

# COORDENAÇÃO BASEADA EM SIMILARIDADE EM MÚLTIPLAS VISÕES COORDENADAS

*Similarity-Based Coordination on multiple coordinated views*

**João Gabriel Gomes Nogueira<sup>1</sup>; Danilo de Medeiros Eler<sup>1</sup>;**

<sup>1</sup> **Faculdade de Ciências e Tecnologia (FCT/UNESP)**

**Departamento da Matemática e Computação**

Presidente Prudente, R. Roberto Símonsens, 305-Brasil

[joaogabriel17.0@hotmail.com](mailto:joaogabriel17.0@hotmail.com); [daniloelerunesp@gmail.com](mailto:daniloelerunesp@gmail.com);

**RESUMO** – Atualmente, muitas aplicações disponibilizam informações textuais, os quais podem ser provenientes de diversas áreas. Pesquisas envolvendo mineração de dados enfrentam o desafio de analisar grandes quantidades de elementos e atributos. Algumas abordagens que utilizam técnicas de visualização permitem que o usuário possa interagir com um conjunto de dados, e consiga observar padrões e tendências de maneira clara e eficiente. Porém, uma única técnica de visualização nem sempre consegue produzir bons resultados para todo tipo de dados, pois cada técnica possui seus pontos fortes e fracos. Na tentativa de mitigar os pontos fracos, diferentes técnicas são utilizadas de forma complementar, concedendo o usuário possuir diferentes visões sobre o mesmo conjunto de dados. Este trabalho tem como por objetivo o desenvolvimento de uma ferramenta que conecte diferentes técnicas de visualizações de dados (no qual isso conhecido como múltiplas visões coordenadas), e que viabilize a interação do usuário com um ou mais conjunto de dados, obtidos de um banco de dados (SQL) ou arquivos texto. Este trabalho também visa reconhecer os pontos fracos e fortes de técnicas de visualizações de dados e entender como o mecanismo de coordenação melhora o reconhecimento de padrões e tendências.

**Palavras-chave:** visualização de dados; múltiplas visões coordenadas; similaridade.

Recebido em: xx/xx/xxx
Revisado em: xx/xx/xxx
Aprovado em: xx/xx/xxx

**ABSTRACT** – Currently, many applications provide textual information, which can be which can be provided by several areas. Researches involving data mining face the challenge of analyze large number of elements and attributes. Some approaches that use visualization techniques allow the user interact with a data set, and note patterns and trends, clearly and efficiently. However, a single visualization technique not always can produce good results for all types of data, because every technique has their strengths and weaknesses. To mitigate weaknesses, different techniques are used complementarity, giving different views of the same data set to the user. This work has as its aim the development a tool to coordinate different data visualization techniques (which it called multiple coordinated views), and allowing the user interaction with one or more data set, obtained from a database (SQL) or text files. This work also aims to recognize the strengths and weaknesses of visualization techniques and understand how the coordination mechanism improves the recognizing patterns and trends.

**Keywords:** data visualization; multiple coordinated views; similarity.

## 1. INTRODUÇÃO

Atualmente, muitas aplicações disponibilizam informações textuais em grandes quantidades, os quais podem ser provenientes de diversas áreas. Pesquisas envolvendo mineração desses dados enfrentam o desafio de lidar com quantidade de elementos e de atributos desses dados. A integração de diferentes áreas e a utilização de uma abordagem visual para exploração e análise de conjuntos de dados é conhecida como Análise Visual (*Visual Analytics*) (Wong e Thomas, 2004).

Em particular, abordagens que utilizam técnicas de visualização são largamente empregadas com o objetivo de dar sentido a esses dados, por meio de representações visuais e de formas de interação (Card et al., 1999). As representações visuais, criadas a partir das informações contidas em um conjunto de dados, procuram realçar o relacionamento entre as instâncias ou entre os atributos que as descrevem. Dessa maneira, o usuário pode interagir sobre o conjunto de dados e descobrir padrões ou confirmar e levantar hipóteses. Além disso, as representações visuais também buscam facilitar a descoberta de anomalias presentes em um conjunto de dados. Algumas dessas técnicas facilitam a análise do conteúdo textual de coleções de documentos, revelando os principais tópicos ou palavras que estão presentes na coleção.

Apesar das diferentes técnicas de visualização de informação colaborarem para a exploração e análise de diferentes conjuntos de dados, elas possuem pontos fortes e fracos, o que não permite a utilização de uma única técnica para produzir bons resultados para

todo tipo de dados. Tomando como exemplo a bem conhecida técnica Coordenadas Paralelas, ela possui a vantagem de exibir muitos atributos simultaneamente e facilita a localização de agrupamentos de dados que tenham o mesmo comportamento. Um ponto fraco da técnica Coordenadas Paralelas é a dificuldade que o usuário tem em analisar um conjunto de dados com milhares de instâncias, não permitindo que seja encontrado com facilidade os agrupamentos. Na tentativa de suavizar os pontos fracos de uma técnica de visualização, diferentes técnicas podem ser utilizadas de forma complementar, possibilitando que o usuário tenha várias visões sobre um conjunto de dados. Uma visão, no contexto deste trabalho, significa uma representação visual de um conjunto de dados por meio de uma técnica de visualização. Utilizando várias técnicas de visualização sobre um conjunto de dados, tem-se várias visões, facilitando sua exploração. Adicionalmente, para auxiliar o usuário na troca entre as visões, pode-se utilizar um mecanismo de coordenação, o qual tem o objetivo de associar os elementos explorados nas diversas visões. A área que estuda como combinar múltiplas visões para obter resultados eficientes de associação entre informações é conhecida como Múltiplas Visões Coordenadas (*Coordinated and Multiple Views – CMV*), ou também *Linked Views*.

## 2. FUNDAMENTAÇÃO TEÓRICA

Entre as áreas da Ciência da Computação, a visualização de informação estuda técnicas para apresentar visualmente dados abstratos, de modo que seja possível identificar padrões e tendências, permitindo que informações importantes possam ser facilmente extraídas e tornando os dados mais compreensíveis.

