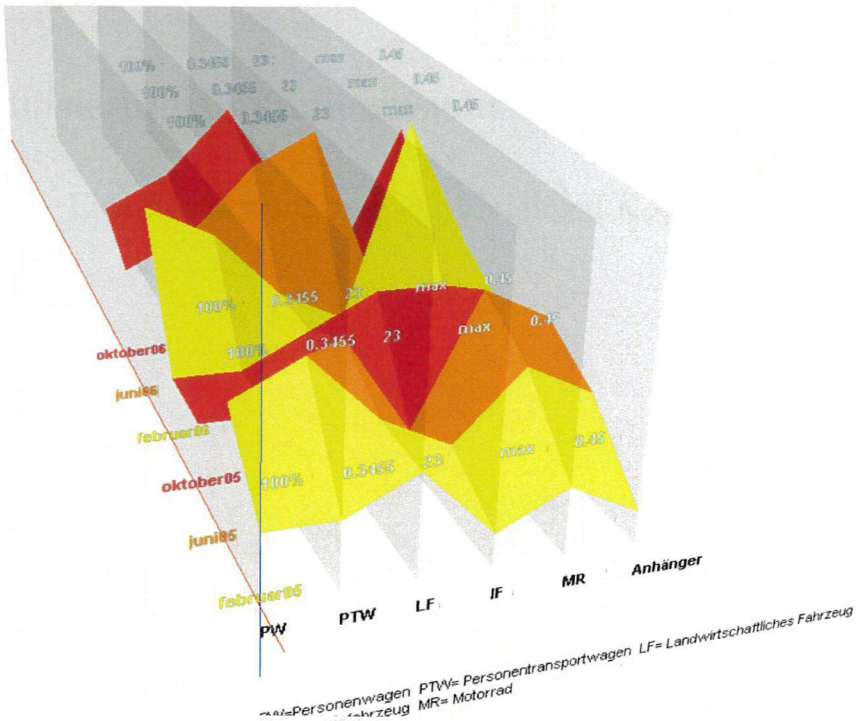
Geographisches Institut Universität Zürich
1742



Nur zum Gebrauch
im Lesesaal!

Visualisierung multidimensionaler Daten mit Zeitbezug

Eine 3D-Erweiterung der Parallelen Koordinaten

Masterarbeit

Stefanie Hautle

Universität Zürich-Irchel
Geographisches Institut
-BIBLIOTHEK-
Winterthurerstrasse 190
CH-8057 Zürich

Geographisches Institut Universität Zürich

Masterarbeit

Visualisierung multidimensionaler Daten mit Zeitbezug

943
025c

Eine 3D-Erweiterung der Parallelen Koordinaten

Stefanie Hautle

Abteilung Geographische Informationsvisualisierung und Analyse GIVA
Geographisches Institut der Universität Zürich

Betreuung:	Dr. Tumasch Reichenbacher
Fakultätsvertretung:	Prof. Dr. Sara Fabrikant

1917643

Inhalt

1 Einleitung 1
1.1 Motivation 1
1.2 Zielsetzungen und Forschungsfragen 3
1.3 Vorgehen 4
1.4 Aufbau der Arbeit 5

2 Theoretische Grundlagen und Stand der Forschung 7
2.1 Daten 7
2.1.1 Datentyp 7
2.1.2 Dimensionalität 8
2.2 Datenvisualisierung 9
2.2.1 Definitionen und Begriffserklärung 9
2.2.2 Anforderungen an eine Visualisierung 11
2.2.3 Gestaltungselemente einer Visualisierung 12
2.3 Visualisierung von Multiparameterdaten 13
2.4 Anforderungen an 3D-Visualisierungen 13
2.5 Parallele Koordinaten 14
2.5.1 Definition der Parallelen Koordinaten 15
2.5.2 Vor- und Nachteile 16
2.5.3 Erweiterte Parallele Koordinaten 18

3 Grundlagen und Herleitung 22
3.1 Die Visualisierung der Zeitabhängigkeit 22
3.1.1 zeitliche Primitiven und Strukturen 23
3.1.2 Besondere Bedeutung des Parameters Zeit 24
3.1.3 Ansätze zur Visualisierung der Zeitabhängigkeit in 3D 24
3.2 Konzept der zeitachsenbasierten 3D Parallelen Koordinaten 27
3.2.1 theoretische Grundlage der Idee 27
3.2.2 geometrische Grundlagen 28
3.2.3 Interaktion 30
3.2.4 Farbwahl 31
3.2.5 Navigation 33

4 Software und Programmiersprache 34
4.1 Java 3D 34
4.1.1 Entwicklungshintergrund 34
4.1.2 Java3D und seine Komponenten 35
4.1.3 Projektionsfläche 36
4.1.4 Das SzenenGraph-Modell 36
4.2 Aufbau und Implementierung des Programms 38

5 Visualisierungsdesign des Prototyps 39
5.1 Kriterien und Rahmenbedingungen 39
5.2 Interaktion im Prototyp 40
5.3 Navigation 41
5.3.1 Übersicht 42
5.3.2 Zoom 42
5.3.3 Auswählen und markieren 42
5.3.4 Translation und Rotation 42
5.4 Designentscheidungen 43
5.4.1 Merkmalsflächen 43
5.4.2 Koordinatenachsen 44
5.4.3 Text 45
5.4.4 Hintergrund 45

6 Visualisierungsbeispiele 46
6.1 Benutzerinterface 46
6.2 Beispiele aus den Prototypen 48
6.2.1 Beispiel 1 48
6.2.2 Beispiel 2 49
6.2.3 Beispiel 3 50
6.2.4 Beispiel 4 52

7 Diskussion 54
7.1. Diskussion der erstellten Beispiele 54
7.2 Die Beantwortung der Fragestellungen 58

8 Schlussfolgerung und Fazit 61
8.1 Erkenntnisse 61
8.2 Verbesserungsmöglichkeiten und Limiten 61
8.3 Offene Fragen 62
8.4 Ausblick 62

Literaturverzeichnis und Onlinequellen 63

Abbildungsverzeichnis

Abbildung 1: Das Vorgehen, zusammengefasst in acht Schritten (Quelle: eigene Darstellung)5
Abbildung 2: Visualisierungspipeline multivariater Daten (Quelle: Schumann u. Müller, 2000: 174)..... 10
Abbildung 3: visuelle Variablen nach Bertin (1974), (Quelle: eigene Darstellung nach Universität Münster, 2007)..... 12
Abbildung 4: Das Grundkonzept der Parallelkoordinaten: parallele Achsen an Stelle von orthogonalen Achsen. Die vier Punkte (links) werde durch die vier Linien (rechts) dargestellt. (Quelle: Siirtola et al.,2009)..... 15
Abbildung 5: Beispiel eines Parallelen Koordinaten Diagramms mit 6 Variablen (Quelle: Technische Universität Wien, 2006)..... 16
Abbildung 6: optisches Wirrwar bei vielen Merkmalslinien (Quelle: Chi-Square, 2001)..... 17
Abbildung 7: Die ESTAT Applikation fügt den Scatterplot (oben links), Karte (unten links), Zeitserien Graph (oben rechts) und das Parallele Koordinaten Diagramm (unten rechts) zusammen. (Quelle: Robinson, 2007). 18
Abbildung 8: Parallele Koordinaten mit Spinogramm (Quelle: Theus,2003)..... 18
Abbildung 9: Parallele Glyphen (Quelle: Fanea et al., 2005) 19
Abbildung 10: Die Technik der "clustered multirelational Parallel Coordinates", (Quelle: Johansson, 2005) 19
Abbildung 11: Parallele Koordinaten mit zusätzlicher Dichteinformation (Quelle: Streit, 2006)..... 20
Abbildung 12: Vom 2D Scatterplot (links) zu dreidimensionalen Parallelen Koordinaten (rechts). (Quelle: Rübel et al., 2008) 20
Abbildung 13: erweiterte PCP(Quelle: Technische Universität Wien, 1999)..... 21
Abbildung 14: (a) lineare Zeitreihe, (b) zyklische Zeitreihe (c) verzweigte Zeitreihe (Quelle: Aigner et al., 2007)..... 23
Abbildung 15: links: 3D Zeitrad (Aigner et al., 2008)..... 25
Abbildung 16: spiralförmige Zeitdarstellung für die Konsumation von12 Lebensmitteln während den Monaten zwischen 1980-1989, rechts: spiralförmige Darstellung prozentualer Anteile von 112 Lebensmitteln von 1980-1989 (Quelle: Carlis u. Konstan, 1998)..... 25
Abbildung 17: Temporal Star Darstellung für Börsenkurse über 8 Wochen (Quelle: Noirhomme-Fraiture, 2002)..... 26
Abbildung 18: Kiviat Tube, (Quelle: Hackstadt u. Malony, 1995)..... 26
Abbildung 19: Beispiel eines Extruded Parallel Coordinate Plot für einen fünfdimensionalen Datensatz (Quelle: Wegenkittl et al., 1997) 27
Abbildung 20: erweiterte Parallele Koordinaten für ein vierdimensionales, dynamsiches System (Quelle: Wegenkittl et al., 1997) 28
Abbildung 21: Merkmalslinien mit der Variable x zu fünf verschiedenen Zeitpunkten (Quelle: Wegenkittl et al., 1997) 28
Abbildung 22: aufgezogene Fläche (Quelle: eigene Darstellung)..... 29
Abbildung 23: dieselben Merkmalslinien wie in Abb. 21 mit dreidimensionalen Parallelen Koordinaten dargestellt (Quelle: Wegenkittl et al., 1997)..... 29
Abbildung 24: Konzept der erweiterten Parallelen Koordinaten für zeitabhängige Daten, (Quelle: eigene Darstellung)..... 30
Abbildung 25: Farbskala mit zwölf Abstufungen (Quelle: eigene Darstellung)..... 32
Abbildung 26: Aufbau der Java Plattform (Quelle: Valentin, 2005) 35

