

Análise Exploratória Visual Orientada a Ferramentas

CARLA MARIA DAL SASSO FREITAS
FLÁVIO RECH WAGNER

Universidade Federal do Rio Grande do Sul - Instituto de Informática
Caixa Postal 15064 91501-970 Porto Alegre, RS
carla, flavio@inf.ufrgs.br

Abstract. The paper presents the architecture of VISTA - an object-oriented system to support the tasks of visual exploratory analysis. Based on the scientific data analysis process and on several case studies, the tasks frequently executed by scientists in analysing their models are identified. Adequate tools are needed to support these tasks. VISTA is constituted by generic tools that can be applied for the analysis of models of different classes of entities.

1 Introdução

A coleta ou produção de dados representando entidades ou fenômenos é a etapa inicial da atividade científica. A partir daí inicia-se o processo de análise dos dados com vistas a estabelecer relações, verificar hipóteses e chegar ao entendimento do objeto de estudo.

À medida que o avanço tecnológico dos meios de coleta e de produção de dados propiciou um aumento considerável no volume e na complexidade dos mesmos, aumentou também a necessidade de exploração dos dados para fins de compreensão dos fenômenos que eles representam. Este é o objetivo principal do que é chamado de análise exploratória de dados [Tukey (1977)], aplicação principal da visualização científica. "Observar os dados para ver o que eles dizem" requer interagir com os dados, selecionar subconjuntos, procurar características, enfim, realizar *guided tours* pela massa de dados [Hurley-Buja (1990), Young-Rheingans (1991)]. As possibilidades de interação caracterizam, portanto, a análise exploratória de dados.

Isto não quer dizer que a visualização científica não se preocupe com a apresentação dos dados de forma a facilitar sua comunicação; de fato, muitas aplicações ditas de visualização são apresentação de dados de forma mais compreensível ao usuário leigo, sem a preocupação com os aspectos interativos de exploração. A diferença entre visualização científica e apresentação visual de dados é que nesta última a preocupação é a comunicação de resultados já previamente entendidos; já a visualização científica preocupa-se em possibilitar o entendimento destes dados [Earnshaw-Wiseman (1992)]. Desta forma, a análise exploratória visual de dados congrega um conjunto de técnicas e ferramentas destinadas a promover novas maneiras de solucionar problemas nas áreas científicas e de engenharia.

O presente trabalho descreve a arquitetura de VISTA, um ambiente de análise exploratória visual concebido a partir do estudo do processo de análise de dados científicos e da observação de situações variadas. As tarefas desempenhadas pelos cientistas são identificadas. Diferentemente dos sistemas conhecidos,

os quais implementam técnicas de visualização, VISTA explora o conceito de ferramentas numa transposição mais direta da realidade do processo de análise de dados científicos. As ferramentas visuais interativas adequadas ao suporte destas tarefas constituem uma "base de ferramentas". VISTA é concebida segundo o paradigma de orientação a objetos. Ferramentas são objetos; entidades e fenômenos são modelados como objetos, constituindo uma "base de entidades". Ferramentas são "aplicadas" a entidades, de forma que em VISTA a interação entre cientistas e os modelos das entidades sob análise é direta.

2 Objetivos, tarefas e ferramentas na análise exploratória visual

A análise de situações variadas, a saber, dados multivariados populacionais, dados multivariados espaciais, dados univariados e dados obtidos de análise de elementos finitos aplicada a objetos 3D permitiu identificar os objetivos da análise exploratória visual [Freitas-Wagner (1993)]. Concomitantemente, a análise destas situações à luz do processo de análise de dados científicos permitiu identificar as tarefas realizadas pelos cientistas na consecução de seus objetivos e, posteriormente, permitiu determinar classes de ferramentas adequadas ao suporte de tais tarefas.

A fundamentação da análise exploratória visual é de que o processo de análise de dados é mais facilmente efetuado sobre representações gráficas das características e propriedades das entidades em estudo. A rigor, tais características e propriedades são "mapeadas" para atributos visuais de representações gráficas diversas. Desta forma, o usuário pode construir um modelo mental da situação, a partir do modelo visual, constatando relações, testando hipóteses, ou seja, realizando todo um processo de investigação, onde a exploração visual e o questionamento são predominantes. A escolha do atributo visual que deve representar uma certa propriedade do objeto é dependente da informação que o usuário deseja extrair naquele contexto e da classe das entidades em estudo e seus atributos. Esta questão já foi abordada em outros

trabalhos [Lohse et al. (1990), Freitas-Wagner (1993)] podendo ser identificadas as seguintes classes de representações visuais: ícones, diagramas e redes, gráficos e tabelas, modelo geométrico e mapas.