Antes de tudo, alguns termos devem ser definidos. Uma amostra, significa a instancia de um dado. E uma amostra possui um ou mais atributos. Os atributos aqui utilizados possuem somente valores numéricos. Por exemplo, uma imagem pode ser considerada uma amostra dentro um conjunto de imagens, e os seus atributos podem ser caracterizados como comprimento, largura, luminosidade, saturação, matiz, entre outros valores numéricos.

Uma técnica de visualização de informação eficiente, consegue apresentar facilmente uma informação a partir de uma quantidade grande dados e atributos.

O primeiro exemplo é mostrado na Figura 1, que mostra a cotação do dólar em um determinado período, onde existe um parâmetro tempo, e a cotação respectiva em função do tempo. A principal informação que pode ser extraída dessa visualização é a valorização ou desvalorização da moeda em certo período.

O segundo exemplo, também conhecido, agora um pouco mais complexo, mostra uma Pirâmide etária, na Figura 2. Aqui, a extração da informação requer uma leitura mais lenta, pois existem diferentes valores para ser analisados.

Conforme maior o número de atributos e dados, maior a dificuldade de leitura, e maior a necessidade de técnicas de visualização que permitam uma extração eficiente de informações.

Por exemplo, quando temos três ou mais parâmetros para descrever uma informação, seria necessário mais de um gráfico.

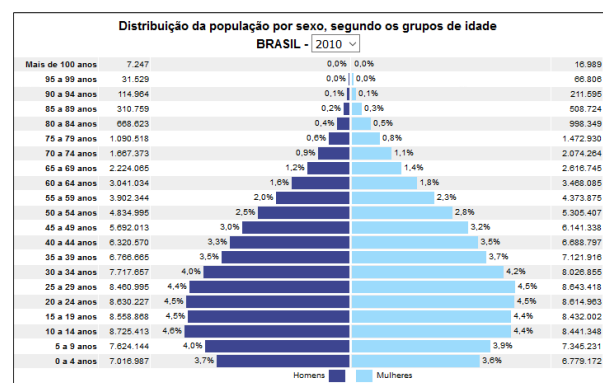
**Figura 1** – Exemplo simples de técnica de visualização de informação de um gráfico bidimensional, no caso está sendo mostrado a cotação do dólar durante um período tempo



Fonte:

<http://economia.uol.com.br/cotacoes/cambio/dolar-comercial-estados-unidos/>

**Figura 2** – Exemplo simples de técnica de visualização de informação de uma Pirâmide etária no ano de 2010 realizada pelo IBGE.



Fonte:

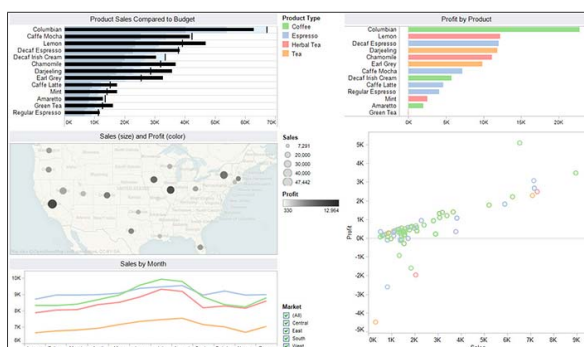
<http://www.censo2010.ibge.gov.br/sinopse/index.php?dados=12>

Uma técnica para melhorar a eficiência da extração de informações é utilizar diferentes técnicas de visualização, que foquem em apresentar uma informação específica. Tal técnica

é conhecida de múltiplas visões. Um exemplo é mostrado na Figura 3, onde existem cinco diferentes visões.

Cada visão utiliza uma técnica de visualização de dado diferente. Desse modo, quando uma visão não consegue apresentar bons resultados para um certo tipo de dado, uma ou mais outras visões podem mitigar esse ponto fraco e garantir a compreensão dos dados.

**Figura 3** – Múltiplas visões de um mesmo conjunto de dados, gerado por software.



Fonte:

<https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception>

Além disso, também pode se utilizar um mecanismo de coordenação, com o objetivo de facilitar a identificação das amostras entre as diferentes visões.

As cores podem ser utilizadas para facilitar a identificação de uma amostra, ou de um grupo de amostras, entre as visões que representam o mesmo conjunto de dados.

Uma coordenação por destaque pode ser realizada por meio de uma chave primária, destacando ou exibindo uma respectiva amostra selecionada em outras visões. Ou a coordenação pode ser realizada por meio similaridade onde as amostras mais similares são destacadas. Mais

detalhes desses mecanismos de coordenação serão apresentados nas próximas sessões.

## 2.1. Técnicas de Visualização

Dentre as técnicas de visualização existem duas que foram implementadas nesse trabalho: Coordenadas Paralelas e Projeção Multidimensional.

### 2.1.1. Coordenadas Paralelas

O gráfico de Coordenadas Paralelas utiliza linhas e colunas para apresentar a relação entre os valores de atributos, onde uma amostra é representada por uma linha, e um atributo é representado por uma coluna.

Um exemplo simples é apresentado pela Figura 4. Nesse exemplo existem cinco amostras (apresentados por cores diferentes) e 4 atributos.

### 2.1.2. Projeção Multidimensional

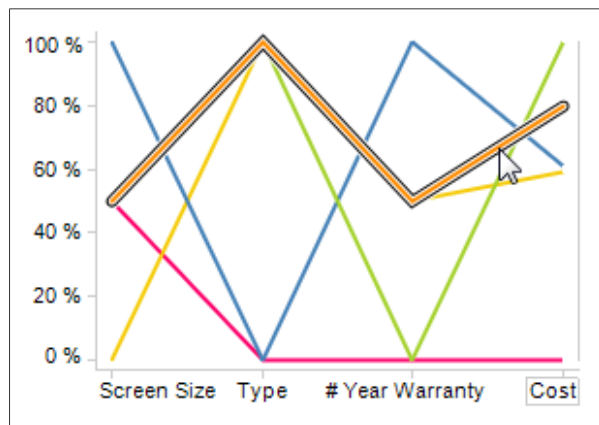
As projeções multidimensionais utilizam a técnica de Redução de Dimensionalidade para gerar pontos que são posicionados em um espaço bidimensional ou tridimensional. Esta técnica permite identificar facilmente a similaridade entre amostras através da distância entre elas. Um exemplo simples é apresentado pela Figura 5. Também conhecido como Scatter Plot ou Gráfico de Dispersão.

