

# E-R Modeling and Visualization of Large Mutual Fund Data

Xiong, F.\*<sup>1</sup>, Prakash, E. C.\*<sup>1</sup> and Ho, K. W.\*<sup>2</sup>

\*1 Center for Graphics and Imaging Technology, Division of Software System, School of Computer Engineering, Nanyang Technological University, Singapore-639798.

\*2 Center for Financial Engineering, Division of Banking and Finance, Nanyang Business School, Nanyang Technological University, Singapore-639798.

Received 3 December 2001.

Revised 12 February 2002.

**Abstract:** Existing visualization techniques used by mutual fund managers focus on the active portfolio management strategies. However, an important aspect of mutual funds is to visualize and understand the fund composition observed at specific points in time. Such a visualization will leverage the fund manager with the power to investigate and to capture forecasting differences and future performance. Our research here is to design a new visualization system that has an Entity-Relationship (E-R) model at the heart of the system and reveals not only the aggregate behavior of the mutual funds, but also the behavior of the mutual fund managers who manage these funds. We find some preliminary evidence that mutual funds in the world market represent a strong geographical pattern, and some potential niche markets are now extracted that stands out in the global scenario such as the returns from funds that belong to a specific country. For the analysis of security distribution, it helps to distinguish the most popular securities from other ordinary ones for an increased realization of profit. Also based on our experience with such huge world fund data to achieve a trade-off between time efficiency and graphics quality, we recommend the approach of splatting visualization, which indicates the influence of a particular security in a fund's portfolio.

**Keywords:** mutual fund, financial visualization, entity-relationship modeling, 3D interactive graphics.

## 1. Introduction

Advanced techniques in both database modeling and visualization have now made it possible for financial manager to understand and interpret data efficiently. To develop an appropriate framework for large data and complex analytical models in financial engineering, visualization designers now face two challenges. The first is financial complexity, which arises from the large data as well as the complex financial models that operate on data. The second is visual complexity, which arises from the lack of structure in financial data as well as from the poor choice of visual presentation.

A mutual fund pools money from hundreds and thousands of investors to construct a portfolio of stocks, bonds, real estates, or other securities, according to its charter (Money 101, Investing in Mutual Fund). Each investor in the fund gets a slice of the total pie. There are more than 16,000 kinds of mutual funds around the world, and their spectrum has come across nearly every main security market, from New York to London, Frankfurt to Tokyo.

A not strict division of mutual funds leads to three main categories: equity funds (stock funds), bond funds and money-market funds. However, there are so many subdivisions below a major class of fund, and more details

exist for every particular division. For instance, there are funds that invest for 'maximum growth', for 'growth' and for 'income'. They will try to make you money by investing in the stock market, but their approaches can be very different. There is a high diversity and variety within mutual fund portfolios. Usually, we pay more attention to the class of equity for its higher earning power.

For equity funds, the financial complexity arises from the diversification of assets allocation. Many of the companies most widely held by equity funds are household names - General Electric, Philip Morris, General Motors, IBM, AT&T, Ford Motor and Chrysler (Laderman, 1995). However, General Electric, the stock most commonly found in investors' portfolios, amounts to only slightly less than six-tenths of 1 percent of the net assets of all equity fund holdings. That is a sign of the diversity in mutual fund construction. Although it proved that mutual fund is safer to dilute risk than any individual stock, there exists a dilemma for the effective management of its highly volatile compositions. One biggest cost in fund business is the uncertainty. The risk increases when part of the portfolio changes responding to a market fluctuation.

Furthermore, the insights and experiences of the financial analysts cannot be totally bulletproof to protect investors from disaster strike. Actually, there is lack of efficient method to observe and analyze the multi-dimensional information generated in every transaction. Without a set of 'clever' risk-management tools, it is easy to get confused facing the mountain-like documents or lose clues dealing with the complicated data files.

Financial visualization focuses on the use of visualization techniques to help people understand financial data. In a number of cases, visualization has been used to analyze and display large volumes of market data in such a way as to extract significant features and results quickly and easily. These techniques include using 'point of view' and 'embodied virtual space' (Strausfeld, 1995) to present mutual fund portfolios, 'tree maps' (Jungmeister and Turo, 1992) and 'market maps' (mutual fund map) to show the hierarchical structure of security market, employing motion and animation for trading analysis and risk viewing in capital markets (Wright, 1995), and using 'world within world' (Feiner and Beshers, 1990) metaphors for exploring  $n$ -dimensional virtual world of financial instruments.