Com base nos estudos de casos desenvolvidos, identificou-se o conjunto de objetivos da análise exploratória num primeiro nível de abstração [Freitas-Wagner (1993), Freitas (1994a)]:

- análise de objetos num contexto;
- análise da estrutura de um objeto;
- análise das propriedades estáticas de um objeto;
- análise das propriedades dinâmicas de um objeto;
- análise comparativa de dois ou mais objetos;
- comunicação de informações.

Cabe observar que objetos são considerados como pertencendo a uma de três grandes classes: sistemas de múltiplas entidades, objetos manufaturáveis e entidades ou fenômenos naturais. Objetos têm propriedades, ou atributos, distingüidos de acordo com a natureza da informação (característica, escalar, vetorial, agregação), a natureza dos valores (alfanumérico, numérico, símbolo), a natureza do domínio (discreto, contínuo, contínuo-discretizado) e a dimensão do domínio (1D, 2D, 3D, nD) [Freitas (1994a)].

Os objetivos citados acima são gerais e para atingí-los é necessária a realização de **operações** ou **tarefas**, algumas gerais, outras específicas para cada caso. Operações são identificadas de forma mais ou menos isolada por diversos autores [Wehrend-Lewis (1990), Beshers-Feiner (1993)]. Entretanto, considerando a análise exploratória visual como a realização da análise de dados científicos representados visualmente, é fundamental partir da observação deste processo. Outros autores [Springmeyer et al. (1992)] caracterizam o processo de análise de dados científicos e estabelecem ações ou tarefas que são desempenhadas por cientistas:

- **interação com representações** congrega ações de geração de representações visuais, exame das representações, orientação da representação visual, consulta a valores numéricos, comparação e classificação;
- **processamento** reúne três outras tarefas: estimativas e transformações, derivação de novas condições e geração de estatísticas;
- **manipulação de dados** reúne ações específicas de navegação, gerência de dados e seleção de subconjuntos de dados;
- **expressão de idéias** corresponde a registro de sessões de análise e descrição;

A consideração do ponto de vista destes últimos autores citados e das necessidades constatadas nos estudos de casos já mencionados levou à classificação das seguintes tarefas como básicas, pois, combinadas, permitem o cumprimento dos objetivos dos usuários no processo de análise exploratória visual [Freitas-Wagner (1994b)]:

- navegação pelo universo de objetos;
- seleção de um conjunto de objetos ou de parte de um objeto;
- consulta a atributos de objetos;
- preparação de processamento (modificação de atributos; inicialização de parâmetros);
- processamento (simulação, cálculos, monitoramento, controle);
- anotação de textos e armazenamento de dados e imagens.

As tarefas identificadas acima devem ser suportadas por recursos visuais interativos. Estes recursos são disponibilizados na forma de "ferramentas" definidas, em sua maioria, como um procedimento interativo sobre uma representação visual.

Ferramentas são classificadas como de

- mapeamento,
- exploração,
- preparação e controle de processamento e
- registro.

As ferramentas de **mapeamento** correspondem às facilidades de indicação de representações visuais para objetos e seus atributos. As ferramentas de **exploração** correspondem às facilidades de suporte à navegação, seleção e consulta dos objetos em estudo. As ferramentas de **preparação e controle de processamento** contemplam as facilidades de análise das informações existentes no modelo, através de métodos numéricos, estatísticos e de simulação. Além da geração de resultados a partir de um processamento específico, englobam a preparação do processamento, sua monitoração e controle. A preparação do processamento diz respeito ao estabelecimento das condições iniciais dos parâmetros; enquanto monitoração e controle do processamento implicam no acompanhamento do processamento, através da observação de atributos das entidades e, eventualmente, intervenção no mesmo. Já as ferramentas da classe **registro** correspondem às facilidades de anotação de textos em representações visuais e armazenamento de resultados numéricos e gráficos, para posterior recuperação ou reprodução em outros meios.