#### 2.1.2.1 Redução de Dimensionalidade

A dimensionalidade de uma amostra é a quantidade de atributos que ela possui. Até o momento presente, é impossível criar um gráfico com mais de 3 dimensões. A Redução de Dimensionalidade tenta utilizar um número menor

de dimensões para descrever o mesmo conjunto de dados.

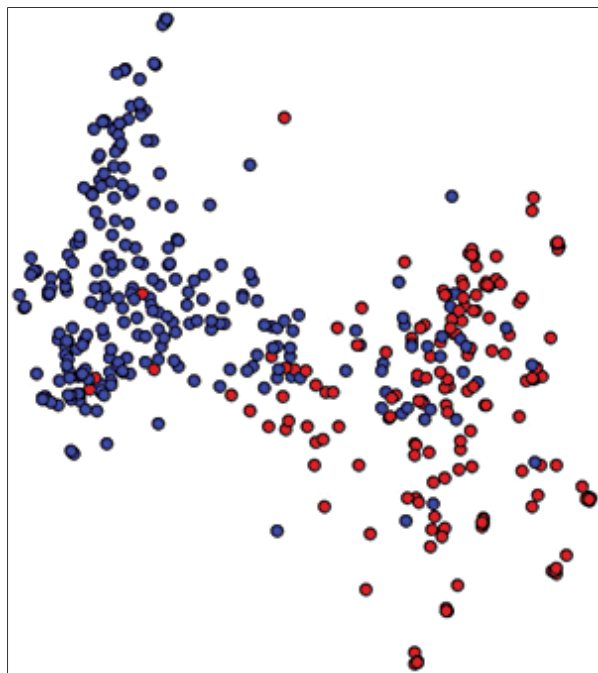
**Figura 4** – Exemplo de uma projeção das Coordenadas Paralelas.



Fonte:

[https://docs.tibco.com/pub/spotfire\\_web\\_player/6.0.0-november-2013/pt-BR/WebHelp/GUID-24AF55A4-E1E3-4F08-B62C-842738904252.html](https://docs.tibco.com/pub/spotfire_web_player/6.0.0-november-2013/pt-BR/WebHelp/GUID-24AF55A4-E1E3-4F08-B62C-842738904252.html)

**Figura 5** – Exemplo de resultado de uma Projeção Multidimensional.



Fonte: Haim Levkowitz, Fernando V. Paulovich, Rosane Minghim, Luis G. Nonato, Least Square Projection: A Fast High-Precision

Multidimensional Projection Technique and Its Application to Document Mapping

### 3. ABORDAGEM PROPOSTA

Observando os trabalhos desenvolvidos na área de Múltiplas Visões Coordenadas (do inglês, Coordinated Multiple Views - CMV) que trabalham com o modelo relacional, nota-se uma carência de técnicas de coordenação baseadas em similaridade, devido o modelo relacional não prover suporte para tal consulta entre as instâncias. Este trabalho tem como principal objetivo propor técnicas de coordenação baseadas em similaridade, por meio de desenvolvimento de uma ferramenta que exiba diferentes visões, em que cada visão apresente uma técnica de visualização de dados diferente. Um ou mais mecanismo de coordenação é utilizado para realizar essa conexão visual entre as diferentes técnicas de visualização por meio do destaque das amostras. Além disso, o usuário poderá interagir com as visões para obter mais detalhes sobre as amostras.

As bases de dados, fornecidas pelo usuário, deverá possuir os mesmos atributos em cada visão e a importação poderá ser realizada a partir de um banco de dados ou arquivo. A partir disso, as informações, como amostras, classe, cores, e atributos, são armazenadas pelo programa em uma estrutura de dados específica para cada visão.

Como já dito anteriormente as técnicas de visualização a serem implementadas nesse trabalho, são Coordenadas Paralelas e Projeção Multidimensional.

Logo após a implementação, bases de dados de diferentes quantidades de amostras e atributos serão utilizadas na validação da

abordagem proposta, visando analisar a eficiência de cada técnica. Assim, é possível identificar as limitações das técnicas que utilizam coordenações baseadas somente nos identificadores e os pontos fracos e fortes de cada técnica utilizada.

## 4. DESENVOLVIMENTO

A ferramenta, resultada desse trabalho, foi desenvolvida em linguagem de programação JAVA, e necessita de uma máquina virtual Java, versão 8 ou superior, para executá-la.

O software utiliza-se de recursos fornecidos da biblioteca Projection Explorer (PEX) para realizar a redução de dimensionalidade utilizada na Projeção Multidimensional.

### 3.1. Exportação da Base de Dados

O primeiro passo a ser realizado para a aplicação das técnicas de projeção é a exportação da base de dados.

#### 3.1.1 Exportação de Arquivo

A exportação pode ser realizada a partir de um arquivo, de formato de texto, com ponto e vírgula separando os atributos e quebra de linha para separar as amostras.

Esse formato é utilizado pela ferramenta Projection Explorer (PEX). Porém, nesse trabalho as três primeiras linhas são ignoradas pois a quarta informa automaticamente o número de atributos, e o número de linhas não necessitam ser definidos estaticamente.

#### 3.1.2 Exportação de Banco de Dados

A exportação via banco de dados utiliza o driver jdbc para realizar a conexão com a base de dados em Mysql.

O usuário deve informar os dados de conexão do MySQL, o nome da tabela a ser importada, qual é o atributo que indica o nome da amostra, quais são os atributos numéricos, e opcionalmente o atributo que caracteriza o grupo do qual a amostra pertence (por exemplo, nome ou Id do Grupo).

### 3.2 Projeções

Logo após a exportação da base dados, já é possível projetar as técnicas de visualização.