Abbildung 27: Schema der Beziehungen zwischen Betrachter , Projektionsfläche und Szenenkamera in Java3D (Quelle: Heinzel, 2003) 36
Abbildung 28: Aufbau Szenegraph (Quelle: Java3D_schnell_tutorial.pdf) 37
Abbildung 29: Ablauf der Implementierung (Quelle: eigene Darstellung) 38
Abbildung 30: Die Auswahlentscheidung für die Entwicklung einer problemspezifischen Visualisierung (Quelle: eigene Darstellung frei nach Lange et al., 2006) 40
Abbildung 31: Navigation control loop (eigene Darstellung nach Ware, 2004) 41
Abbildung 32: Parameterform der Ebenengleichung (Quelle: Gut, 2009) 43
Abbildung 33: Benutzeroberfläche der Applikation 46
Abbildung 34: Auswahl eines spezifischen Merkmals über die Pfeilbox „Time Axis“ 47
Abbildung 35: Detailansicht GUI des Prototypen 47
Abbildung 36: Beispiel 1 ParallelCoordinatePlot3d 48
Abbildung 37: Beispiel 2 ParallelCoordinatePlot3d, mit Legende 49
Abbildung 38: Beispiel 2 ParallelCoordinatePlot3d, Ansicht von schräg oben, ohne Legende 49
Abbildung 39: Beispiel 3 ParallelCoordinatePlot3d, Ansicht von schräg vorne, mit Extrema und Hilfslinien 50
Abbildung 40: Beispiel 3 ParallelCoordinatePlot3d, Ansicht von schräg vorne 50
Abbildung 41: Beispiel 3 ParallelCoordinatePlote3d, Ansicht von oben, mit eingeblendeten Merkmalslinien 51
Abbildung 42: Beispiel 3 ParallelCoordinatePlot3d, Frontansicht, mit eingeblendeten Merkmalslinien 51
Abbildung 43: Beispiel 4 ParallelCoordinatePlot3d, zyklischer Datensatz 52
Abbildung 44: Beispiel 4 ParallelCoordinatePlot3d, Seitenansicht, zeitlicher Verlauf und topologische Oberfläche 52
Abbildung 45: Durch eine bestimmte Anzahl von Kontrollpunkten oder Kontrollpolygonen kann eine komplexe Fläche beschrieben werden. (Quelle: CADauno, 2009) 55
Abbildung 46: keine Korrelation (lila), positive Korrelatin (blau), negative Koorelation (rot) (Quelle: eigene Darstellung nach Electronic Visualization Laboratory, 2001) 56
Abbildung 47: herkömmliche Darstellung der Variable Zeit (Variable: YEAR), (Quelle: Siirtola, 2000) 57
Abbildung 48: neue Darstellungsform mit der Variable Zeit auf der x-Achse 57

Tabellenverzeichnis

Tabelle 1: Einflussfaktoren zur Nutzung von Farbskalen (Quelle: eigene Darstellung nach Thomas, 2003) 31
Tabelle 2: Umsetzung der Anforderungen an einen dreidimensionale Visualisierung (Quelle: eigene Darstellung) 55

„The real voyage of discovery consists not in seeking new landscapes,
but in having new eyes."

[Marcel Proust, französischer Schriftsteller, 1871 - 1922]

Danksagung

Ich bedanke mich bei allen Personen, die mich bei der Umsetzung meiner Masterarbeit unterstützt haben. Besonders geht dieser Dank an: Dr. Tumasch Reichenbacher und Prof. Dr. Sara Fabrikant vom Geographischen Institut der Universität Zürich. Ihre Unterstützung, kompetente Betreuung und kritischen Hinweise halfen bei der Umsetzung der Arbeit. Besonders danke ich auch meinem Vater, meiner Schwester und Thomas Bächler, die mich während des ganzen Studiums stets unterstützt und begleitet haben.

Ich habe mich für das Thema dieser Arbeit entschieden, um einen vertieften Einblick in die geographische Informationsvisualisierung und die Datenanalyse zu erhalten. Durch die praktische Umsetzung und Implementierung des theoretischen Konzepts war es mir möglich, einen Einblick in die Programmierarbeit zu gewinnen und mich mit der Softwareerweiterung Java3D auseinander zu setzen.

Appenzell, den 25. Januar 2010 Stefanie Hautle

Zusammenfassung

Der Entwurf geeigneter visueller Darstellungstechniken für eine problembezogene und fachspezifische Informationsdarstellung ist Gegenstand aktueller Forschungsprojekte. Durch die rasante technische Entwicklung von Sensoren und Überwachungssystemen fallen immer mehr multidimensionale Datensätze, mit einer Vielzahl an abhängigen Parametern, an. Viele dieser Datensätze weisen einen Zeitbezug auf, dessen Visualisierung und Analyse besondere Darstellungsmethoden erfordert. Das Ziel solcher Visualisierungsmethoden ist es, möglichst viele der Parameter gleichzeitig darzustellen, so dass sie für den Betrachter zusammen erfassbar werden. Auf diese Weise können auch Zusammenhänge erkannt werden, die in Einzeldarstellungen der Parameter, verloren gehen können. Die Visualisierungsmethode der *Parallelen Koordinaten* eignet sich besonders gut für diese verlustfreie und gleichzeitige Darstellung mehrerer Attributwerte. Mit dieser Methode lassen sich n-dimensionale Datenräume in die zweidimensionale Ebene abbilden. Ein Nachteil der herkömmlichen Methode besteht allerdings darin, dass sich dem Betrachter eines Parallelen Koordinaten Diagramms, zeitliche Muster oder Verläufe nur schwer oder gar nicht erschliessen. In der vorliegenden Arbeit wird deshalb untersucht ob sich eine dreidimensionale Darstellung der Parallelen Koordinaten so konzipieren lässt, dass der Zeitparameter in multidimensionalen Datensätzen effizient und schnell erkannt werden kann. Dafür soll die von Wegenkittl et al. (1997) vorgeschlagene Methode der *Erweiterten Parallelen Koordinaten,* weiterentwickelt und ausgebaut werden. Ein problembezogenes, theoretisches Konzept widmet sich der Darstellungsmöglichkeit einer achsenbasierten, dreidimensionalen Visualisierungstechnik für zeitabhängige, multidimensionale Datensätze. Die praktische Umsetzung des Konzepts erfolgt mittels der objektorientierten, 3D-fähigen Softwareumgebung Java3D. Ein speziell für die Fragestellung entwickelter Prototyp generiert vier Visualisierungsbeispiele multidimensionaler Beispielsdatensätze. Durch ein interaktives und logisch strukturiertes Benutzerinterface werden die graphischen und theoretischen Anforderungen an eine dreidimensionale Visualisierung umgesetzt und für den Betrachter anwendbar. Die entworfenen Visualisierungsbeispiele sind aussagekräftig und ermöglichen einen neuen, vertieften Einblick in die zeitliche Struktur eines multidimensionalen Datensatzes. Periodizitäten und zeitliche Muster sind schnell, und gegenüber der herkömmlichen Methode der Parallelen Koordinaten, einfacher zu erkennen. Es werden neue Analysewerkzeuge vorgeschlagen, die sich in einer zukünftigen Weiterentwicklung untersuchen und einbauen lassen.

Schlagworte: Visualisierung von Zeitabhängigkeit, PCP, Parallele Koordinaten, multidimensionale Datensätze, Java3D

1 Einleitung

Dieses Kapitel erläutert die Problemstellung und liefert die Einführung in die Thematik der vorliegenden Arbeit. Zu Beginn werden die Bedeutung und der Stellenwert des Themas in einem grösseren Kontext aufgezeigt, während im zweiten Abschnitt das Ziel der Arbeit erläutert und die Forschungsfragen formuliert werden. Ebenfalls werden der Aufbau und das Vorgehen dieser Arbeit beschrieben.

1.1 Motivation

In den letzten Jahrzehnten werden in fast allen Wissenschaftsbereichen riesige Datenmengen gespeichert, die analysiert und verstanden werden müssen. Beim Aufzeichnen von Prozessen oder dem Überwachen von Vorgängen werden mittels komplexen Sensoren und Systemen, alle anfallenden Parameter gespeichert. Die Datensätze, welche dadurch gewonnen werden, sind multidimensional. Um das Informationspotenzial solcher Datensätze nutzen zu können, ist eine effiziente visuelle Datenanalyse nötig, die sich im Fall von multidimensionalen Datensätzen aber oft als unübersichtlich und anspruchsvoll erweist. Deshalb erscheint es besonders wichtig, eine geeignete Visualisierungstechnik für spezifische Fragestellungen zu finden. Nur eine zweckmässige Visualisierung kann dem Betrachter multidimensionaler Daten helfen, die Informationsgewinnung zu vereinfachen. Die Grundidee einer solchen Visualisierung liegt darin, die Daten möglichst einfach und intuitiv für das menschliche, visuelle System aufzubereiten. Durch eine geeignete graphische Darstellung der Daten kann dem Betrachter ein Einblick in die innere Struktur der Daten gewährt werden. Es wird ihm ermöglicht Schlussfolgerungen aus den Daten zu ziehen, sowie direkt mit den Daten zu interagieren (Keim, 2002). Die Datenvisualisierung ermöglicht dem Menschen ein tieferes Verständnis eines Datensatzes und erlaubt es ihm, neue Hypothesen und Annahmen zu formulieren. Sie hat deshalb eine wichtige Bedeutung in der Datenexploration, Datenanalyse und beim Abfragen von Informationen.

Die Thematik der Visualisierung multidimensionaler Datensätze stellt ein breites, aktuelles Forschungsfeld mit vielen Anwendungsbereichen dar. Zahlreiche Forschungsgruppen beschäftigen sich mit unterschiedlichen Fragen und Methoden der Visualisierung multidimensionaler Daten. Eine Übersicht über mehrdimensionale Visualisierungstechniken gibt Grinstein (2001). Siirtola (2007) liefert eine Zusammenfassung über interaktive, mehrdimensionale Visualisierungsmethoden. Multidimensionale Daten weisen oftmals einen Zeitbezug auf. Es existieren Forschungsarbeiten in unterschiedlichen Wissenschaftsbereichen, die sich neuen, leistungsfähigen Visualisierungsmethoden für multidimensionale, zeitbezogene Datensätze widmen. Eine gute Übersicht geben Aigner et al. (2007) oder Daassi et al. (2006). Mit einer interaktiven Visualisierung für periodische Zeitserien-Daten beschäftigen sich Carlis u. Konstan (1998). Tominski et al. (2005) schlägt eine 3D-

Visualisierung für zeitabhängige Daten auf Karten vor, Buono et al. (2005) widmen sich der interaktiven Mustererkennung in Zeitserien. Andrienko u. Andrienko (2005) untersuchen die räumliche Verteilung zeitlicher Verläufe, Guo et al. (2006) ermitteln Zeit-Raum-Muster in multidimensionalen Daten und Hochheiser und Shneiderman (2004) beschreiben die „Timebox" als Möglichkeit für interaktive Abfragen zeitlicher Muster.