3 A arquitetura de VISTA

VISTA é organizada a partir de uma base de ferramentas e uma base de entidades, manipuladas através de uma interface icônica (Figuras 1 e 2). A abordagem de implementação é orientada a objetos. Tanto entidades como ferramentas são objetos. Existe, portanto, uma hierarquia de classes para a modelagem das entidades e, analogamente, hierarquias de ferramentas. A Figura 3 apresenta o modelo de objetos utilizado para descrever as classes de entidades tratadas em VISTA, no nível mais alto de abstração. Este modelo foi desenvolvido seguindo a metodologia OMT [Rumbaugh et al. (1991)].

A hierarquia de classes de entidades formaliza a classificação anteriormente exposta e discutida [Freitas-Wagner (1993)], incorporando as diferenças decorrentes

da natureza dos atributos, natureza dos valores, natureza e dimensão do domínio. Entidades da classe sistema são modelados como agregações de entidades primitivas ou de sub-sistemas. Objetos manufaturáveis são agregações de objetos componentes que, por sua vez, são constituídos por elementos geométricos. Entidades naturais podem ser de duas classes: amostras espaciais ou amostras populacionais (conjunto). A classe Amostra-espacial permite modelar dados amostrados espacialmente a respeito de algum fenômeno, enquanto Conjunto modela dados coletados de entidades tipo populações. Ambas as classes Elemento-geométrico e Elemento-espacial são compostas a partir de classes básicas como Ponto, Célula-2D e Célula-3D. Os atributos das entidades aparecem como variáveis das instâncias das classes utilizadas na sua representação.

As hierarquias de ferramentas contemplam as diferentes classes identificadas, compartilhando, entretanto, recursos básicos. A classe dos **mapeamentos** corresponde a ferramentas que permitem estabelecer a qualquer momento qual representação visual deve ser associada a uma entidade, atributo de entidade ou conjunto de atributos de entidades (ver Figuras 4 e 5). Representações alternativas podem estar ativas ao mesmo tempo, permitindo observações simultâneas dos atributos das entidades. Cada instância de mapeamento corresponde, na realidade, a um relacionamento dinâmico, que é estabelecido e, posteriormente, eliminado. As diferentes representações visuais, equivalentes em resultados, às técnicas de visualização amplamente conhecidas [Brodie et al. (1992)], são parametrizadas através dos mapeamentos. Por esta razão, a classe Mapeamento aparece como uma generalização de classes especializadas para cada tipo de representação visual.

No que diz respeito à navegação, a classe básica é o **cursor**. Conforme a navegação seja num espaço ou num conjunto de objetos, o recurso básico empregado é o **cursor espacial** ou o **cursor individual**. Um cursor espacial pode ser um **focalizador** (barra de *scroll* em janelas), para entidades com domínio 2D, ou uma **câmera sintética**, no caso de entidades definidas no espaço 3D. Cabe salientar que na aplicação de um mapeamento sobre uma entidade, uma ferramenta de navegação é também ativada.

As ferramentas que permitem a seleção de entidades são classes de **seletor**: um **seletor individual**, para indicação de entidade(s) exibida(s) na tela; **seletor espacial**, para delimitação física de uma parte do conjunto de entidades, e **seletor lógico**, pela determinação de um critério sobre os valores das propriedades das entidades. A classe seletor espacial é uma generalização de sub-classes dependentes da dimensão do domínio da entidade e da dimensão do sub-domínio desejado, como, por exemplo, uma janela, quando se quer limitar uma área a ser observada sobre uma região 2D ou sobre uma linha; um volume, quando o objeto está definido no espaço 3D e se quer limitar um

sub-volume; um plano de corte aplicado a um volume quando se quer obter um plano ou um subvolume; uma linha quando se quer limitar a consulta a esta categoria de entidade geométrica; um cursor espacial de ponto, quando se quer selecionar um ponto apenas do espaço-domínio.

Já um seletor lógico permite selecionar um subconjunto dos dados a partir da especificação de uma restrição sobre os valores de um ou mais atributos. A reprodução de uma situação de aplicação de um seletor lógico sobre uma entidade representando um volume de dados tomográficos, seguida da aplicação de um mapeamento para superfície, pode ser observada na Figura 6. Nesta figura é, também, exemplificado o processo interativo básico de um usuário de *VISTA*: tomar um ícone de uma ferramenta e aplicá-lo à entidade representada visualmente ou selecionada. Eventualmente, a ferramenta necessita de parâmetros e os solicita através de uma caixa de diálogo usual.