In this paper, we attempt to introduce a possible solution to analyze the mutual fund market based on entity relationship modeling and visualization. Work has been done to understand the current fund market and acquire the basic trading/management strategies. Work also has been done to explore and apply novel visualization techniques to facilitate our financial investigation of mutual fund market.

## 2. E-R Modeling in Visualization System Design

### 2.1 Entity-Relationship Modeling

Since its introduction by Chen (1977), the entity-relationship model (ERM) has become widely accepted in the area of database design and software engineering. In the ERM, as usual, we represent entities as fixed-sized rectangles, relationships as fixed-sized diamond-shaped drawings, and attributes as fixed-sized ovals. The connections between the entities, relationships and attributes are represented as continuous lines. The numeric notion associated with a line indicates that it is a connection from 1 to many, many to 1, or many to many. These simple and clear structuring concepts allow users to naturally model the affairs in real world. With the further contributions of other authors (Teorey, 1990), the ERM has gone beyond the original model to help database designers capture more complex semantics from conceptual schema to logical schema.

### 2.2 Proposed Solution

It is well known that visualization can serve as a memory generation of pictorial representation of relationships. However, there should be a bridge to link the cognition of end users with that of visualization designers. It will not be able to give the expected benefits unless discrete materials are arranged in a capable format. Thus we refer to E-R modeling as the logical design and interpretation tool for relational database visualization. The simplicity and readability of E-R approaches makes it even more convenient for end users to master all the required techniques, targeting at the 'pith' of visualization. Armed with the 3D interactive graphics, we present the information contexts and their characteristics by a multiple 'point of view'. Three techniques are included, differentiated by rendering process as 3D scatter plots, map visualization and splatting visualization .

### 3. Mutual Fund Visualization Techniques

Visualization plays an important role in the investigation of mutual funds. Some of the relevant techniques include:

- **3D Scatter Plots:** In scatter plots, 3D coordinates  $(x, y, z)$  represent selected data attributes. The direct mapping represents the original data distribution and show how one parameter is affected by another. By this means, it is clear to see the clustering of data. The size, shape and color of points help to distinguish different data properties. Users can rotate the graph to observe the details from all directions.

In our approach of mutual fund visualization, the  $X$ ,  $Y$  and  $Z$  axes are usually mapped to the primitive elements of funds and their enclosed securities. Color mapping is applied to highlight the scalar data (key values), based on an user-defined data segmentation.

- **Map Visualization:** When the source data contains geographical components, map visualization will be typically used. It represents the data over a three-dimensional landscape, using bar charts or polygonal surface, to create a spatially-related visualization. By a zoom in/zoom out operation, the analysts will be able to navigate in the 3D cyberworld.

In our applications, the regions on a landscape are associated with the countries where mutual funds or their securities published. Here each graphical object has recognizable shape and position. Its height and color can be mapped to statistic values such as number of funds in a country, average fund value in a country, and total value of securities in a country.

- **Splatting Visualization:** Splatting visualization is a subbranch of the field well known as volume visualization. Becker (1997) has proposed an approach on volume rendering for relational data. We prefer to test volume visualization's performance for financial instruments, and provide a complete view of data clustering or congregation. In our project, we employ Gaussian splatting. The sampling is performed by creating a special influence, or a splatting function  $SF(x, y, z)$  that will distribute the data value of each point over the surrounding region. Each splat is a fuzzy ellipse, while the shape of the ellipse can be controlled by vector or normal data and the size of the splat can be controlled by scalar magnitude.

A general form of  $SF(x, y, z)$  is the uniform Gaussian distribution shown below:

$$SF(X, Y, Z) = S \cdot e^{-f(r/R)^2}$$

In this equation,  $S$  is a scale factor that multiplies the exponential,  $f$  is the exponent scale factor ( $f \geq 0$ ),  $r$  is the distance between any point and the Gaussian center point and  $R$  is the radius of influence of the Gaussian, where  $r \leq R$ .

### 4. Implementation Steps

In this section, we will outline the major steps of designing logical architecture in visualization system. Then two examples are given to exhibit the mutual fund visualization using E-R modeling.