Eine wichtige und oft verwendete Visualisierungsmethode, welche sich für die Darstellung multidimensionaler Daten eignet, sind die „**Parallelen Koordinaten**". Das grundlegende Konzept der Parallelen Koordinaten Diagramme (engl: Parallel Coordinate Plot), welches erstmals in Inselberg (1985) vorgestellt wurde, besteht darin, multidimensionale Datenpunkte in einer zweidimensionalen Graphik mittels einer sequentiellen Abfolge von parallelen Linien zu visualisieren. Diese Methode hat den Vorteil, dass sie alle Variablen gleichzeitig darstellen kann (vgl. Inselberg, 1990 und Inselberg, 2009). Zahlreiche weitere wissenschaftliche Publikationen beschäftigen sich mit verschiedenen Aspekten der Methode der Parallelen Koordinaten: Ellis und Dix (2006) beschreiben ihre Verwendbarkeit im Zusammenhang mit grossen Datensätzen, Edsall (2003) erläutert den Gebrauch in der geographischen Visualisierung, Moustafa und Wegman (2002) optimieren die Methode mathematisch, Unwin et al. (2003) widmen sich der Methode im Bezug auf die explorative Datenanalyse. Nur ganz wenige Forschungsgruppen haben sich bisher mit dem schwierigen Vergleich zeitlicher Verläufe befasst, der die Methode der Parallelen Koordinaten mit sich bringt. Die Problemstellung der vorliegenden Arbeit bildet sich aus diesem Nachteil der Methode. Basierend auf die Arbeiten von Fanea et al. (2005), Streit et al. (2006), Johansson et al. (2005) und Johansson et al. (2006), soll nun eine dreidimensionale Erweiterung entwickelt werden um die Zeitdimension multidimensionaler Daten visualisieren zu können. Mit der entwickelten Erweiterung sollen Fragestellungen bezüglich zeitlicher Dimensionen, Muster oder Zusammenhängen in einem Datensatz intuitiver und schneller beantworten werden können. Daraus ergibt sich das Ziel, welches im folgenden Abschnitt formuliert wird.

1.2 Zielsetzungen und Forschungsfragen

Das Ziel der vorliegenden Arbeit liegt darin, die Visualisierungsmethode der Parallelen Koordinaten so zu erweitern, dass zeitliche Verläufe in einem multidimensionalen Datensatz erkennbar werden und untersucht werden können. Die Analyse zeitlicher Strukturen und Muster erweist sich mit der herkömmlichen Methode der Parallelen Koordinaten Diagramme als nur schwer oder gar nicht durchführbar. Durch eine zusätzliche Dimension in der Darstellung kann der Zeitparameter in mehrdimensionalen Datensätzen besser, schneller und einfacher erfasst werden. Basierend auf dem theoretischen Konzept von Wegenkittl et al. (1997) und der Fragestellung dieser Arbeit, soll ein Prototyp entwickelt werden, der vier verschiedene Visualisierungsbeispiele generiert. Diese Beispiele werden aufgezeigt und diskutiert. Eine systematische Evaluation der entwickelten Visualisierungsbeispiele ist nicht Bestandteil dieser Arbeit.

Damit eine Erweiterung, wie sie im vorangegangenen Abschnitt beschrieben wurde, vorgenommen werden kann, muss zuerst die Frage nach der Besonderheit des Parameters Zeit im Allgemeinen gestellt werden.

> Forschungsfrage 1:
> *Ist „Zeit" eine „normale" Variable, die analog zu anderen Variablen behandelt werden kann?*

Die zweite Forschungsfrage widmet sich den bereits bekannten Visualisierungstechniken für zeitabhängige, multidimensionale Daten.

> Forschungsfrage 2:
> *Welche dreidimensionalen, achsenbasierte Visualisierungstechniken können für eine Darstellung zeitabhängiger, multidimensionaler Daten verwendet werden?*

Aufgrund der erarbeiteten theoretischen Grundlagen und Überlegungen soll die Methode der Parallelen Koordinaten schliesslich so erweitert werden, dass zeitliche Verläufe in einem multidimensionalen Datensatz visualisiert werden können. Mit einer geeigneten, 3D-fähigen Software wird ein Prototyp entwickelt.

> Forschungsfrage 3:
> *Wie kann die Visualisierungsmethode der Parallelen Koordinaten um eine „Zeitdimension" erweitert werden?*

Mit dem implementierten Prototyp sollen vier Visualisierungsbeispiele eines multidimensionalen Datensatzes dargestellt und diskutiert werden. Es soll untersucht werden inwiefern diese Erweiterung neue und bessere Analysemöglichkeiten in der visuellen Datenexploration bietet kann.

> Forschungsfrage 4:
> *Welche Vorteile und Möglichkeiten erschliessen sich dem Betrachter eines um die Zeitdimension erweiterten Parallelen Koordinaten Diagramms?*

1.3 Vorgehen

Um die herkömmliche Visualisierungsmethode der Parallelen Koordinaten so zu erweitern, dass die Zeitdimension der visualisierten Daten sichtbar wird, werden zuerst die theoretischen Grundlagen erarbeitet. Es soll erläutert werden inwiefern die Eigenschaften der zu visualisierenden Daten eine Rolle spielen, was Datenvisualisierung bedeutet und welche Anforderungen damit verknüpft sind. Es muss geklärt werden, was beim Visualisieren von Multiparameterdaten zu beachten ist und ob sich die Anforderungen an eine Visualisierung bei einer dreidimensionalen Darstellung ändern. Schliesslich wird die herkömmliche Methode der Parallelen Koordinaten vorgestellt und deren Vor- und Nachteile erläutert. Verschiedene Konzepte von bereits umgesetzten Erweiterungen konventioneller Parallelen Koordinaten Diagramme sollen kurz vorgestellt und erläutert werden.

Um die erste Forschungsfrage *Ist „Zeit" eine „normale" Variable, die analog zu anderen Variablen behandelt werden kann?* beantworten zu können, muss der Aufbau und die Struktur der „Zeit" aufgezeigt und diskutiert werden.

Zur Beantwortung der zweiten Fragestellung *Welche dreidimensionalen, achsenbasierte Visualisierungstechniken können für eine Darstellung zeitabhängiger, multidimensionaler Daten verwendet werden?* werden aus der aktuellen Literatur Visualisierungsmethoden, die sich für die Darstellung von zeitabhängigen Daten eignen, aufgezeigt und diskutiert.

Basierend auf diesen Ansätzen wird in einem weiteren Schritt ein Konzept entworfen, das die dreidimensionale Erweiterung der Methode hinsichtlich der Visualisierung von Daten mit Zeitbezug, ermöglicht. Das soll die Beantwortung der dritten Forschungsfrage *Wie kann die Visualisierungsmethode der Parallelen Koordinaten um eine „Zeitdimension" erweitert werden?* ermöglichen.

Es wird analysiert, welche geometrischen Konzepte geltend gemacht und welche Anforderungen an eine Visualisierung eingehalten werden müssen. Des Weiteren soll untersucht werden, welche Interaktionen für eine 3D- Darstellung dieser Art Sinn machen, und welche graphischen Gestaltungsmittel verwendet werden können. Mit den Erkenntnissen aus diesen Überlegungen wird ein Prototyp implementiert. Mit Hilfe der Softwareumgebung Java3D soll exemplarisch eine Realisierung des entwickelten Konzepts vorgenommen werden. Visualisierte Beispielbilder zeigen dann die Vorteile, die für die visuelle Analyse zeitabhängiger Daten mittels Paralleler Koordinaten entstehen, auf. Diese Visualisierungsbeispiele werden diskutiert, analysiert und auf neue Analysemöglichkeiten hin untersucht. Dieses Vorgehen führt schliesslich zur Beantwortung der vierten Forschungsfrage: *Welche Vorteile und Möglichkeiten erschliessen sich dem Betrachter eines um die Zeitdimension erweiterten Parallelen Koordinaten Diagramms?*

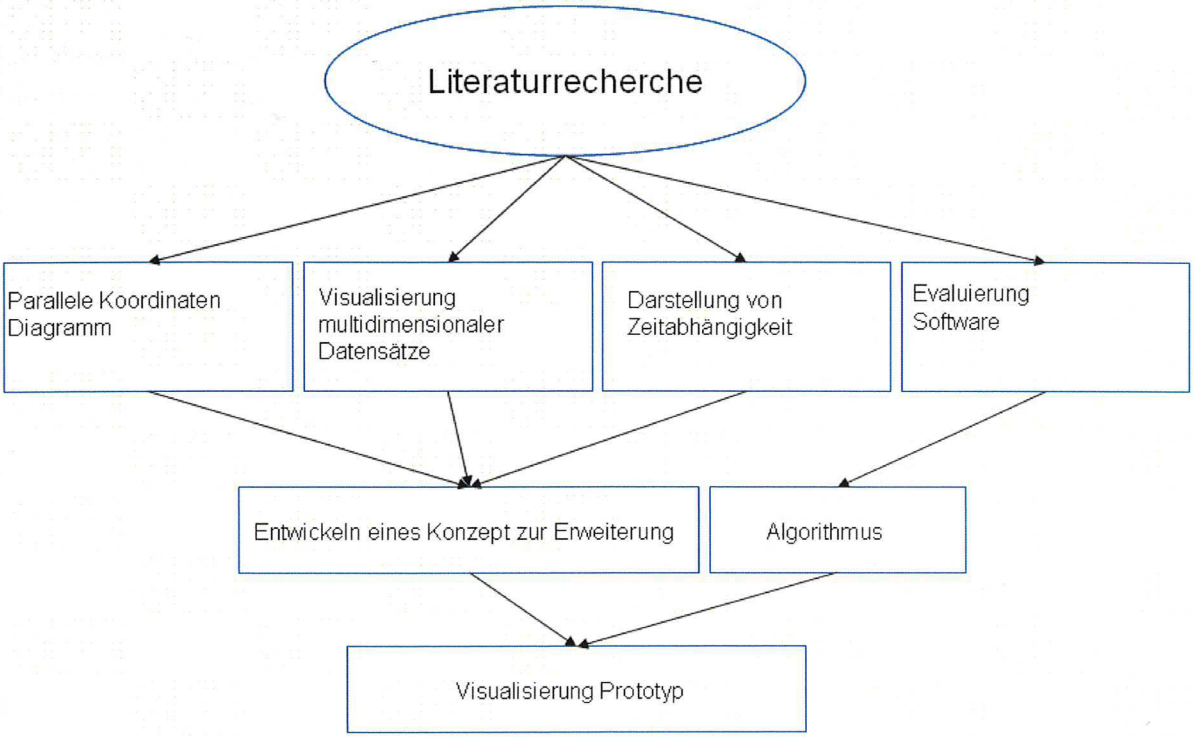


Abbildung 1: Das Vorgehen, zusammengefasst in acht Schritten (Quelle: eigene Darstellung)