As ferramentas de consulta estão organizadas como uma hierarquia, onde especializações levam em conta a natureza da informação consultada: **leitor**, que serve a atributos tipo característica, permitindo apenas a exibição dos valores associados à entidade indicada, ou seja, presentes no modelo; **medidor escalar**, para o caso de obtenção de valores escalares, os quais, em geral, são calculados por interpolação ou aproximação, com base naqueles presentes no modelo; **medidor vetorial**, análogo ao medidor escalar, para grandezas vetoriais.

Há, entretanto, grandezas que são derivadas de atributos escalares ou vetoriais. Neste caso, é necessário um conjunto de **avaliadores**, que façam uso de funções de cálculo dependentes da aplicação. Estes avaliadores produzem novos dados a partir dos existentes e atuam sobre um sub-espaço ou subconjunto de entidades/atributos selecionados. Esta classe de recursos é, portanto, dependente da natureza das entidades, da natureza e dimensão do domínio dos atributos e constitui a classe genérica **processamento**. Nesta classe genérica de processamento, são também incluídas ferramentas de controle de simulação, quando for o caso.

As representações visuais geradas pelas ferramentas de mapeamento devem poder ser manipuladas graficamente e armazenadas para análise posterior ou, simplesmente, para apresentação. As ferramentas que reúnem os recursos necessários para tal são duas: ferramentas para anotação e para armazenamento de dados e imagens. **Anotação** permite apor um texto a uma representação visual de forma semelhante a uma legenda. O **armazenamento** permite indicar tanto dados como representações visuais correntemente exibidas que devem ser armazenados em arquivo para posterior recuperação.

4 Comparação com outras propostas

Sistemas de visualização científica como AVS [Upson et al. (1989)], Iris Explorer [Silicon Graphics (1991)] e MPGS [Cray Research (1991)] oferecem técnicas de

visualização que exigem como parâmetros informações que aqui associamos conceitualmente às ferramentas de mapeamento e exploração e, mais especificamente nesta última classe, navegação e seleção. Neste aspecto reside a maior contribuição do presente trabalho: parte-se dos objetivos e tarefas da análise exploratória visual e não das técnicas de visualização que se deseja oferecer. VISTA é, pois, baseada num referencial de ferramentas que todo sistema de apoio a esta classe de atividade deve prover.

Do ponto de vista de utilização, AVS e Iris Explorer são construtores de aplicativos, não sendo destinados propriamente a um usuário final como é o caso de MPGS. VISTA destina-se a ambas as classes de usuários pois a construção de aplicações em VISTA requer apenas a inclusão de novas ferramentas na "base de ferramentas" e o seu uso requer apenas a sua manipulação através da interface icônica.

No que se refere ao uso da abordagem orientada a objetos, alguns sistemas adotam este paradigma de implementação. O sistema proposto por Brittain et al. (1990) oferece como ferramentas várias técnicas de visualização tradicionalmente conhecidas (isosuperfície, visualização volumétrica, emissão de partículas, etc). Entretanto, não há, por exemplo, a distinção conceitual entre seleção de sub-domínios e respectivas representações visuais. É basicamente orientado à utilização interativa, não propiciando extensão fácil, nem sua utilização na construção de aplicativos. VISAGE [Schroeder et al. (1992)] é outro sistema de visualização orientado a objetos. É baseado na abordagem *data flow*, com duas grandes classes de objetos: dados e processos. Os objetos da classe dados permitem a modelagem de uma descrição geométrica com dados escalares e vetoriais associados; os objetos da classe processo tomam dados, processam segundo a técnica de visualização pretendida e geram descrições geométricas para *renderers*. Em VISTA, entidades e ferramentas são objetos. A aplicação de ferramentas a entidades equivale a troca de mensagens entre objetos, não havendo ao nível do usuário final distinção entre ferramentas da mesma classe aplicadas a entidades estruturalmente distintas.

VISTA pode ser utilizado por um usuário final, de forma interativa, uma vez que ferramentas podem ser escolhidas da base de ferramentas e aplicadas tanto sobre as representações icônicas das entidades como a uma representação visual previamente obtida através de um mapeamento. VISTA pode, ainda, ser utilizado como um construtor de aplicativos pela simples incorporação de especializações das ferramentas existentes. Novas ferramentas assim criadas são incorporadas à base de ferramentas e passam a estar disponíveis aos usuários.

