solução mais adequada.

c) Caracterização dos dados de entrada:

Os programas comportam dois tipos de entradas: por banco de dados relacional e por descrição de pesquisa lógica. A entrada por BD relacional é atualmente uma caracterização comum a programas de mapeamento, como SIG e Atlas. Porém não há armazenamento de objetos gráficos nesses bancos, uma vez que a representação é gerada automaticamente. A caracterização por consulta lógica se assemelha às consultas SQL desenvolvidos nos programas SIG. O usuário insere uma sentença de questionamento composta de operadores lógicos e campos do BD, e o programa retorna somente aqueles dados que atendam a esse questionamento. A representação é realizada pela aplicação de regras de composição e primitivas gráficas sobre os dados selecionados.

d) Critérios de expressividade e eficiência:

Esses critérios são bastante subjetivos ficando a cargo do autor do mapa fazer a avaliação final dos critérios e fazer suas descrições nos programas. Segundo Mackinglay (1986), esses critérios identificam as linguagens gráficas que expressam a informação desejada de modo que todo o conjunto de dados envolvido na representação possa ser codificado adequadamente. Esse critério exige como condição a existência de uma sentença que codifique todos os dados do conjunto corretamente e somente os dados desse conjunto. Se forem codificados dados incorretos nessa sentença, então essa sentença continua válida, porém os programas têm de ser capazes de formularem outras sentenças mais adequadas. Dessa forma uma lista de possíveis representações é criada para cada intervenção do usuário. O critério de efetividade é usado para ordenar as possíveis representações de modo que a mais efetiva possa ser a primeira escolha. Mas esse critério, da forma como é aplicado nos programas analisados, considera somente os dados e a capacidade do programa em representá-los. Não há considerações sobre as possíveis interpretações pelo usuário.

O critério de eficiência não fica bem definido pelos autores. Mackinglay (1986) afirma que os critérios de eficiência identificam qual das linguagens gráficas é a mais eficaz em explorar as potencialidades do meio de saída, digital ou analógico, e do sistema visual humano em determinada situação. Não há nenhuma alusão de como esse critério pode ser aplicado para avaliação tomando por base o sistema visual humano. Outro ponto é sua definição, totalmente subjetiva. O próprio autor afirma que esse critério pode ser avaliado em termos de diferentes fatores como acuracidade e velocidade computacional na criação da representação ou mesmo custo financeiro. Assim esse critério pode assumir quaisquer valores dependendo da aplicação e de quem faz a análise desse critério. Com isso a comparação entre critérios fica prejudicada.

Portanto, quanto mais complexa seja a aplicação para a representação a ser criada, mais sensível é a escolha e a formulação dos critérios de expressividade e eficiência a ser inserida nesses programas.

e) Enfoque:

O aspecto que mais diferencia os programas de ViSC é o enfoque dado pelos seus idealizadores. Mackinglay (APT, 1986) procura escolher a visualização mais eficiente baseado nos critérios de expressividade e eficiência. Conforme descrito acima, esses critérios apesar de subjetivos são utilizados por todos os demais autores analisados. Casner (BOZ, 1991) propõe que a melhor visualização para um conjunto de dados se dá através da escolha do operador perceptivo mais eficiente. Essa escolha é feita com relações topológicas, chamadas de operadores perceptivos e agrupadas como linguagens gráficas, associada à relação entre os dados. Roth et al (SAGE, 1994) propõem que o objetivo do usuário na visualização é o fator decisivo para a escolha do tipo de representação. O usuário tem de fazer a especificação de modo particular ao usar o programa, inserindo a maior quantidade possível de informação descritiva suplementar aos dados para escolha adequada do tipo de representação. Senay e Ignatius (VISTA, 1994) criaram um programa baseado no conhecimento de percepção visual. Esse programa é o único, segundo a descrição dos próprios autores até aquele momento, que permite interatividade com o usuário de modo que este possa alterar a representação antes da sua finalização.

4 CONSIDERAÇÕES FINAIS

Os programas de ViSC analisados foram escolhidos pela procura por sistemas que emulam o conhecimento específico em visualização em ambiente digital para a construção automatizada de representações gráficas. Nos programas analisados, as semelhanças significativas podem ser resumidas ao paradigma adotado. Segundo o paradigma de geração e teste, a geração ocorre quando os dados são tratados, classificados e é escolhida uma simbologia para cada classe. Cada visualização antes de ser mostrada ao usuário é testada segundo os critérios definidos pelos autores. Esses critérios, chamados de expressividade e eficiência, atuam como classificadores de resultados, ou seja, esses critérios elencam quais as melhores representações e em que ordem devem ser disponibilizadas ao usuário. Dessa forma uma lista de possíveis representações é criada sendo apresentadas uma a uma até o usuário aceitar uma das opções.