1.4 Aufbau der Arbeit

Der Aufbau dieser Arbeit widerspiegelt das eben aufgezeigte Vorgehen. Zu Beginn eines Kapitels wird dessen Gliederung jeweils kurz beschrieben. Im folgenden Kapitel *Theoretische Grundlagen und Stand der Forschung* werden die im Zusammenhang mit der vorliegenden Themenstellung wichtigen Begriffe definiert und aktuelle Forschungsarbeiten erwähnt. Zu Beginn des Kapitels werden die Eigenschaften von Daten aufgezeigt, welche die Basis jeder Visualisierung bilden. Es werden die Grundlagen der Visualisierung vermittelt, sowie die daran verknüpften Anforderungen, sowie die möglichen Gestaltungsmittel einer Darstellung erläutert. Ebenfalls wird in diesem Kapitel auf die Visualisierung von Multiparameterdaten eingegangen. Im letzten Abschnitt des wird das Grundkonzept der Visualisierungsmethode „Parallele Koordinaten Diagramme“, die in der vorliegenden Arbeit erweitert und werden soll, vorgestellt und deren Eigenschaften erläutert.

Im dritten Kapitel *Grundlagen und Herleitung* wird beschrieben, wie die Zeitabhängigkeit von Variablen visualisiert werden kann und worin sich der Zeitparameter von anderen Variablen unterscheidet. Anschliessend wird das Konzept eines dreidimensionalen Parallelen Koordinaten Diagramms skizziert. Es wird aufgezeigt, welche Überlegungen gemacht werden müssen und welche interaktiven Techniken umsetzbar sind. Die Implementierung eines Prototyps soll zeigen, dass das Konzept realisierbar ist.

Im Kapitel 4 wird die *Softwareumgebung und Programmiersprache* Java3D vorgestellt. Mit Hilfe einer Darstellung wird das implementierte Programm erläutert, der Aufbau und die Funktionsweise des Prototyps veranschaulicht.

Das *Visualisierungsdesign des Prototyps* wird im Kapitel 5 beschrieben. Hier werden die Kriterien und Rahmenbedingungen für die Entwicklung festgelegt. Ebenfalls wird aufgezeigt wie der Benutzer mit der Visualisierung interagieren und durch den Datenraum navigieren kann. Zum Schluss werden die Designentscheidungen, die für die praktische Umsetzung des Konzepts und der Erstellung eines Prototyps getroffen werden müssen, erläutert.

Im Kapitel *Visualisierungsbeispiele* werden die vier generierten Beispiele, welche anhand verschiedener, exemplarischer Datensätze erstellt wurden, gezeigt und erklärt.

Im Kapitel *Diskussion* werden die visualisierten Beispielbilder diskutiert und die gestellten Forschungsfragen beantwortet. Es wird aufgezeigt, auf welche Fragestellungen mit der erweiterten Methode der Parallelen Koordinaten eingegangen werden kann, wie der Parameter „Zeit" dargestellt wird und welche neuen Möglichkeiten zur Datenraumanalyse sich daraus ergeben.

In den *Schlussfolgerungen und Fazit* werden die Ergebnisse und die Erkenntnisse zusammengefasst. Es werden Verbesserungsmöglichkeiten vorgeschlagen, die einen Ausblick auf machbare Entwicklungen der vorgeschlagenen Visualisierungsmethode ermöglichen.

2 Theoretische Grundlagen und Stand der Forschung

Ziel dieses Kapitels ist die Vermittlung der theoretischen Grundlagen der Datenvisualisierung von multidimensionalen Daten. Dazu wird zunächst auf den Begriff der „Daten" und deren Typisierung eingegangen. Anschliessend werden Begrifflichkeiten zu der Datenvisualisierung erklärt und die Anforderungen, die an eine Visualisierung gestellt werden, formuliert. Es soll untersucht werden inwiefern sich die Anforderungen an dreidimensionale Visualisierungen von den allgemeinen Anforderungen an eine Visualisierung, unterscheiden. Im letzten Abschnitt folgt die Erklärung des Konzepts der „Parallelen Koordinaten", auf dem die vorliegende Arbeit aufbaut, und für deren Fragestellungen von zentraler Bedeutung ist.

2.1 Daten

Die Grundlage jeder Visualisierung bilden die verwendeten Daten. Als Daten werden logisch strukturierte Informationseinheiten bezeichnet, welche Beobachtungsresultate repräsentieren oder Messungen von Prozessen beschreiben (vgl. Andrienko u. Andrienko, 2006: 17-18). Gemäss Schumann u. Müller (2000: 27) unterscheiden sich Daten in ihrer Struktur, ihrer Dimensionalität und durch den Datentyp voneinander. Erst durch die gezielte Analyse können beobachtete Prozesse und Phänomene in Daten, beurteilt und verstanden werden. Andrienko u. Andrienko (2006: 17) unterscheiden zwei Datentypen: Referenzen und Attribute, welche auch als unabhängige und abhängige Variablen bezeichnet werden können. Eine zusammengefasste oder gruppierte Einheit von unabhängigen und abhängigen Variablen wird als Datensatz bezeichnet. Ein Datensatz kann auf einem abstrakten Level als eine Korrespondenz zwischen Referenzen, z.B. Werte von unabhängigen Variablen (Zeit, Raum, Bestand) und Charakteristika, z.B. Attributwerte angesehen werden (Andrienko u. Andrienko, 2006:17). Schumann u. Müller (2000: 35) erachten folgende Aspekte der Datenmerkmale[1] als wichtig für die Datenvisualisierung:

- Datentyp
- Dimensionalität
- Wertebereich
- Strukturierung

2.1.1 Datentyp

Grundsätzlich werden Daten in **quantitative** und **qualitative** Daten unterteilt. Als quantitative Daten gelten messbare, numerische Werte, während nicht numerische und nicht standarisierte Informationseinheiten als qualitative Daten bezeichnet werden. Nebst dieser

[1] Die in einem Beobachtungsraum gemessenen, berechneten oder entworfenen Grössen werden als Merkmale bezeichnet (Schumann u. Müller, 2000: 35).

Unterteilung ist gleichzeitig auch die Unterscheidung zwischen **strukturierten** (z.B. Werte in Tabellen) und **unstrukturierten** (ohne Einteilung oder Aufbau) Datentypen möglich.

2.1.2 Dimensionalität

Die Dimensionalität von Daten spielt in der vorliegenden Arbeit eine bedeutende Rolle. Um sie zu bestimmen kann die Einteilung von Shneiderman (1996) verwendet werden:

- **1-dimensional**: Lineare Datentypen wie Textdaten, alphanumerische Listen von Namen, Quellcode von Programmen etc. Die Messung und Beurteilung von Daten dieses Datentyps können allein anhand der Länge vorgenommen werden.
- **2-dimensional**: Geographische Karten, Zeitungen, Pläne gehören zu den zweidimensionalen Daten. Sie können z.B. anhand ihrer Länge, Breite, Anzahl, Alter oder Breite gemessen werden. Die zwei abhängigen Variablen der Daten können auf je einer Achse abgetragen werden und so in einem Diagramm dargestellt werden.
- **3-dimensional**: Dreidimensionale Daten repräsentieren reale Objekte, wie beispielsweise chemische Moleküle oder Gebäude. Sie lassen sich typischerweise durch ihre Höhe, Breite und Länge messen. 3D-Computergraphiken werden auch zu diesem Datentyp gezählt.
- **multidimensional**: Tabellarische Daten aus relationalen Datenbanken und Statistiken mit mehreren Attributen müssen in mehreren Dimensionen visualisiert werden. Daten mit n unterschiedlichen Attributen können als Punkte in einem n-dimensionalen Raum dargestellt werden. Zu den Hauptaufgaben der Analyse solcher Datentypen gehört das Auffinden von Mustern und Strukturen, das Aufdecken von Korrelationen und Beziehungen, Lücken oder Ausreissern.
- **hierarchisch geordnet**: Strukturierte Datensammlungen mit hierarchischen Verbindungen, wie beispielsweise Telefonverbindungen oder Baumdiagramme.
- **Netzwerke**: Als Netzwerke gelten strukturierte Datensammlungen ohne hierarchischen Aufbau.
- **temporal**: Daten mit Zeitbezug werden als temporale Daten bezeichnet.

Die Fragestellungen, welche in der vorliegenden Arbeit beantwortet werden, betreffen ausschliesslich den multidimensionalen Datentyp. Es wird nicht unterschieden zwischen temporalen und multidimensionalen Daten, welche ein zeitliches Attribut beinhalten. Von multidimensionalen Daten wird gesprochen, sobald für einen Beobachtungspunkt mehrere Merkmale in mehreren Ausprägungen vorliegen. Der Begriff „multi“ soll in dieser Arbeit analog zu Harris (1999) verwendet werden, indem zwischen „1“ bzw. „2 und mehr“ unterschieden wird. Unter multidimensionalen Daten verstehen wir also Daten mit mindestens zwei abhängigen Variablen *n* (Schumann, 2000: 171). Mehrdimensionale Daten werden zusammen mit den multivariaten Daten, unter dem Begriff Multiparameterdaten, zusammengefasst. Während mehrdimensionale Daten typischerweise mehrere unabhängige Variablen eines Beobachtungsraum beschreiben, werden unter multivariaten Daten abhängige Variablen in einem bestimmten Merkmalsraum verstanden. Nach Schumann (2000: 172) sind die Begriffe multidimensionale Daten und multivariate Daten in einander überführbar und können deshalb synonym verwendet werden. In den folgenden Kapiteln wird deshalb meistens von mehrdimensionalen oder multidimensionalen Daten gesprochen.