#### 4.1 Data Source

We have used the real mutual fund data from our collaborator 'Thomson Financial Securities Data'. In the following examples, we focus on mutual funds that invest in worldwide stock markets. The scope of our samples includes 16,495 funds (more than 21,000 securities) scattered in 107 countries. The net asset value (NAV) varies from a few cents to a few thousand of dollars, as well as the tremendous changes in current shares held. For the market value of every fund, interesting and informative patterns are discovered through different combinations of securities. These individual clues have been collected, amplified and captured in this paper for a reference. Our original mutual fund data comprise a description about 1) the basic information of mutual funds, 2) the basic information of countries, 3) the detailed information of securities and 4) the component of mutual fund and the annual change in shares held.

#### 4.2 Visualization Tool

To achieve a high quality graphics output, we use the SGI MineSet (SGI™ MineSet™) as the development tool. It provides graphical toolkit to support data mining and visualization. The integration of diverse visualization tools allows users to choose the most appropriate method for a given problem. The client-server architecture performs most of the compute intensive tasks on a powerful server while the processed results are returned to the client for visualization. It is convenient to visually explore multivariate relational data. So that we will take it in financial visualization, especially for mutual fund market.

#### 4.3 Major Steps in Architecture Design

The entity-relationship approach to the design of logical architecture consists of the following steps:

1. Identify entity types.
2. Identify relationship types.
3. Draw an E-R diagram with entity and relationship types.
4. Identify value types and attributes.
5. Translate the E-R diagram into a visualization pipeline.
6. Generate the target visualization.

#### 4.4 Case Study

##### 4.4.1 System overview

Based on the definition of E-R diagram, the flow chart at top level is given in Fig. 1. For any mutual fund in the world market, it belongs to only one country, or one sector. On the other hand, mutual funds are often categorized by the market capitalization of the stocks held in their portfolios. Some fund managers use a value approach to stocks, searching for stocks that are undervalued when compared to other similar companies. Thus market capitalization provides such a way to see the prospectus. Customers trade their preferable funds in an open market, and once the transaction ends, a contract of brokerage is automatically built between the company where the fund was issued and the investor. Professional fund managers are hired by those companies to engage their intelligence and insight in any investment. Various kinds of securities construct the pyramid of mutual fund, where each brick has its own brand of country or industrial sector. And that building is not static, even a gentle breeze from the financial market will shake the vulnerable equilibrium. Sometimes it leads to a reshuffle of assets by adding or removing some shares currently held.

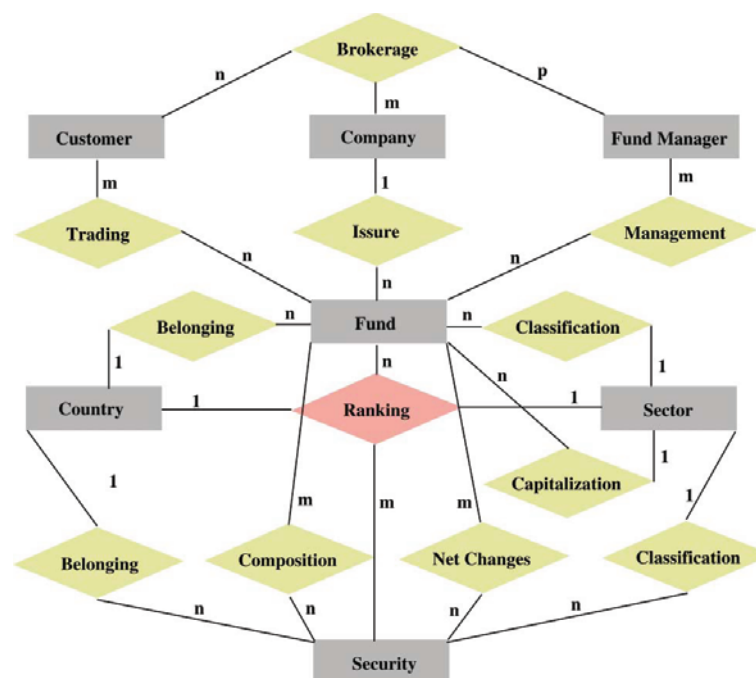


Fig. 1. The E-R diagram of world mutual fund market (top level).

#### 4.4.2 Case study I: Fund/security distribution in countries

The first example is the geographical distribution of funds or securities in the world. We know that every fund or security should belong to a country, but what we do not know is how large that country's fund/security market is and which country has the most profitable financial instruments. A classification is obviously helpful to solve this problem. While that is not enough, financial analysts may require more information on the regional distribution, especially from geographical interpretation.