#### 3.2.1 Coordenadas Paralelas

A primeira visão que a ferramenta exibe logo após a exportação é a das Coordenadas Paralelas. Com isso, para cada amostra é calculada a posição relativa ao valor máximo e mínimo do atributo em relação as outras amostras. A equação dada para calcular a posição relativa na coluna é dada pela equação 1

$$PR(t) = \frac{\text{valor}(t) - \min_{[f_0]}(t)}{\max(t) - \min_{[f_0]}(t)} \quad (1)$$

Tendo o valor da posição relativa basta multiplicá-la pela largura da coluna para encontrar a posição absoluta.

$$PA = PR(t) * \text{Largura} \quad (2)$$

#### 3.2.2 Projeção Multidimensional

Com a utilização da biblioteca PEX, para calcular a redução de dimensionalidade, os pontos de cada amostra são posicionados nas suas respectivas posições em relação ao tamanho da visão (comprimento e largura da janela).

### 3.3 Múltiplas Visões



A técnica das Coordenadas Paralelas, e a Projeção Multidimensional podem utilizar a mesma base dados, ou utilizar base de dados diferentes, desde que possua o mesmo número de atributos.

### 3.4 Mecanismo de Coordenação

Para facilitar a identificação de amostras em ambas visões, é possível destacar uma amostra através da interação do usuário. E ao destacar uma amostra em uma visão, uma outra amostra (ou a mesma, dependendo o mecanismo de coordenação utilizado) é destacada na outra visão.

A interação do usuário com uma amostra descrita nessa sessão, pode ser uma simplesmente mover o mouse sobre a amostra.

#### 3.4.1 Coordenação por chave primária

A coordenação por chave primária utiliza somente do identificador da amostra entre as visões para realizar o destaque.

Quando o usuário interagir com uma amostra em uma visão, a mesma amostra é destacada na outra visão.

O problema dessa coordenação é que não pode ser utilizada entre bases de dados diferentes, pois mesmo que o identificador seja idêntico, as amostras não são similares.

#### 3.4.2 Coordenação por similaridade

A coordenação por similaridade utiliza da distância euclidiana entre as amostras.

Quando o usuário interagir com uma amostra em uma visão,  $n$  amostras mais similares serão destacadas na outra visão, o valor de  $n$  é definido pelo usuário.

A vantagem de não necessitar da chave primária, garante que tal mecanismo de coordenação possa ser utilizado entre base de dados diferentes.

### 3.4.2 Cores

Além da coordenação por destaque, a codificação por cor de cada amostra facilita a identificação da mesma amostra, ou do mesmo grupo de amostra na outra visão, sem a necessidade de interação do usuário.

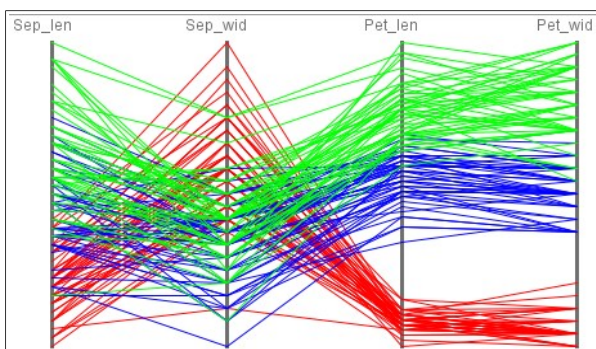
## 5. EXPERIMENTOS

Para validar as projeções das técnicas de visualização, diferentes bases de dados para foram exportadas.

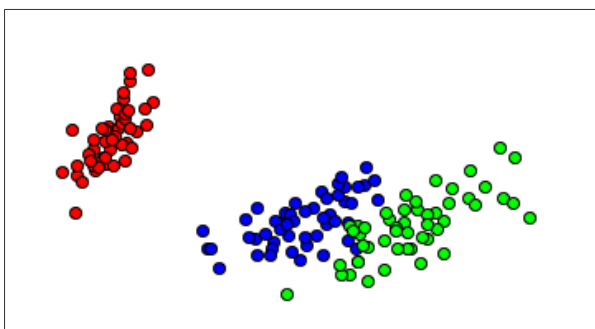
Inicialmente, as técnicas de visualização devem estar corretamente projetadas em cada visão. As Figura 6 e 7, apresentam o resultado das Coordenadas Paralelas e da Projeção Multidimensional, respectivamente, para a bem conhecida base de dados “Fisher's Iris”, que armazena características sobre uma espécie de flor (mais detalhes em <https://www.kaggle.com/uciml/iris>), com um total de 150 amostras divididos entre 3 grupos com 50 amostras cada, onde as amostras possuem 4 atributos. Cada grupo de amostras é representado uma mesma cor em ambas as visões. Na visão das Coordenadas Paralelas é possível identificar que o grupo em azul e o grupo em verde possuem valores muito similares para os atributos de Comprimento e Largura da Sépala (Sep\_len e Sep\_wid, respectivamente), enquanto o grupo em vermelho difere bastante dos demais grupos em relação aos atributos de Comprimento e Largura da Pétala (Pet\_wid e Pet\_len respectivamente).

Enquanto na Projeção Multidimensional, já não é possível ver detalhes sobre os valores dos atributos, é possível facilmente identificar quais grupos de amostras são mais similares, por meio do posicionamento dos pontos no plano de projeção (quanto mais próximo mais similares são as instâncias dos dados). No caso, o grupo em azul e em verde possuem amostras similares e não é tão trivial definir padrões distintos para esses grupos, sendo que ao escolher uma amostra qualquer, os similares podem pertencer a um grupo diferente, como apresentado nas Figuras 8 e 9. Isso não ocorre com amostras do grupo vermelho que ao selecionar uma amostra, os primeiros mais similares pertencem ao mesmo grupo, como apresentado na Figura 10. O destaque da ferramenta é feito pela coloração das amostras mais importantes e descoloração das demais.