2.2 Datenvisualisierung

2.2.1 Definitionen und Begriffserklärung

Es existiert eine Vielzahl an unterschiedlichen Begriffen um die visuelle Repräsentation von Daten zu beschreiben. Aber da alle Definitionen das Visualisieren der in den Daten enthaltenen Information beinhalten, können sie unter dem allgemeinen Begriff **Datenvisualisierung** zusammengefasst werden.

Die zentrale Aufgabe der Datenvisualisierung besteht darin, abstrakte Daten oder Zusammenhänge in einer aussagekräftigen, graphischen Form darzustellen, also Informationen zu visualisieren. Man bezeichnet sie deshalb auch als Informationsvisualisierung. Die Vorteile der grafischen Darstellung von Daten, gegenüber beispielsweise einer tabellarischen Darstellung, beruhen auf der bemerkenswerten Fähigkeit des menschlichen Gehirns zur Verarbeitung grafischer Informationen. Über die Hälfte[2] der Sinneszellen eines Menschen dienen dazu, Informationen aus seiner Umwelt durch visuelle Signale wahrzunehmen (Inselberg, 2009: 1). Bis zu 100 Megabyte graphischer Information kann der Mensch pro Sekunde verarbeiten (Kokjer, 1987).

> *"Information visualization can be defined as the use of computer-supported, interactive, visual representations of abstract data (...) to amplify cognition"* (Card et al., 1999).

Durch die visuelle Repräsentation grosser Datenmengen wird dem menschlichen Benutzer ein Einblick in die innere Struktur der Daten ermöglicht und der Austausch von Arbeitsergebnissen vereinfacht. Primäres Ziel der Informationsvisualisierung ist, im Gegensatz zur wissenschaftlichen Visualisierung[3], nicht die Abbildung selbst, sondern vielmehr die Möglichkeit, Muster und Zusammenhänge im Datenraum aufzudecken. Innere, vorher verborgene Zusammenhänge können durch eine geeignete Visualisierung vom Betrachter intuitiv erfasst, verstanden und bewertet werden. Das Finden der wertvollen, bzw. für eine bestimmte Fragestellung relevanten Informationen, ist in unstrukturierten, multidimensionalen Datensätzen ist keine leichte Aufgabe.

Das Wort „visualisieren" bedeutet u.a. „erkennbar machen für den Geist oder die Vorstellung". Visualisierung ist also eine kognitive Aktivität zur Bildung eines mentalen Modells. Diese Definition ist deshalb auch zugleich eine der wichtigsten Anforderungen an eine Visualisierungsmethode: sie muss Daten und die damit zusammenhängenden Prozesse und Phänomene *für die Vorstellung* des Forschenden *erkennbar machen* (Andrienko u. Andrienko, 2006: 166). Auf weitere Anforderungen an eine Visualisierung wird im Kapitel 2.2.2 eingegangen.

[2] wenn visuelle Signale als „ Lichtstimulation der Augen" definiert werden und die Anzahl Nervenzellen der Gross- und Kleinhirnrinde (Kortex) als Gesamtanzahl angenommen wird.

[3] Der Begriff „wissenschaftliche Visualisierung" wurde 1987 von McCormick, DeFanti und Brown geprägt. Er bezeichnet eine Methode bei der physisch Daten durch visuelle Attribute abgebildet werden.

Die Datenvisualisierung steht in unmittelbarem Zusammenhang mit der Analyse von Daten und wird in der Forschung während verschiedener Arbeitsphasen eingesetzt:

- Explorative Analyse
- Konfirmative Analyse
- Präsentation

Am Anfang eines Projektes bzw. einer Fragestellung wird beispielsweise häufig eine *explorative* Datenanalyse durchgeführt. Die Datenvisualisierung bildet häufig den Ausgangspunkt einer Untersuchung. In dieser Phase können mit Hilfe der Datenvisualisierung allfällige Hinweise auf Zusammenhänge untersucht werden oder ungerichtet Strukturen und Informationen in den Daten aufgedeckt werden. Im Anschluss an die explorative Datenanalyse erfolgt in der *konfirmativen* Untersuchung (statistische oder mathematische Datenanalyse) eine weitere Visualisierung um eine aufgestellte Hypothese zu prüfen. Darauf kann erneut eine visuelle Datenanalyse mit modifizierten (z.B. durch Skalierung) oder reduzierten Daten (z.B. eine Teilmenge) erfolgen (Elsner, 1997: 1). Zum Schluss werden die Erkenntnisse zusammengetragen und in einer *Präsentation* dargestellt und zugänglich gemacht. Die Datenvisualisierung ist also während dem gesamten Forschungsprozess von wichtiger Bedeutung.

Um abstrakte Daten graphisch veranschaulichen zu können, muss eine Abbildung auf geometrische Beschreibungsformen gefunden werden. Die Konvertierung von Daten in eine graphische Darstellung durchläuft mehrere Stufen, die in einer Visualisierungspipeline angeordnet sind (vgl. Abbildung 2).

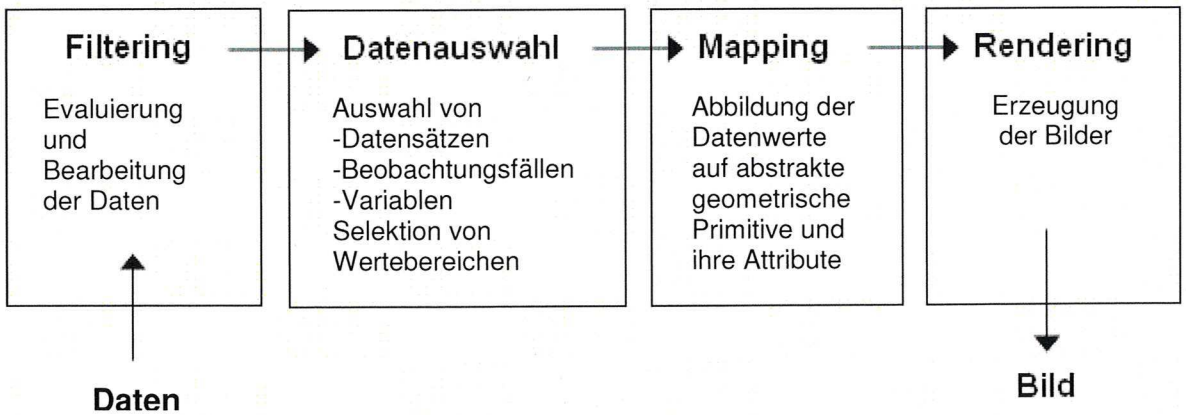


Abbildung 2: Visualisierungspipeline multivariater Daten (Quelle: Schumann u. Müller, 2000: 174)

Während dem *Filtering*, das auch als Datenaufbereitung bezeichnet werden kann, durchlaufen die Rohdaten verschiedene Operationen, welche die Daten aufbereiten und evaluieren. Zu dieser *Datenauswahl* gehören beispielsweise das Reduzieren und das Vervollständigen des Datensatzes, sowie das Berechnen von Extrema. Nach dieser vorbereitenden Arbeit wird der Datensatz eingegrenzt und einzelne Wertebereiche selektiert. Erst während dem *Mapping* werden die Datenwerte auf geometrische Primitve und den dazugehörigen Attributen abgebildet. Das *Rendering* beschreibt schliesslich das Erzeugen der Bilder aus den Primitiven und ihren Attributen (Holzhüter, 2005).

2.2.2 Anforderungen an eine Visualisierung

Wie bereits erwähnt wurde, liegt das Hauptziel einer Visualisierung darin, die Abbildung von Daten, Modellen oder Konzepten auf geometrische Primitive und deren Attribute zu ermöglichen (Tominski et al., 2002). Um dieses Ziel zu erreichen, muss eine Visualisierung bestimmte Anforderungskriterien erfüllen. Eine der bekanntesten Formulierung für diese Kriterien, geht auf Mackinlay (1986) zurück und besagt, dass eine Visualisierung möglichst

- expressiv (ausdrucksfähig)
- effektiv und
- angemessen sein sollte.

Als wichtigste Vorraussetzung für eine Visualisierung gilt die möglichst unverfälschte Wiedergabe der darzustellenden Datenmenge. Es sollen nur die tatsächlich in den Daten enthaltenen Informationen dargestellt werden. Auf diese Weise wird dem Betrachter eine unverfälschte Vermittlung aller relevanten Aspekte garantiert und das Risiko einer Fehlinterpretation minimiert. Das Kriterium der *Expressivität* muss erfüllt sein, um eine qualitativ gute und angemessene Darstellung der Informationen zu erreichen (Schumann u. Müller, 2000: 9).

Die *Effektivität* einer Visualisierung beschreibt deren Eignung unter Berücksichtigung der Zielsetzung und des optimalen Anwendungskontextes. Das Effektivitätskriterium gibt deshalb Aufschluss über die Fähigkeit einer Darstellungsform, die in ihr enthaltenen Informationen möglichst verständlich zu veranschaulichen und auf eine intuitive Weise an den Betrachter zu vermitteln. Weil die Effektivität nebst den Daten von weiteren Einflussfaktoren abhängt, ist die effektivste Darstellung jedoch nicht zwangsläufig auch die beste Darstellungsform.

Dieser Tatsache widmet sich die dritte Anforderung an eine Visualisierung: die *Angemessenheit*. Da die Generierung einer Visualisierung mit Kosten und einem gewissen Rechen- und Ressourcenaufwand verbunden ist, stellt sich neben der Frage der Effektivität auch die Frage der Angemessenheit. Die Angemessenheit einer visuellen Darstellung bestimmt dem zufolge den Aufwand und die Kosten für die Durchführung des Visualisierungsprozess (Schumann u. Müller, 2000: 12).

Diese drei Kriterien bestimmen zusammen die Eignung einer graphischen Darstellung für ein festgelegtes Anwendungsziel und beeinflussen die Qualität der Visualisierung.