To design the E-R diagram, let us see Fig. 1 again. Here we would like to discuss the relationship of 'Belonging' for both fund and security. It is a many-to-one mapping, and a detailed description is given in Fig. 2(a). We see that the primary entity types are fund, security, country, and identified relationship is distribution (belonging). The next step is to determine attributes and value types for the entity relationship. For fund, it has the following value types as number, name and management company. For security, its value types are number, name, price and shares outstanding (means at what extent a company has capitalized its property in the market). But there are still some values which are not obviously achievable. For instance, the price of one share of a fund is determined by the net asset value, or NAV, which is the total value of the securities the fund owns divided by the number of shares outstanding. In this case, we add an extra value type as statistic data to represent those calculated features. For the definition of attributes, it is analogous to that of value types and can be solely identified.

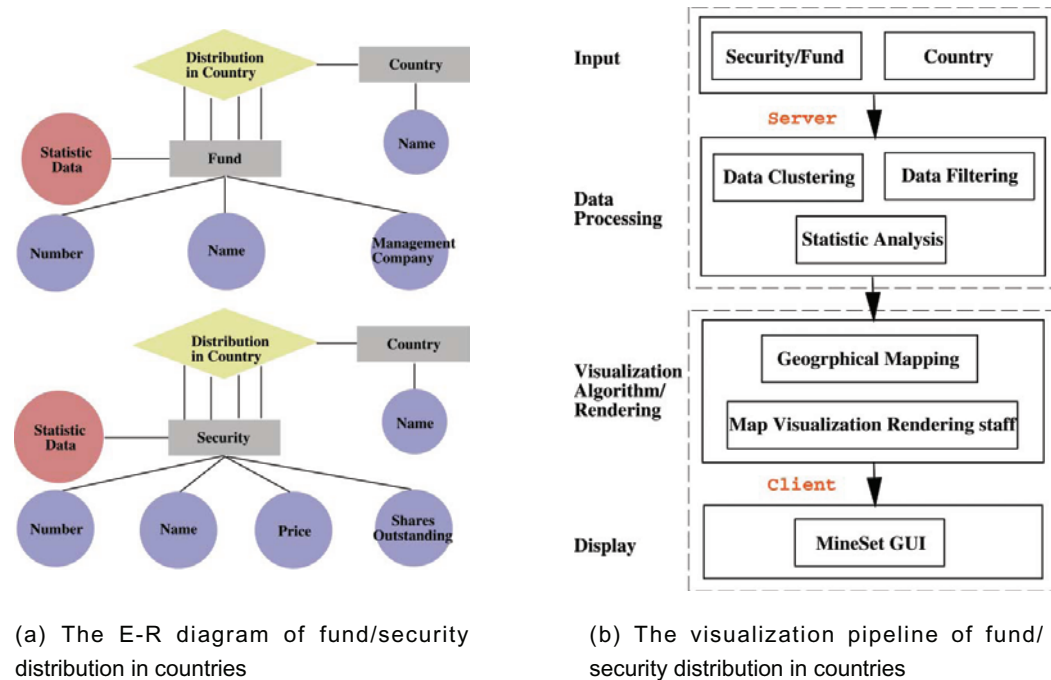


Fig. 2. Case study I: E-R modeling and visualization pipeline.

So far as we know, E-R modeling is only a logical architecture design. A translation from the enterprise view to a real visualization pipeline will be included. In Fig. 2(b), the input of visualization pipeline consists of both fund/security data and a country list. For Fig. 4, it is a three-dimensional landscape of European mutual fund market. Data items are associated with graphical bar-chart objects in the virtual map. However, the objects have recognizable spatial shapes and positions, such as those found in geographical maps. Figures 4(a) and 4(b) both represent the distribution of mutual fund in European countries. Here funds are classified by their country codes, then aggregated to get statistical information. For any country, there are three significant variables to look into: the number of funds, the market value of funds and the average market value of funds. In Fig. 4(a), two parameters (number and market value) are mapped to the height and color of graphical objects, and in Fig. 4(b), the meaning of color changes to be average market value. This tool displays quantitative and relational characteristics of the spatially oriented data.