Referências

- Beshers, C. & Feiner, S. Autovisual: Rule-based design of interactive multivariate visualizations. *IEEE Computer Graphics and Applications*, 13(4):41-49, julho 1993.
- Brittain, D.L.; Aler, J.; Wilson, M. & Wang, S.C. Design of an end-user data visualization system. In: VISUALIZATION'90. San Francisco, 23-26 Out., 1990. *Proceedings*. Los Alamitos, IEEE, 1990. pp.323-328.
- Brodie, K.W.; Carpenter, L.A.; Earnshaw, R.A.; Gallop, J.R.; Hubbard, R.J.; Mumford, A.M.; Osland, C.D. & Quarendon, P. (eds.) *Scientific Visualization - Techniques and Applications*. Berlin, Springer-Verlag, 1992.
- Cray Research. *Multipurpose Graphic System User Manual*. Eagan, 1991.
- Earnshaw, R.A. & Wiseman, N. *An Introductory Guide to Scientific Visualization*. Berlin, Springer-Verlag, 1992.
- Freitas, C.M.D.S. & Wagner, F.R. A methodology for selecting visual representations in scientific and simulation applications. In: SIMPÓSIO BRASILEIRO DE COMPUTAÇÃO GRÁFICA E PROCESSAMENTO DE IMAGENS. Recife, Outubro 19-22, 1993. *Anais*. Rio de Janeiro, SBC, 1993. pp. 89-97
- Freitas, C.M.D.S. *Uma abordagem unificada para análise exploratória simulação interativa visual*. Porto Alegre, PGCC/UFRGS, 1994. (Tese de doutorado)
- Freitas, C.M.D.S. & Wagner, F.R. Ferramentas de suporte às tarefas da análise exploratória visual. (submetido à Revista de Informática Teórica e Aplicada)
- Hurley, C. & Buja, A. Analyzing high-dimensional data with motion graphics. *SIAM J. Sci. & Statist. Comput.*, 11, 1193-1211, 1990.
- Rumbaugh, J.; Blaha, M.; Premerlani, W.; Eddy, F. & Lorenzen, W. *Object-oriented Modeling and Design*. Englewood Cliffs, N.J., Prentice-Hall, 1991.
- Schroeder, W.J.; Lorenzen, W.E.; Montanaro, G.D. & Volpe, C.R. VISAGE: an object-oriented scientific visualization system. In: VISUALIZATION'92. Boston, 19-23 Out., 1992. *Proceedings*. Los Alamitos, IEEE, 1992. pp. 219-226.
- Silicon Graphics. *Iris Explorer User's Guide*. La Jolla, CA, 1991.
- Springmeyer, R.R.; Blattner, M.M. & Max, N.L. A characterization of the scientific data analysis process. In: VISUALIZATION'92, Boston, Mass., Outubro 19-23, 1992. *Proceedings*. IEEE, Los Alamitos, 1992. pp. 235-242.
- Tukey, J.W. *Exploratory Data Analysis*. Reading, MA, Addison-Wesley, 1977.
- Upson, C.; Faulhaber Jr, T.; Kamins, D.; Laidlaw, D.; Schlegel, D.; Vroom, J.; Gurwitz, R. & Van Dam, A. The application visualization system: a computational environment for scientific visualization. *IEEE Computer Graphics & Applications*, 9(4):30-42, July 1989.

Wehrend, S. & Lewis, C. A problem-oriented classification of visualization techniques. In: VISUALIZATION'90. San Francisco, Outubro 23-26, 1990. *Proceedings*. IEEE, Los Alamitos, 1990. pp. 139-143.

Young, F.W. & Rheingans, P. Visualizing structure in high-dimensional multivariate data. *IBM Journal of Research and Development*, 35(1/2):97-107, jan./mar. 1991.