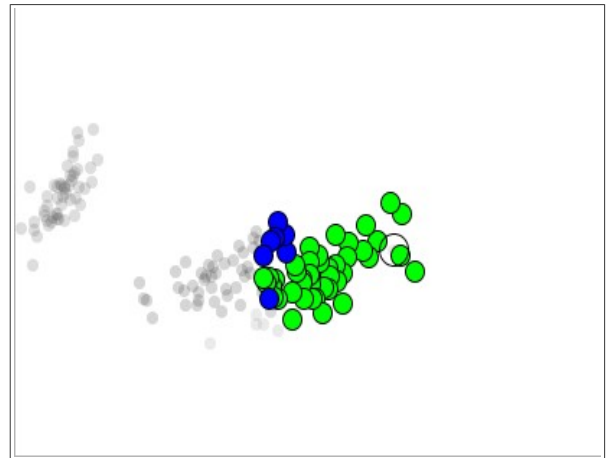
**Figura 6** – Resultado das Coordenadas Paralelas da base de dados Iris.



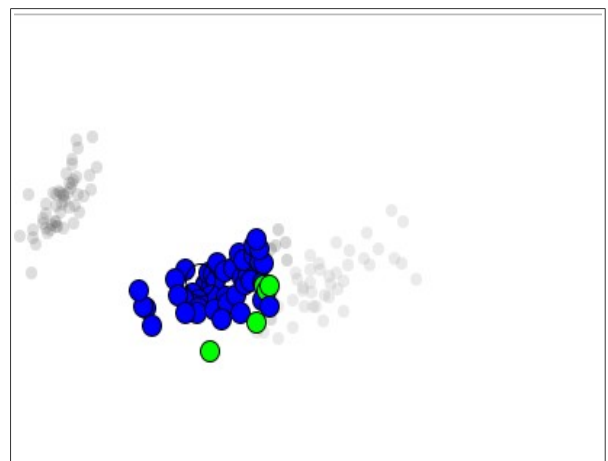
**Figura 7.** Resultado da Projeção Multidimensional da base de dados Iris.



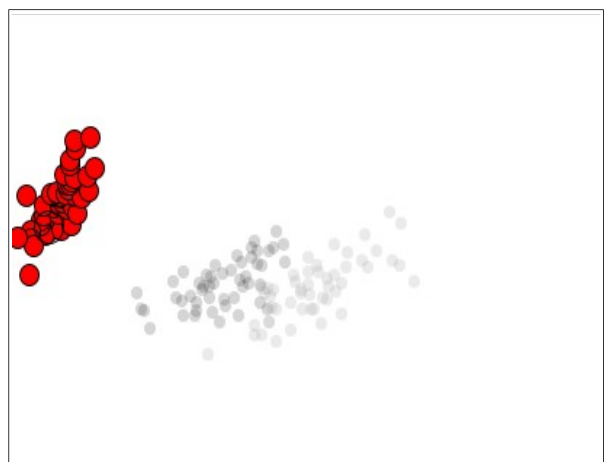
**Figura 8.** Visualizando as 50 amostras mais similares de uma amostra qualquer do grupo verde da base de dados Iris.



**Figura 9.** Visualizando as 50 amostras mais similares de uma amostra qualquer do grupo azul da base de dados Iris.



**Figura 10.** Visualizando os 50 amostras mais similares de uma amostra qualquer do grupo vermelho da base de dados Iris.





Outro conjunto de dados bem conhecido e que foi utilizado para a validação de abordagem proposta é formado por imagens de texturas de Brodatz (Handbook of Image and Video Processing, pg 455), que analisa algumas características de texturas em imagens. Foram extraídas 16 características utilizando filtro de Gabor[citar] e a projeção desse espaço de características é apresentado nas Figuras 11 e 12. Esta base possui 70 amostras divididas em 7 grupos com 10 amostras e cada amostra possui 16 atributos.

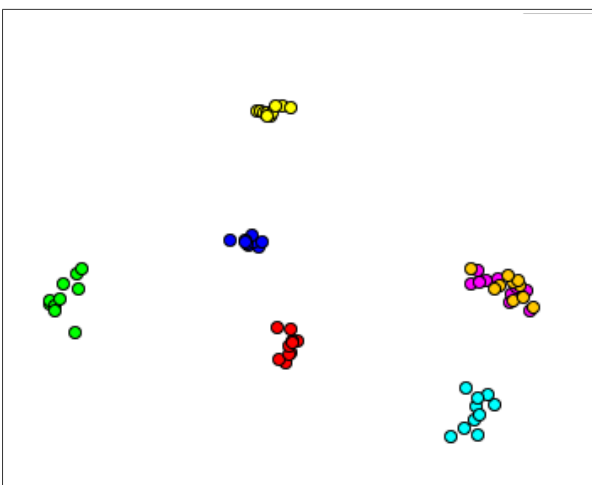
O grupo rosa e laranja de amostras de textura são bastantes similares (como apresentado na Figura 12 ), logo não é trivial identificar padrões entre as amostras desses grupos. Detalhes sobre a similaridade entre os atributos das amostras é apresentada na Figura 11, onde é visível que a maioria dos grupos de amostras não apresentam muitas diferenças em relação a característica gabor3, e que a gabor13 é a principal característica que pode definir o padrão de cada grupo.

A terceira base de dados usada para o

**Figura 11** – Resultado das Coordenadas Paralelas da base de dados Brodatz.



**Figura 12.** Resultado da Projeção Multidimensional da base de dados Brodatz.



teste é um conjunto de características de imagens médicas, 157 amostras, 28 atributos e 7 grupos. O resultado das visões está na Figura 13. A ferramenta também pode exibir somente as amostras de um grupo, permitindo o usuário verificar a distribuição das coordenadas no plano, como apresentado na Figura 14. A coordenação entre as visões por identificador destaca a mesma amostra em ambas as visões como apresentado na Figura 15.