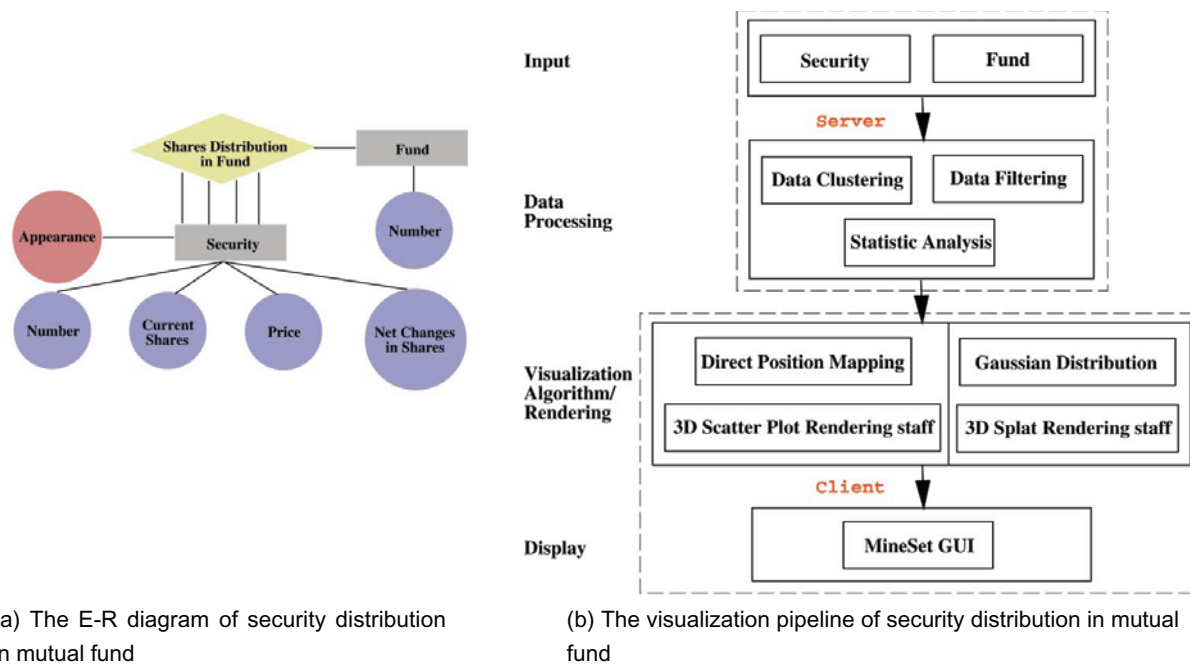
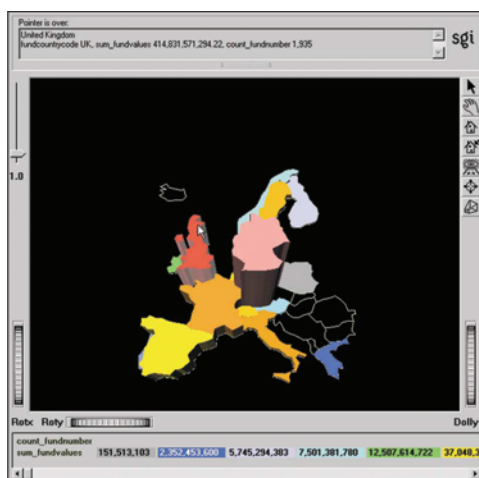
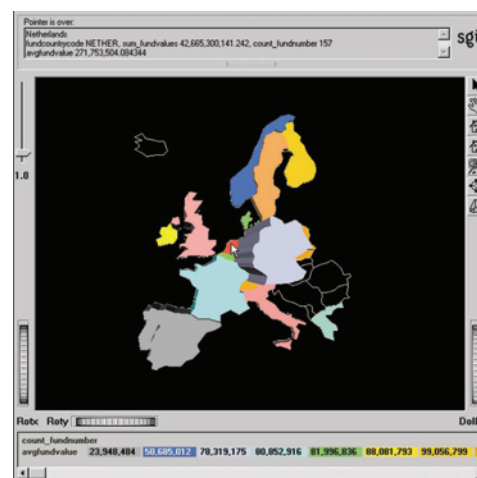


Fig. 3. Case study II: Security distribution in mutual fund.



(a) FDMV: Fund distribution of the market value (Europe)



(b) FDAMV: Fund distribution of the average market value (Europe)

Fig. 4. Case study I: Map visualization of fund/security distribution in countries.