> *„Die Qualität einer Visualisierung definiert sich durch den Grad, in dem die bildliche Darstellung das kommunikative Ziel der Präsentation erreicht. Sie lässt sich als das Verhältnis von der vom Betrachter in einem bestimmten Zeitraum wahrgenommenen Information zu der im gleichen Zeitraum zu vermittelnden Information beschreiben. Die Qualität einer Visualisierung ist somit ...abhängig von den Charakteristika der zugrunde liegenden Daten und ihren Eigenschaften, dem Bearbeitungsziel, den Eigenschaften des Darstellungsmediums sowie den Wahrnehmungskapazitäten und Erfahrungen des Betrachters.“* (Schumann u. Müller 2000: 7).

2.2.3 Gestaltungselemente einer Visualisierung

In einer Visualisierung werden die Informationen eines Datensatzes mit Hilfe graphischer Elemente, wie Punkte, Linien, Kurven und geometrischen Formen in Verbindung mit Darstellungsattributen wie Farbe und Textur, repräsentiert. Um diese graphischen Elemente gezielt und effektiv zur Erzeugung von Darstellungen bzw. Visualisierungen einzusetzen, entwickelte Bertin (1974) die Theorie der graphischen Semiotik, eine Wissenschaft der graphischen Zeichen. Als „... Mittel der graphischen Darstellung zur Transkription von Ähnlichkeits-, Ordnungs- und Proportionalitätsbeziehungen ..., die das Auge in Bezug auf Flecken wahrnehmen kann" (Bertin, 1982: 186) definiert Bertin acht verschiedene visuelle Variablen: Grösse, Helligkeitswert, Muster/Textur, Farbe, Orientierung/Richtung und die Form. MacEachren (1994) ergänzt diese Systematik indem er die visuelle Variable Farbe weiter unterteilt in Schärfe, Sättigung und Helligkeit. Obwohl die acht visuellen Variablen von Bertin auch für dreidimensionale Darstellungen gelten, wurden sie im Zuge der Entwicklung von interaktiven, mehrdimensionalen Darstellungen (Animationen etc.) laufend erweitert. DiBaise et al. (1991) und Dransch et al. (2000) fügen weitere Variablen dazu: Zeitpunkt, Veränderungsraten, Ordnung, Synchronisation, Dauer. Die Abbildung 3 zeigt die von Bertin entwickelten, visuellen Variablen (ohne die Variable Position des Elements):

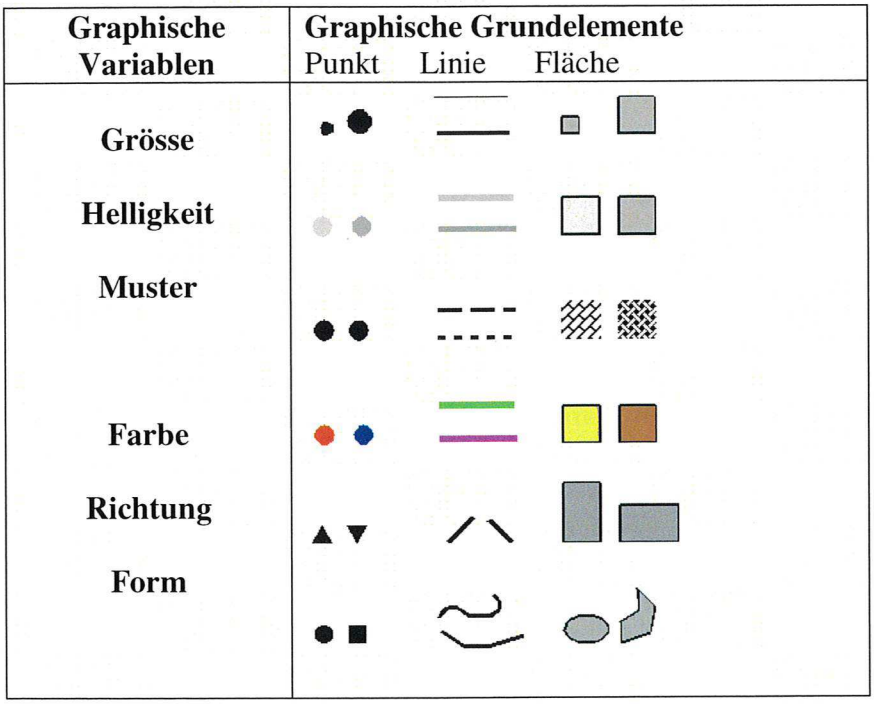


Abbildung 3: visuelle Variablen nach Bertin (1974), (Quelle: eigene Darstellung nach Universität Münster, 2007)

Durch die Verwendung dieser Variablen kann gemäss Hake et al. (2002: 107) eine objektive Gliederung der Darstellung, eine verbesserte Anschaulichkeit auf der Basis von Assoziationen, sowie die subjektive Bewertung durch Betonen und in den Hintergrundstellen bestimmter Elemente, erreicht werden.

2.3 Visualisierung von Multiparameterdaten

Multiparameterdaten definieren sich dadurch, dass für jeden Beobachtungspunkt Werte für mehrere abhängige Variablen existieren (vgl. Kapitel 2.1.2). Da diese Werte erst zusammen den gesamten Datensatz bilden, müssen sie in einer graphischen Darstellung auch als Einheit erfassbar sein (Schumann u. Müller, 2000: 173). Gemäss Schumann u. Müller (2000: 171) liegt das Ziel der Visualisierung von Multiparameterdaten darin, „möglichst viele Parameterwerte gleichzeitig darzustellen, um Zusammenhänge besser erkennen zu können (...)". Nur durch die Darstellung der Gesamtheit aller Werte kann gewährleistet werden, dass wichtige Zusammenhänge im Datensatz nicht verloren gehen. Trotz der erkennbaren Grundgesamtheit sollen aber einzelne, den Betrachter interessierende Beobachtungsfälle, identifizierbar bleiben. Die Auswahl und das Extrahieren eines einzelnen Merkmals kennzeichnet eine qualitativ gute Visualisierung für multidimensionale Datensätze. Die, für einen bestimmten Datensatz, geeignete Visualisierungsmethode wird aufgrund der darzustellenden Variablen ausgewählt. Um diese Auswahl treffen zu können, müssen folgende Fragen zu diesen Variablen gestellt werden:

- Welche Werte gehören zu einem Beobachtungspunkt?
- Welche Werte liegen für einen Beobachtungsfall vor?
- Welche Werte nimmt die abhängige Variable an?

2.4 Anforderungen an 3D-Visualisierungen

Multidimensionale Daten werden immer häufiger auch in entsprechend multidimensionalen Darstellungen visualisiert. Hauptsächlich dreidimensionale Visualisierungsformen haben in den letzten Jahrzehnten einen grossen Aufschwung erlebt und sind in den Mittelpunkt aktueller Forschungsprojekte gerückt. 3D-Visualisierungen erscheinen u.a. geeignet für die Darstellung multidimensionaler Datenräume geeignet, weil der Mensch auch in einer dreidimensionalen Umwelt lebt und es seinen kognitiven Fähigkeiten entspricht, Dinge dreidimensional zu betrachten und zu interpretieren. Die neue Möglichkeiten, die sich bei der Verwendung einer dritten räumlichen Dimension in einer Visualisierung eröffnen, können aber nur genutzt und ausgeschöpft werden, wenn der Betrachter die Geometrien und Attribute auch richtig erkennen kann. Die Länge und Position von Linien im Raum, sowie die Form von geometrischen Primitiven, müssen korrekt eingeschätzt und interpretiert werden können. Diese Voraussetzung ist deshalb besonders wichtig, da bei der Darstellung von dreidimensionalen, graphischen Elementen mit einem 2D-Medium (Bildschirm) perspektivische Verzerrungen und Verdeckungen entstehen können (Schumann u. Müller, 2000: 175). Die drei Hauptanforderungen an 3D-Visualisierungen formuliert Holzhüter (2005) wie folgt:

- **Räumliche Trennung**: geometrische Objekte im Raum sollten so angeordnet werden, dass sie sich nicht gegenseitig überzeichnen oder verdecken.

- **Räumliche Identifizierbarkeit**: alle Achsen und deren Funktionen müssen eindeutig identifizierbar und voneinander unterscheidbar sein. Der Verlauf von Formen muss für den Betrachter verfolgbar sein, dass alle visualisierten Daten beurteilt werden können.

- **Unterstützung der Erkennbarkeit**: Die Visualisierung sollte möglichst viele Funktionen und Möglichkeiten beinhalten, die das Erkennen von Strukturen in der zu untersuchenden Datenmenge unterstützen. In der Regel handelt es sich hierbei um

    - die Darstellung aller Ausprägungen eines definierten Zeitpunktes oder Bereichs
    - die Darstellung des zeitlichen Verlaufs eines Merkmals
    - der Vergleich zweier Merkmale
    - die Darstellung der Häufigkeiten
    - die Gesamtdarstellung aller Merkmale

Diese drei Kriterien sind voneinander abhängig und nicht immer leicht voneinander zu unterscheiden. Beispielsweise ist die räumliche Identifizierbarkeit bedingt durch die Trennung der einzelnen geometrischen Objekte. Aber erst mit der Anforderung der Erkennbarkeit wird gewährleistet, dass räumlich voneinander getrennte Objekte auch im dreidimensionalen Raum sichtbar bleiben.

Für die Visualisierung multidimensionaler Daten existieren unterschiedliche Konzepte und Ideen. Grinstein et al. (2001) geben einen Überblick.

2.5 Parallele Koordinaten

Das Konzept der Parallelen Koordinaten[4] wurde erstmals von Inselberg (1985) vorgestellt, welcher die Methode bis heute mathematisch untersucht und weiterentwickelt. Wegman (1990) schlägt die Methode der Parallelen Koordinaten erstmals für die explorative Analyse vor und diskutiert sie im Rahmen der Datenanalyse (Unwin et al., 2006: 144). Seit der formalen Einführung vor über 20 Jahren haben sich die Parallelen Koordinaten zu einer anerkannten und effizienten Visualisierungsmethode entwickelt um komplexe, multidimensionale, numerische Daten in einer zweidimensionalen Graphik darzustellen.