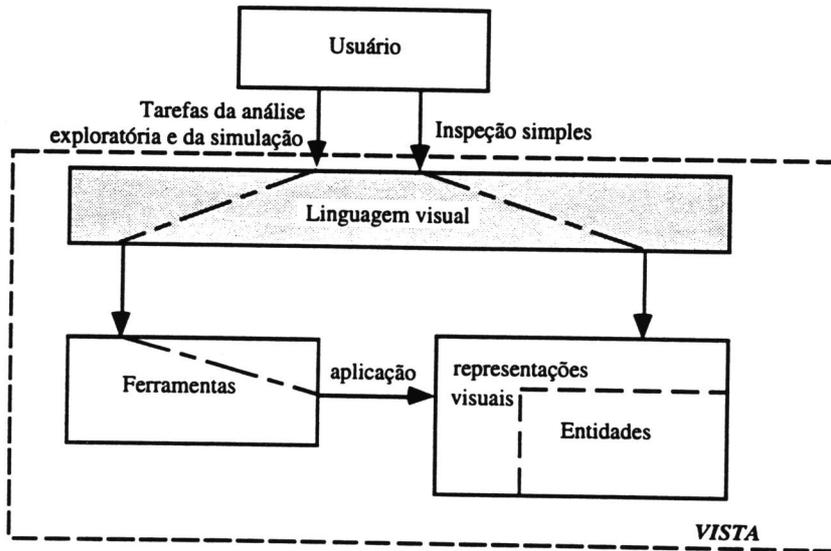


Figura 1 - Estrutura geral de VISTA

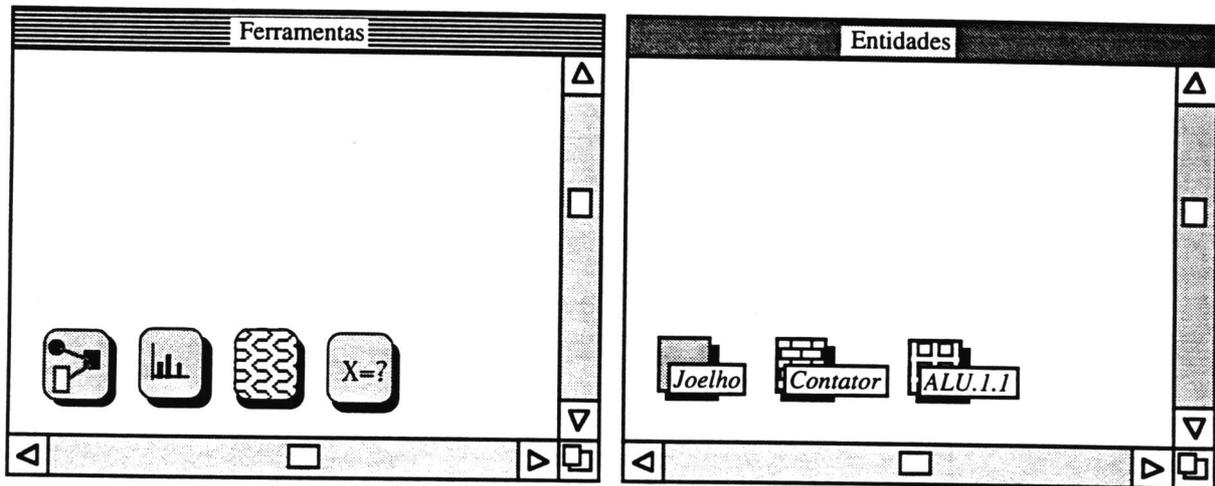


Figura 2: Base de ferramentas e base de entidades com representações icônicas.