#### 4.4.3 Case study II: Shares distribution in mutual fund

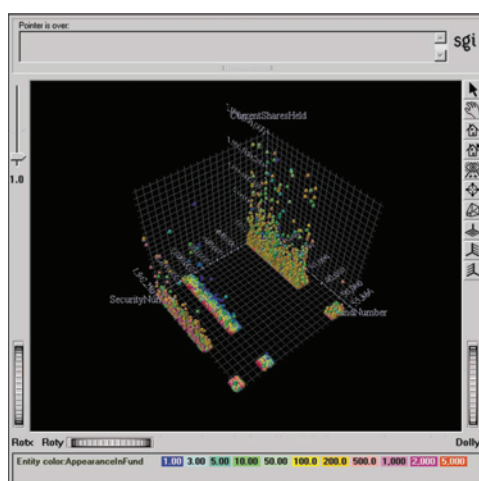
The second example is the distribution of shares in mutual fund. Investors need not worry about keeping the track of stock holdings - that is the fund manager's job. The selection of stocks in local or global region needs to be visualized. A 3D visualization of world fund market will provide us an intuitionistic view at the most popular securities.

Similar to the previous example, the design of E-R modeling identifies entity types, relationship types and value types/attributes. In Fig. 3(a), primary entity types are a) the fund, b) the security and the relationship type is distribution (composition). It is a many-to-many mapping, that is, each security can be involved in more than one fund, and each fund can involve more than one security. A security entity has the following value types: number, price, current shares and net changes in shares. The last two items mean that how many shares of that security are

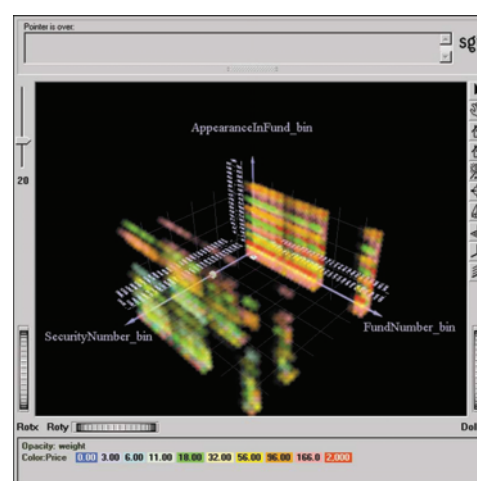
currently held in a fund, and how much the quantity has changed since the prior report. Another entity is fund, whose value types are number and name. Additionally, we would like to introduce a value type called 'appearance', which indicates how often one security will appear in the world fund market, leading to the 'hottest' choice of fund managers. The definition of attributes is the same with that of value types.

Using the translation rules discussed in 'case study I', the E-R diagram in Fig. 3(a) can be translated into the visualization pipeline shown in Fig. 3(b).

From Fig. 5(a), we can observe that the three axes correspond to fund number, security number and current shares held. Using color mapping, scalar data (appearance in fund) is distinct and the trend of variation is perceivable. Besides a bird's-eye view of the world market, analysts can select a few specified areas for intensive exploration. Nevertheless, the drawback of 3D scatter plots is: with the size of the dataset increasing, the visibility of individual points will be gradually diminished. An alternative to scatter plots visualization is the use of splatting. This method leads an effective discrimination between the high frequency and ordinary data by setting color and density. Figure 5(b) shows the security distribution in mutual fund. Three axes  $X$ ,  $Y$ ,  $Z$  stand for security number, fund number and appearance in fund; scalar data price is mapped to color.



(a) 3D scatter plots of security distribution in world fund market



(b) Splat visualization of security distribution in world fund market

Fig. 5. Case study II: Security distribution in mutual fund.