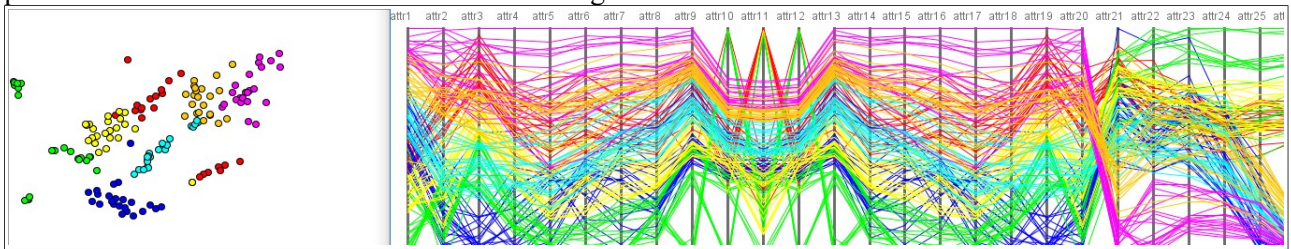
A quarta base de dados, também é um conjunto de características de imagens médicas com 28 atributos, e é apresentada na Figura 16. Essa base foi dividida em duas

partes, a primeira parte é apresentada na visão da Projeção Multidimensional (à direita na Figura 16) e possui 138 amostras e divididos em 6 grupos com 23 amostras, e a segunda parte é apresentada na visão das Coordenadas Paralelas (à esquerda na Figura 16) e possui 132 amostras e divididos em 6 grupos com 22

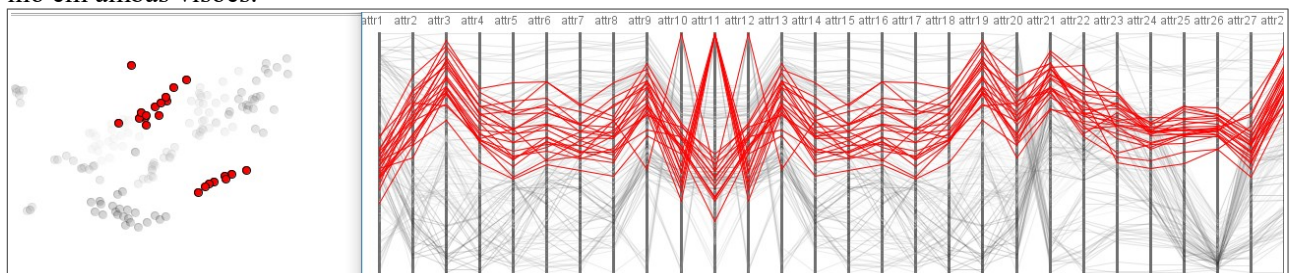
amostras. É visível na Projeção Multidimensional que a maioria dos grupos possuem amostras bastantes similares entre si, e amostras que estão fora do padrão das demais.

Para exemplificar a coordenação entre visões por similaridade, as Figuras 17 e 18,

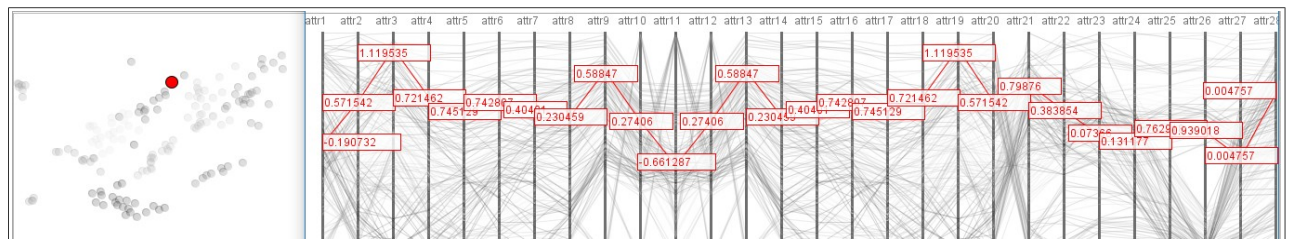
**Figura 13** – Resultado da Projeção Multidimensional (à direita) e das Coordenadas Paralelas (à esquerda) da primeira base de dados com características de imagens médica.



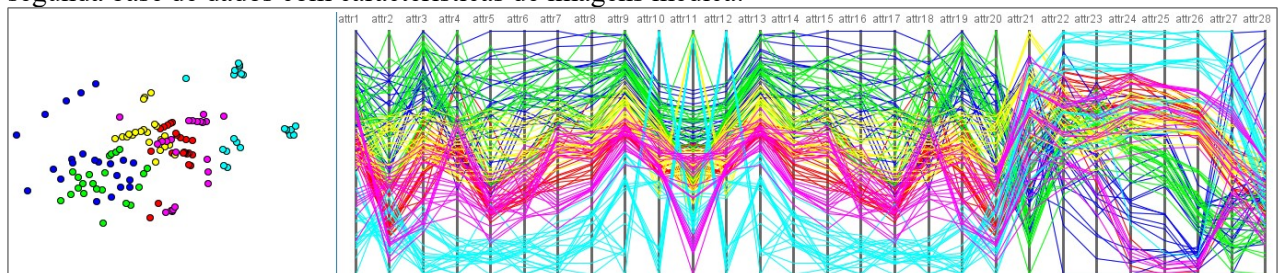
**Figura 14** – Resultado da Projeção Multidimensional (à direita) e das Coordenadas Paralelas (à esquerda) da primeira base de dados com características de imagens médica, destacando amostras do grupo em vermelho em ambas visões.



**Figura 15** – Resultado da Projeção Multidimensional (à direita) e das Coordenadas Paralelas (à esquerda) da primeira base de dados com características de imagens médicas, destacando a mesma amostra em vermelho em ambas visões.

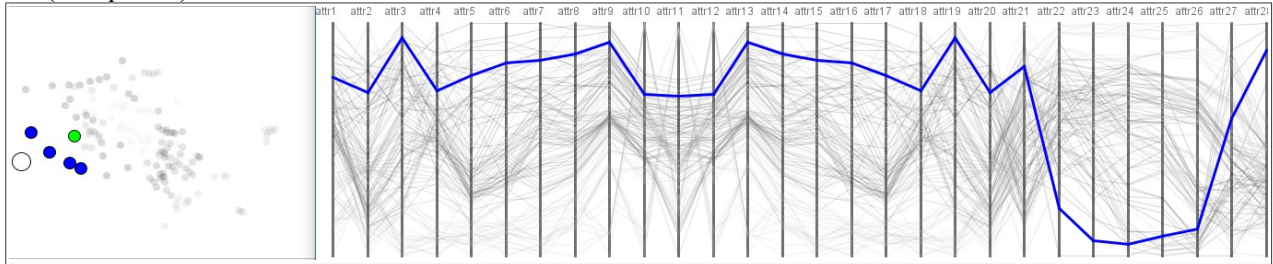


**Figura 16** – Resultado da Projeção Multidimensional (à direita) e das Coordenadas Paralela (à esquerda) da segunda base de dados com características de imagens médica.