Durch das Aufziehen mehrerer paralleler Merkmalsachsen, kann mit dieser Visualisierungsmethode eine Vielzahl an Variablen gleichzeitig dargestellt werden. Auf diese Weise überführen Parallele Koordinaten multidimensionale Beziehungen in zweidimensionale Muster bzw. Formen, die sich besonders gut für die visuelle Datenraumerkundung eignen. Durch die graphische Darstellung der multidimensionalen Beziehungen zwischen den Variablen einzelner Objekte, unterscheiden sich die Parallelen Koordinaten von anderen Visualisierungsmethoden, die oftmals nur begrenzte Punktmengen aufweisen (Inselberg, 1997).

[4] engl. Parallel Coordinates

Das Ziel und den Zweck der Parallelen Koordinaten liegt gemäss Inselberg (2009: 3) in der Visualisierung von multivariaten bzw. mehrdimensionalen Problemen, ohne Verlust der in den Daten enthaltenen Information. Zusätzlich dazu zeichnen sich die Parallelen Koordinaten durch folgende Eigenschaften aus (Inselberg, 1997):

- niedere Komplexität der Repräsentation, da die Anzahl der Achsen der Anzahl an Dimensionen (Variablen) entspricht
- gilt für alle ganzzahligen *N*
- jede Variable wird gleich behandelt
- das dargestellte Objekt kann wieder erkannt werden durch projektive Transformationen (z.B. Rotation, Translation, Skalierung, Perspektive…)
- die Darstellung vermittelt intuitiv die Informationen über die Eigenschaften des dargestellten n-dimensionalen Objektes
- die Methode basiert auf strengen, mathematischen Regeln und Algorithmen

Im folgenden Abschnitt wird kurz das Konzept der Parallelen Koordinaten kurz erläutert. Die geometrische Einzelheiten, sowie die mathematische Herleitung der Methode sind nicht Bestandteil dieser Arbeit. Darüber existiert genügend wissenschaftliche Literatur (vgl. Inselberg, 2009; Wegman, 1990).

2.5.1 Definition der Parallelen Koordinaten

Bei den parallelen Koordinaten handelt es sich um eine Visualisierungsmethode n-dimensionaler Daten, bei der die Merkmalsachsen parallel und äquidistant zueinander angeordnet werden. Für jede positive Ganzzahl *N* wird im euklidischen, *n*-dimensionalen Raum R^n ein Koordinatensystem aufgezogen. Auf der y-Achse werden *n* Kopien der originalen Linie äquidistant und im rechten Winkel zur x-Achse dargestellt. Sie bilden die Merkmalsachsen des parallelen Koordinatensystems, die alle dieselbe positive Orientierung der y-Achse haben. Jede Dimension wird so auf einer eigenen Achse abgetragen.

Ein Punkt ξ mit den Koordinaten $(x_1,x_2,\ldots x_n)$, der im kartesischen Koordinatensystem als Punkt eingetragen wird, stellt im Parallelen Koordinaten Diagramm eine polygonale Verbindungslinie, deren *n* Eckpunkte auf *n* horizontalen oder vertikalen Achsen liegen, dar (siehe Abb. 4). Es besteht also eine Ähnlichkeit zwischen den Punkten in R^N und den polygonalen Verbindungslinien mit den Eckpunkten auf $x_1,x_2,\ldots, x_N$.

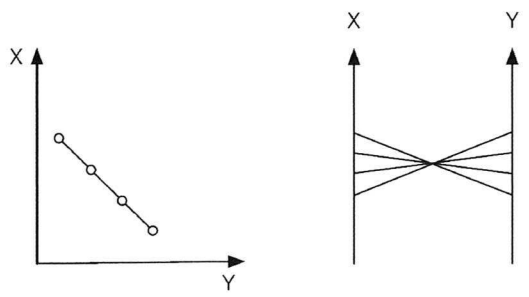


Abbildung 4: Das Grundkonzept der Parallelkoordinaten: parallele Achsen an Stelle von orthogonalen Achsen. Die vier Punkte (links) werde durch die vier Linien (rechts) dargestellt. (Quelle: Siirtola et al.,2009)

Die maximalen und minimalen Werte der einzelnen Attribute werden üblicherweise zu den oberen und unteren Endpunkten auf den Merkmalsachsen skaliert. Anhand der entstehenden Merkmalslinien in einem Parallelen Koordinaten Diagramm, können multivariate Signaturen für jede einzelne Beobachtung, sowie die Beziehungen zu den anderen Variablen, abgelesen werden (Edsall, 2003: 607).

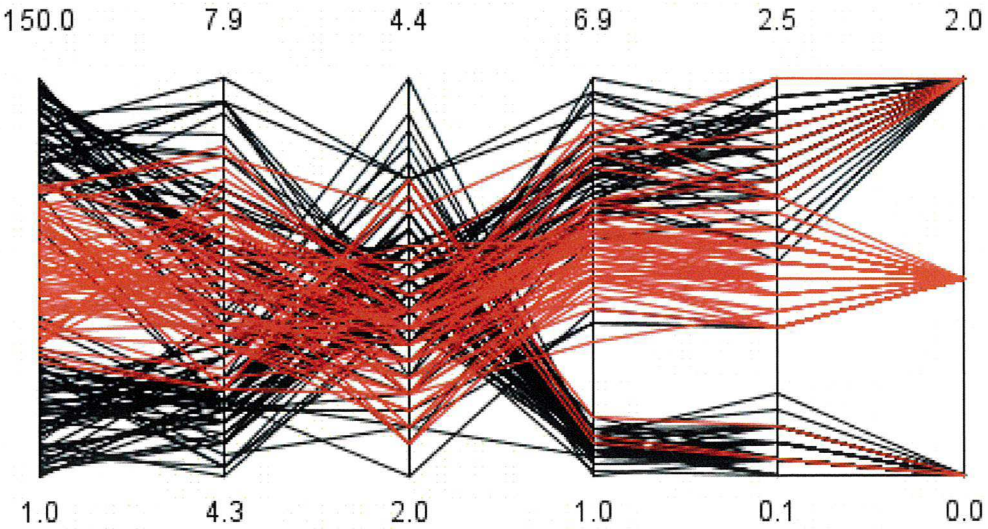


Abbildung 5: Beispiel eines Parallelen Koordinaten Diagramms mit 6 Variablen (Quelle: Technische Universität Wien, 2006)

Chen et al. (2008, 164) sieht die Anwendungsmöglichkeiten der Parallelen Koordinaten in der explorativen Datenanalyse hauptsächlich in folgenden drei Bereichen:

- **Übersicht verschaffen**: Keine andere statistische Darstellung kann so viel Information (Beobachtungen und Variablen) auf einen Blick visualisieren. Damit eignet sich die Methode der parallelen Koordinaten vor allem dazu, sich einen ersten Überblick über den Datensatz zu verschaffen.

- **porträtieren, hervorheben**: Durch die Möglichkeit des Hervorhebens einer einzelnen Beobachtung (*highlighting*) kann die Darstellung in Parallelen Koordinaten auch dazu verwendet werden, um ein Profil eines bestimmten Falles oder einer Gruppe zu erfassen und mit dem Rest des Datensatzes zu vergleichen.

- **monitoring**: Es besteht die Möglichkeit bestimmte Merkmale einer Teilmenge des Datensatzes in Bezug zur Gesamtheit zu betrachten.

2.5.2 Vor- und Nachteile

Parallele Koordinaten Diagramme eignen sich besonders dazu, um die Ergebnisse multidimensionaler Verfahren zu interpretieren. Dazu kann das Aufspüren von Ausreissern, die Anhäufung bestimmter Werte (Cluster) oder die Klassierung der Werte in einem multivariaten Kontext, gehören (Chen et al., 2008: 167). Gemäss Schumann u. Müller (2000: 186) realisieren parallele Koordinaten eine verlustfreie und eindeutige Abbildung eines n-dimensionalen Raumes in die Ebene. Ein weiterer Vorteil der Parallelen Koordinaten

Darstellung ergibt sich aus einer ihrer charakteristischen Eigenschaften: Alle Dimensionen (bzw. Variablen) werden gleich behandelt. Diese Tatsache lässt jederzeit eine Umsortierung der Dimensionen zu, wodurch sich der Datensatz auf verschiedene Weisen betrachten lässt. Durch die Umsortierung können neue Muster und Zusammenhänge in den Daten erkannt werden, die durch eine andere Visualisierungsmethode vielleicht unentdeckt geblieben wären.

Ein Nachteil der Parallelen Koordinaten Diagramme wird beim Visualisieren grosser Datensätze deutlich. Trotz der Fähigkeit, eine Vielzahl an Variablen gleichzeitig visualisieren zu können, stösst die Methode bei wenigen hunderten Linien an ihre Grenzen (vgl. Abb. 6). Die vielen Überschneidungen der Merkmalslinien führen zu einem optischen Wirrwarr (Clutter, Overflow) und überdecken die eigentliche Struktur des Datensatzes. In den letzten Jahren wurde mit grossem Bemühen daran gearbeitet, Verfahren zu entwickeln, um die Unleserlichkeit beim Darstellen vieler Attribute, zu reduzieren (Ellis u. Dix ,2006).

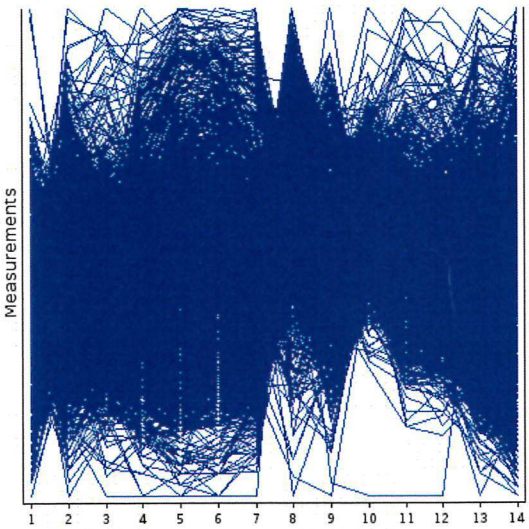


Abbildung 6: optisches Wirrwar bei vielen Merkmalslinien (Quelle: Chi-Square, 2001)

Der oben genannte Vorteil der möglichen Umsortierung der einzelnen Merkmalsachsen erweist sich zugleich als Nachteil der Methode. Das Parallele Koordinaten Diagramm zeigt eine bestimmte Abfolge einzelner Merkmale in einem Datenraum. Die erste Merkmalslinie wird aufgezogen durch die Achsen 1 und 2, die nächste durch die Achsen 2 und 3 und wiederum die nächste durch die Achsen 3 und 4 usw. Das bedeutet, dass eine interessante Struktur, wie beispielsweise eine Korrelation, in der Merkmalsebene 3 und 5 erst aufgedeckt werden kann, wenn die Merkmalsachen in die richtige Reihenfolge gebracht worden sind. Spezifische Softwareapplikationen, wie GEOVISTA, GGOBI etc., bieten dem Benutzer jedoch eine Hilfestellung für die interaktive Umsortierung der Merkmalsachsen an.