#### 4.5 Discussion and Findings

- Mutual Fund Relationships: The use of E-R modeling in visualization design is a probably new branch to set the architecture. It provides more insight than did traditional method which simply links data variables to graphics elements. From case study I and II, we learn that it is now the E-R diagram that tells designers what kind of relationship is achievable, so as to the construction of visualization pipeline.
- Mutual Fund Visualization: Additionally, the E-R modeling currently helps the designers to evaluate different visualization techniques.
- Mutual Fund Visualization: This includes map visualization, 3D scatter plots and splatting as explained below:
  1. European Market - Map Visualization: In Case I, the relationship refers to the distribution of security/fund in different countries, where geographical properties are linked with data properties. Map visualization proves to be appropriate in the representation. From Figs. 4(a) and (b), we can see that although Germany has the largest number of funds (= 4,095), the wealth of its fund market is less than that of the the United Kingdom, which obtains a portion of 35.3% in the European market. As to the average fund value, it seems that Netherland owns the richest mutual funds, followed by Italy and UK.
  2. Security Distribution - 3D Scatter Plots: Case II emphasizes on the relationship of security distribution in world fund market. Figure 5(a) represents the fund market at a high level of details, by mapping every fund to a spatial point.
  3. Security Distribution - Splatting Visualization: The stereograph illustrated in Fig. 5(b) reveals that there

is no direct link between a security's price and its liability to be held by mutual funds, but from a statistical point of view, those most 'favorable' securities will appear in the range from \$18 to \$166.

#### 4.6 Conclusion and Future Work

Financial visualization is not just the 'pretty graphics'. Far wider and deeper applications are studied currently in the area of common security, market index, option portfolio, as well as world mutual fund market. It is actually the case that cyberworlds (Kunii and Luciani, 1998) built as worlds of information need to be visually explored. The composition is achieved by the theoretical framework presented in this report. Nevertheless, financial analysts might be interested in a more practical implementation tool, which requires further improvements in both visualization techniques and model capacity.

#### References

- Becker, B. G., Volume Rendering for Relational Data, Proceedings of IEEE Symposium on Information Visualization, (1997), 87-90, 124.
- Chen, P., Data Base Management No.6: The Entity-Relationship Approach to Logical Data Base Design, (1977), Q.E.D. Information Sciences, Inc.
- Feiner, S. and Beshers, C., Worlds within Worlds: Metaphors for Exploring n-Dimensional Virtual Worlds, Proceedings of the ACM SIGGRAPH Symposium, (1990), 76-83.
- Investing in Mutual Fund: Top 10 Things to Know, <http://www.money.com/money/101/lessons/8/topten.html>.
- Jungmeister, W.-A. and Turo, D., Adapting Treemaps to Stock Portfolio Visualization, Research Technical Report CAR-TR-648, University of Maryland, Center for Automation, (1992).
- Kunii, T. L. and Luciani, A., Cyberworlds, (1998), Springer-Verlag, Tokyo.
- Laderman, M. J., Business Week: Guide to Mutual Funds, Fifth Annual Edition, (1995), McGraw-Hill, Inc.
- Mutual Fund Map, <http://www.smartmoney.com/maps>.
- SGI<sup>TM</sup> MineSet<sup>TM</sup>, <http://www.sgi.com/software/mineset.html>.
- Strausfeld, L., Financial Viewpoints: Using Point-of-View to Enable Understanding of Information, Proceedings of ACM CHI'95 Conference on Human Factors in Computing Systems, (1995), 208-209.
- Teorey, T. J., Database Modeling and Design: the Entity-Relationship Approach, (1990), Morgan Kaufmann Publishers, Inc.
- Wright, W., Research Report: Information Animation Applications in the Capital Markets, Proceedings of InfoVis'95, IEEE Symposium on Information Visualization, (New York), 19-25, color plates (1995), 136-137

#### Author Profile



Fei Xiong: He is currently a Master student of School of Computer Engineering, Nanyang Technological University. He received his BS in computer science and BE in Economics from Peking University, China, in 1998. His research interests are visualization of financial engineering, computational finance, 3D geometric modeling for very large data, high performance computing and portfolio management for security/mutual fund market.



Edmond Cyril Prakash: He is an Assistant Professor of Computer Science at the Nanyang Technological University, Singapore. He received his BE, ME and Ph.D degrees from Annamalai University, Anna University and Indian Institute of Science respectively. His research focus is on exploring the applications of virtual reality in engineering, science, medicine and finance. He also serves as a Deputy Director for the Financial Engineering program at NTU, Singapore.



Kim Wai Ho: He is an Associate Professor of Banking & Finance at the Nanyang Technological University, Singapore. He received his BS, MFin, Ph.D degrees from Imperial College, RMIT, and Nanyang Technological University respectively. His research focus is on corporate finance, valuation, investment, financial statement analysis and financial modeling. He serves as the Director of the Center for Financial Engineering.