Entretanto esses critérios são definidos de forma bastante subjetiva e o resultado fica condicionado ao conhecimento e a experiência do usuário e podem assumir quaisquer valores dependendo da aplicação e de quem faz a análise desses critérios. Portanto, quanto mais complexa seja a aplicação para a representação que se está criando, mais sensível é a escolha e a formulação do critério para ser inserida nesses programas. Com isso a comparação de critérios entre representações fica prejudicada e algumas representações não podem ser comparadas através dos mesmos. Como cada autor cria suas próprias regras de composição, cada programa passa a ter um sistema particular para escolha de regras para montagem e desenho de visualizações. Essas regras são confusas e seus objetivos nem sempre muito claros. Por terem função similar à topologia de dados geográficos, essas regras podem ser adaptadas à topologia como usadas em SIG, porém expandido o conceito para dados não geográficos.

Outro cuidado se refere à complexidade na definição das primitivas gráficas pelos autores. Essa definição deve ser foco para os programas de ViSC, ou seja, esses programas devem considerar, entre outras questões, o uso das primitivas gráficas como definido pela Cartografia . Como principal implicação da confusão de primitivas gráficas e variáveis visuais é o surgimento de representações ambíguas que não transmitem a informação da maneira mais adequada e que podem ocultar discrepâncias presentes nos dados.

Por outro lado, a construção automatizada de visualizações segundo o paradigma utilizado pode auxiliar a construção de sistemas de visualização cartográfica auxiliando o usuário na criação das mesmas e limitando a possibilidade de equívocos durante a execução do projeto cartográfico temático. Usuários com pouco conhecimento em Cartografia podem se utilizar das gerações automatizadas e testar as múltiplas representações geradas até encontrar uma que melhor expresse as relações entre os dados em cada etapa do processo de exploração visual, sendo assegurado o uso adequado dos classificadores, variáveis visuais e tipos de representação.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ANDRIENKO, G. L. & ANDRIENKO, N. V. Interactive maps for visual data exploration. **International Journal Geographic Information Science**, vol 13, No 4, 1999, pp 355-374.

ANDRIENKO, G. L. & ANDRIENKO, N. V. Tutorial Presented FHG AIS (Fraunhofer Institute for Autonomous Intelligent Systems) **ISESS conference**, Sesimbra Portugal, <http://www.ais.fraunhofer.de/and>. Acesso em 09/10/06.

CASNER, S. M. A Task-Analytic Approach to the Automated Design of Graphic Presentations. **ACM Transactions on Graphics (TOG)** archive, 1991. Volume 10 , Issue 2, p. 111 – 151. ACM Press New York, NY, USA.. ISSN:0730-0301 <http://doi.acm.org/10.1145/108360.108361>. Acesso em 21/05/2007.

DOMIK, G. 1999. Computer Graphics, Visualization and Image Processing. Palestra de ViSC. <wwwcs.uni-paderborn.de>. Acesso em outubro de 2006.

MACKINGLAY, J. Automating the Design of Graphical Presentations of Relational Information. **Transactions on Graphics (TOG)** archive, Volume 5, Issue 2, p. 110 – 141. ACM Press New York, NY, USA 1986.ISSN:0730-0301 <http://doi.acm.org/10.1145/22949.22950>. Acesso em 21/05/2007.

ROTH, S. F, KOLOJEJCHICK, J., MATTIS, J., CHUAH, M. C., GOLDSTEIN, J., JUAREZ, O. Sage Tools: A Knowledge-Based Environment for Designing and Perusing Data Visualizations. **Conference on Human Factors in Computing Systems**. 1994a. Boston, Massachusetts, USA. P. 27 - 28 ISBN:0-89791-651-4. <http://doi.acm.org/10.1145/259963.259992>. Acesso em 14/07/2007.

ROTH, S. F, KOLOJEJCHICK, J., MATTIS, J., GOLDSTEIN, J., Interactive Graphic Design Using Automatic Presentation Knowledge. **Conference on Human Factors in Computing Systems.** 1994b. Boston, Massachusetts, USA. P. 112 – 117. ISBN:0-89791-650-6. <http://doi.acm.org/10.1145/191666.191719>. Acesso em 14/04/2007.

SENAY, H. & IGNATIUS, E.; A Knowledge-Based System For Visualization Design Computer Graphics And Applications. 1994. **IEEE**, Volume 14, Issue 6, p. 36 – 47, doi: 10.1109/38.329093 acessado em <http://ieeexplore.ieee.org/iel1/38/7784/00329093.pdf>. Acesso em 25/05/2007.