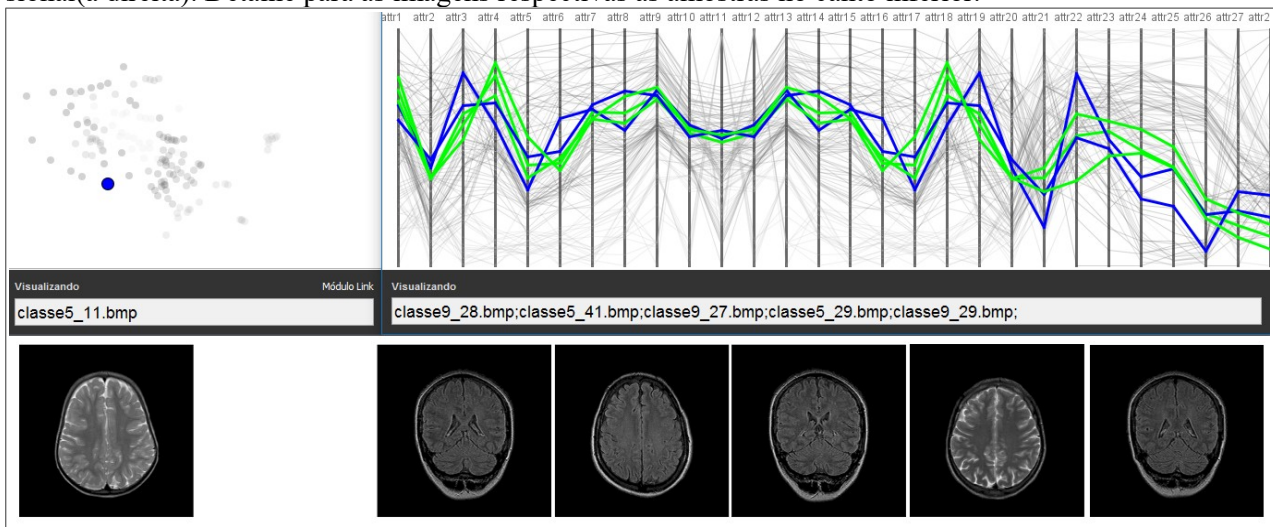




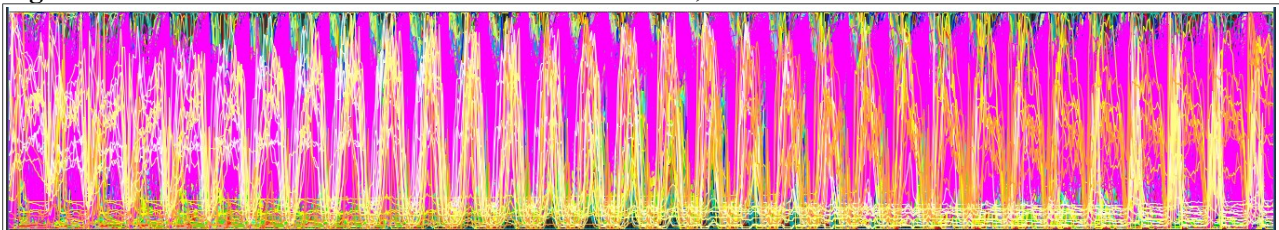
**Figura 17** – Resultado da Projeção Multidimensional (à direita) e das Coordenadas Paralelas (à esquerda) da segunda base de dados com características de imagens médica, destacando as cinco amostras mais similares na Projeção Multidimensional (à direita) selecionando uma amostra na visão das Coordenadas Paralelas (à esquerda).



**Figura 18** – Resultado da Projeção Multidimensional (à direita) e das Coordenadas Paralelas (à esquerda) da segunda base de dados com características de imagens médica, destacando as cinco amostras mais similares na visão das Coordenadas Paralelas (à esquerda), selecionando uma amostra na Projeção Multidimensional (à direita). Detalhe para as imagens respectivas as amostras no canto inferior.



**Figura 19** – Resultado da visão das Coordenadas Paralelas, de uma base de dado com 1024 atributos.



apresentam a mesma base dados de imagens médicas da Figura 16, com os destaques realizados pela ferramenta. Na Figura 17 na visão das Coordenadas Paralelas (à direita), ao selecionar uma amostra por meio do cursor, as cinco amostras mais similares são destacadas na visão da Projeção Multidimensional (à esquerda).

Já na Figura 18, o mesmo processo da Figura 17 é representado, porém invertendo a função da visão que seleciona amostra com a função da visão que destaca as amostras mais similares. Detalhe para essa Figura está na descrição em baixo da projeção, que informa o nome das amostras destacadas (a ferramenta disponibiliza essa informação para qualquer projeção). Em particular ao selecionar a

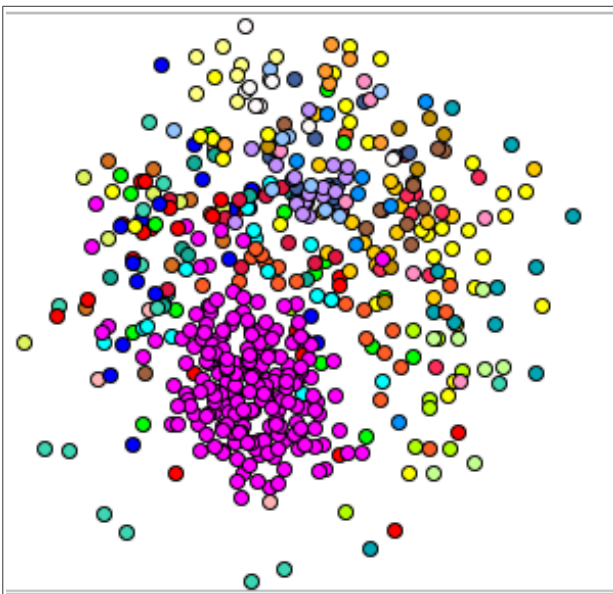
amostra ou imagem classe5\_11.bmp, as cinco amostras mais similares (de acordo com os atributos extraídos das imagens) são destacados. As imagens, respectiva as amostras destacadas, são apresentados na parte inferior da Figura.