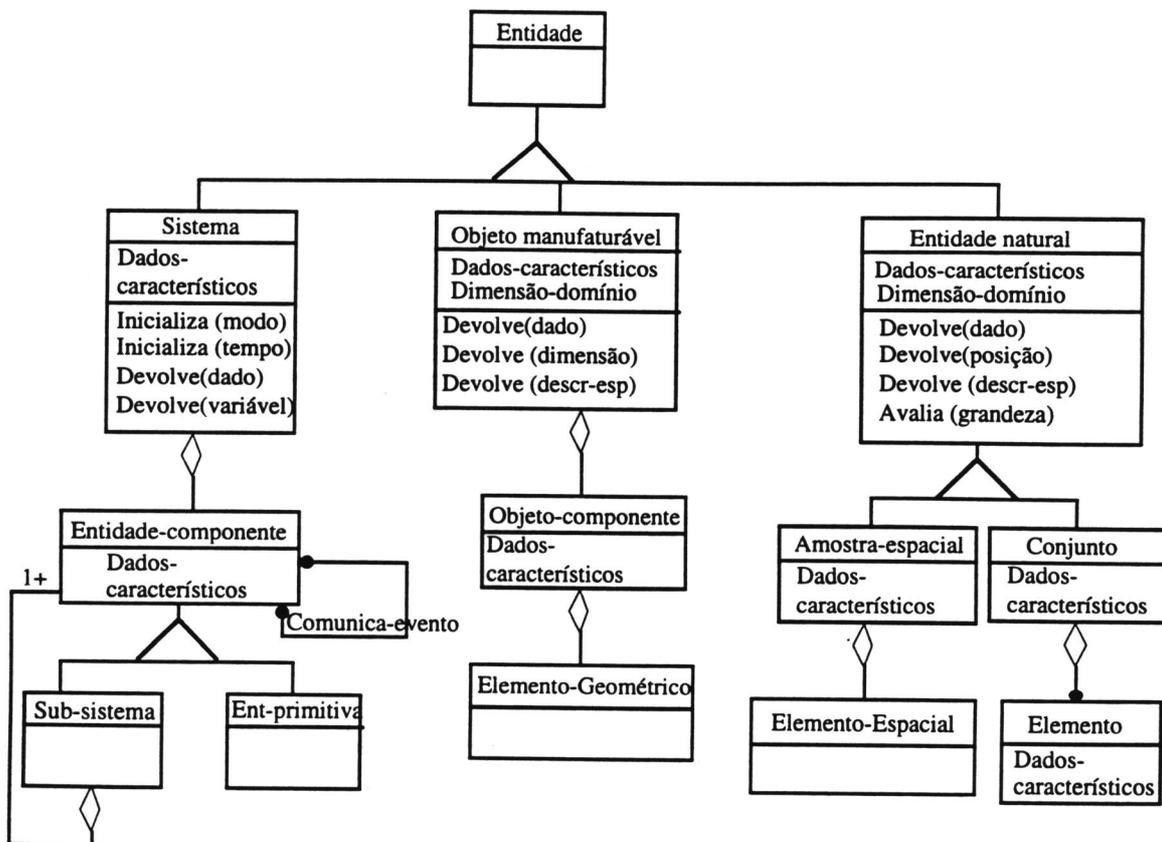


Figura 3: Modelo da classe Entidade.

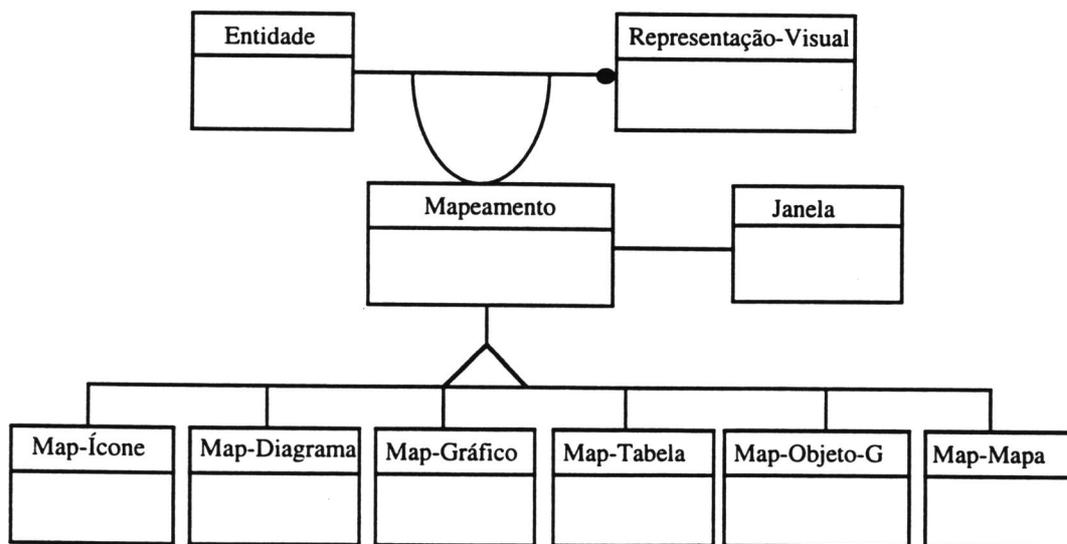


Figura 4: Ferramenta Mapeamento representada como uma associação entre Entidade e Representação-Visual.