## 6. LIMITAÇÕES

As explorações com as bases de dados apresentadas na seção anterior não apresentaram problemas com o tempo de execução e facilidade na exploração.

A quinta base de dados é mais complexa que as anteriores e possui 537 amostras, 35 grupos e 1024 atributos. O resultado é apresentado nas Figuras 19 e 20. A ferramenta apresenta uma demora para a geração da imagem das Coordenadas Paralelas, devido a grande quantidade de atributos. Além disso, muitas amostras ficam sobrepostas, dificultando a seleção de uma

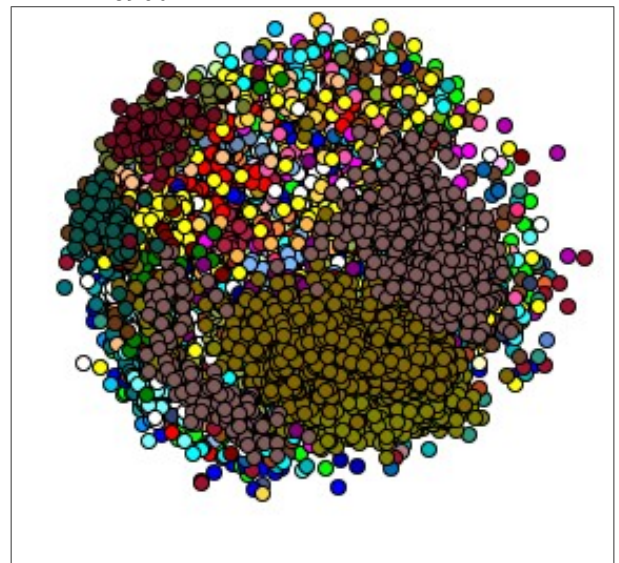
**Figura 20** – Resultado da Projeção Multidimensional com 1024 atributos.



amostra específica. Portanto a técnica de Coordenadas Paralelas, implementada neste trabalho torna-se ineficaz quando a base de dados possui uma quantidade grande de atributos. Já a Projeção Multidimensional não é afetada por este problema, logo é possível utilizar esta visão para selecionar a amostra desejada, exemplificando a ideia das visões coordenadas, onde uma visão pode complementar outra.

Porém a Projeção Multidimensional também torna-se ineficaz, com várias amostras sobrepostas, quando a base de dados possui uma quantidade grande de amostras como apresentado na Figura 21. Essa base de dados possui 8960 amostras e 112 grupos. A grande quantidade de grupos de amostra também dificulta a visualização, pois é difícil definir 112 cores distintas o suficiente para representar os grupos.

**Figura 21** - Resultado da Projeção Multidimensional com 8960 amostras.



Para mitigar o problema de grande quantidades de amostra, pode se alterar algoritmo de redução de dimensionalidade

para distribuir melhor as amostras, tentar reduzir a quantidade de amostras, aumentar a espaço de projeção ou utilizar uma terceira técnica de visualização de dados e complementar com uma coordenação entre as visões.

## **7. CONCLUSÕES E TRABALHOS FUTUROS**

No início deste trabalho, foi estudado o conceito das técnicas de visualização de dados e como suas representações visuais podem facilitar a leitura de uma base de dados. Também foi considerado que cada técnica possui pontos fracos e fortes, e que com a coordenação de uma ou mais visões, esse problema poderia ser mitigado. O conceito de coordenação por chave primaria e por similaridade foi introduzido como uma técnica coordenação.

A partir disso foi desenvolvido uma ferramenta que exemplifique a coordenação entre duas visões, em particular, Coordenadas Paralelas e Projeção Multidimensional. A base de dados poderia ser importada por arquivo ou Banco de Dados em MySQL.

Durante a fase de testes, base de dados simples (com poucos atributos e amostras) foram utilizadas para validar as projeções das técnicas de visualização. Com intuito de verificar a eficiência das técnicas de visualização, base de dados maiores foram utilizadas.

Quanto maior a quantidade de atributos, as Coordenadas Paralelas tornaram-se ineficaz ao selecionar uma amostra específica, porém a Projeção Multidimensional, por meio da coordenação, era possível selecionar a amostra e destacá-la no meio de várias outras amostras Coordenadas Paralelas.

A coordenação por similaridade também permitiu visualizar as amostras similares com facilidade.

Os testes também permitiram visualizar as limitações da ferramenta em projetar as técnicas de visualização para base de dados com grande quantidade de amostras.

O problema de tempo de projeção das Coordenadas Paralelas, pode ser solucionado alterando a estrutura de dados pré-alocada para a aplicação da técnica. Implementar a própria técnica de redução de dimensionalidade pode reduzir o tempo da geração das coordenadas para Projeção Multidimensional.

Para mitigar o problema da eficiência em extrair informações de grandes base de dados, novas técnicas de visualização podem ser utilizadas e coordenadas com as projeções já existente. Adicionar opções para melhorar interação do usuário com a base de dados, como ampliação, ocultar algum grupo de amostras e pesquisar uma amostra pelo nome, pode aumentar a eficiência das técnicas.

## 8. REFERÊNCIAS

MANJUNATH, B.S., MA, W.Y.: Texture features for browsing and retrieval of image data. **IEEE Transactions on Pattern Analysis and Machine Intelligence** <  
<http://dx.doi.org/10.1109/34.531803> > 18(8), 837–842 (1996). Acesso xx março 2017.