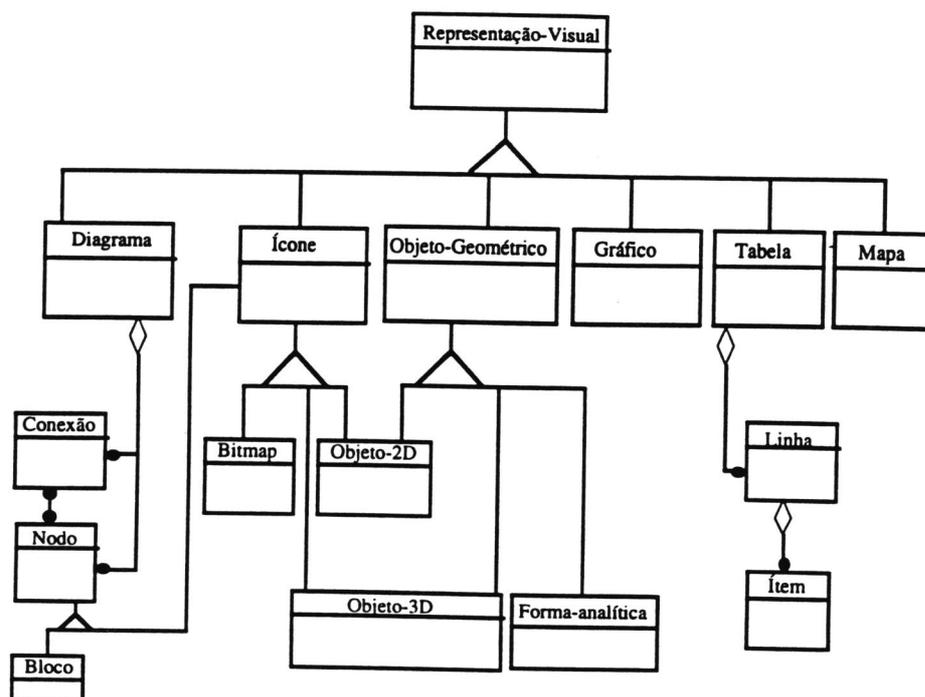


Figura 5: Hierarquia de classes para Representação-Visual.

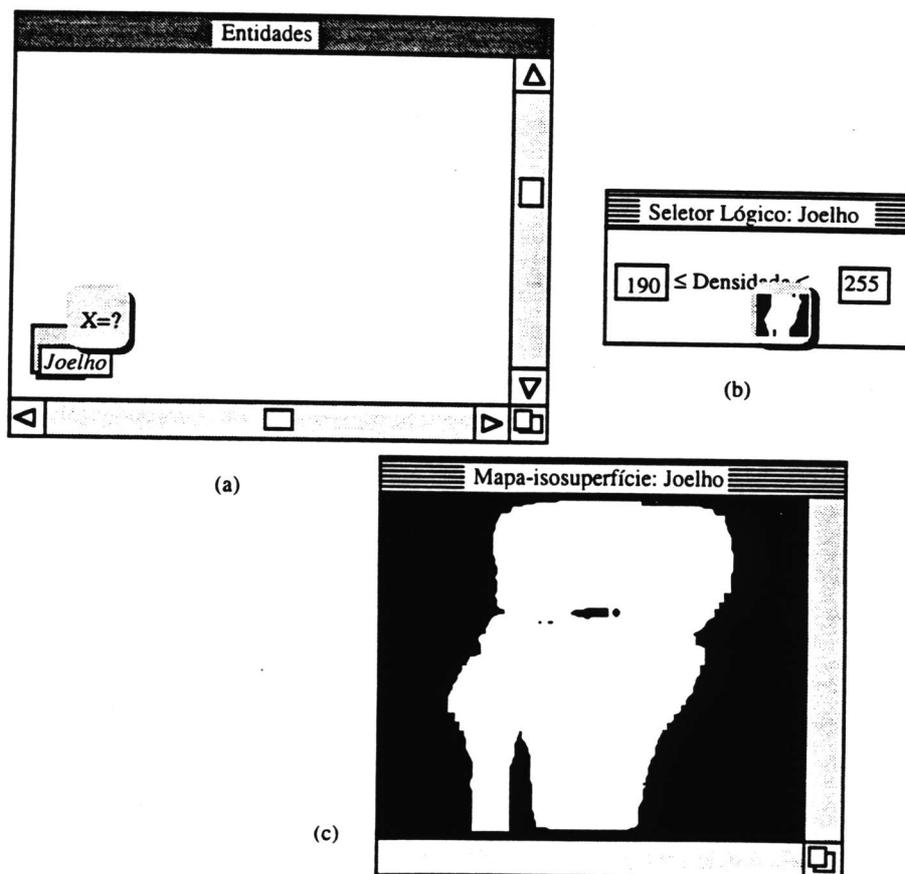


Figura 6: Aplicação de um seletor lógico sobre um objeto que representa um volume de dados tomográficos, seguida de um mapeamento para superfície da região selecionada no volume.