Ebenfalls setzen die teils komplexen Formen und Muster eines Parallelen Koordinaten Diagramms Fachwissen und ein geübtes Auge voraus, um die Graphik richtig und sinnvoll zu interpretieren oder eine visuelle Datenanalyse durchzuführen. Zugleich kann trotz der Sichtbarkeit aller Dimensionen nicht immer der ganze Datenraum auf einen Blick erfasst werden.

2.5.3 Erweiterte Parallele Koordinaten

Um die oben erwähnten Nachteile aufzuheben und die Lesbarkeit der herkömmlichen Methode der Parallelen Koordinaten zu verbessern, existieren verschiedene, projektbezogene Ansätze und Erweiterungen. Hauptsächlich durch die Kombination mit anderen Visualisierungsmethoden für multidimensionale Daten, kann den Nachteilen der Parallelen Koordinaten entgegen gewirkt werden. Die konventionelle Darstellung in Verbindung mit der Darstellung von Scatterplots (vgl. Robinson, 2007) erlaubt dem Betrachter beispielsweise eine genaue Analyse einzelner Merkmale und des gesamten Datenraums zugleich.

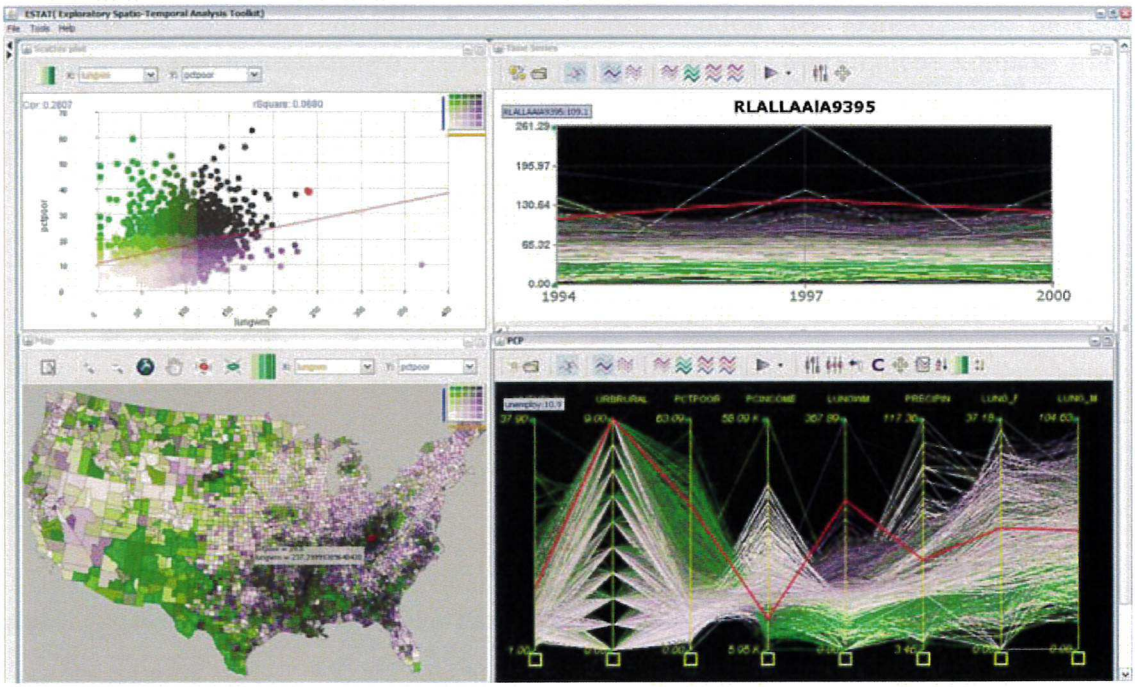


Abbildung 7: Die ESTAT Applikation fügt den Scatterplot (oben links), Karte (unten links), Zeitserien Graph (oben rechts) und das Parallele Koordinaten Diagramm (unten rechts) zusammen. (Quelle: Robinson, 2007).

Theus (2003) ergänzt die Merkmalslinien in den Parallelen Koordinaten mit Spinogrammen.

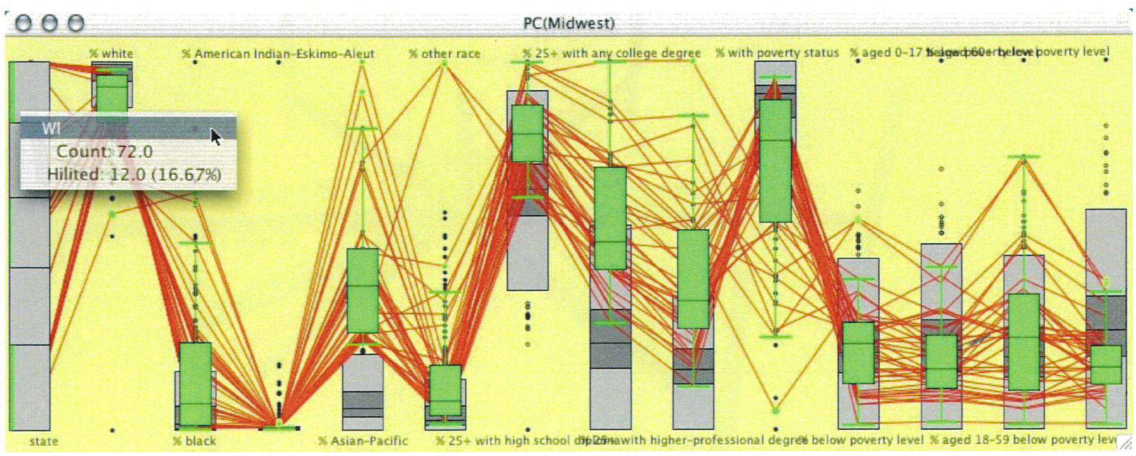


Abbildung 8: Parallele Koordinaten mit Spinogramm (Quelle: Theus,2003)

Nebst den Kombinationen von Parallelen Koordinaten mit anderen Visualisierungsmethoden, existieren weitere Ansätze um die Lesbarkeit der herkömmlichen Methode zu verbessern. Aktuelle Forschungsarbeiten beschäftigen sich damit, zweidimensionale Parallele Koordinaten, um eine Dimension zu ergänzen. Durch das Hinzufügen einer dritten Dimension wird versucht, eine effizientere Analysemöglichkeit eines Datensatzes zu finden. Mit solchen „dreidimensionalen" Erweiterungen der herkömmlichen Methoden beschäftigen sich verschiedene Forschungsgruppen:

Fanea et al. (2005) verbinden beispielsweise die zweidimensionalen Parallelen Koordinaten mit sog. StarGlyphen. Durch die räumliche Aufgliederung und Anordnung der Parallelen Koordinaten wird eine dreidimensionale Repräsentation des Datenraumes möglich (vgl. Abbildung 9). Korrelationen zwischen den einzelnen Merkmalen können mit dieser Visualisierungstechnik jedoch nur schwierig erkannt oder aufgedeckt werden.

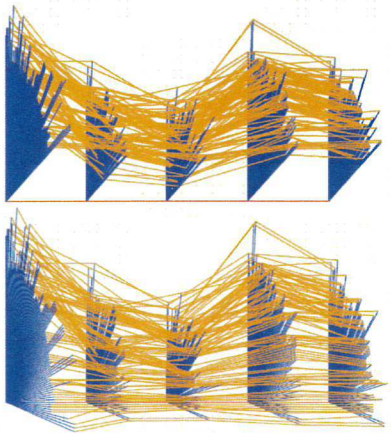


Abbildung 9: Parallele Glyphen (Quelle: Fanea et al., 2005)

Johansson et al. (2005) entwerfen eine gruppierte, multirelationale 3D-Erweiterungsmöglichkeit der Parallelen Koordinaten (Abbildung 10).

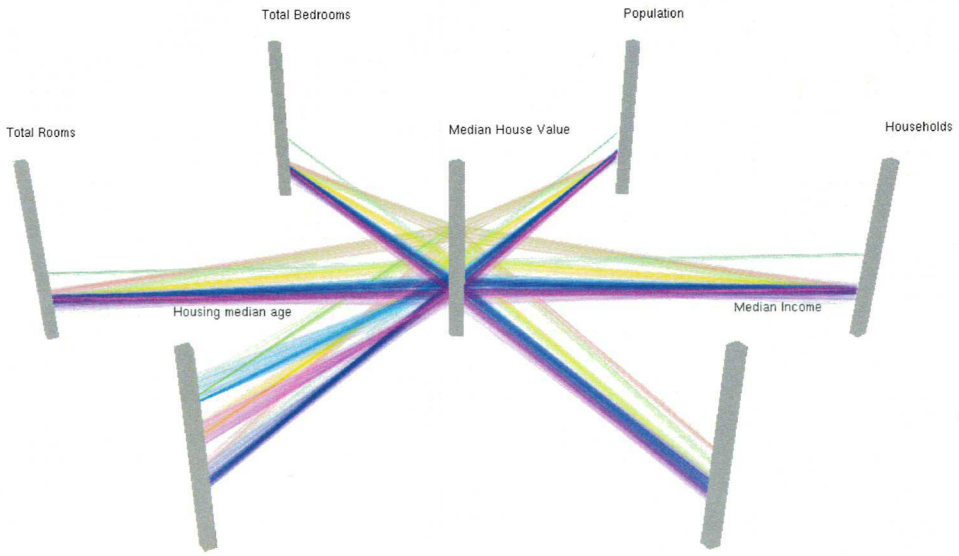


Abbildung 10: Die Technik der "clustered multirelational Parallel Coordinates", (Quelle: Johansson, 2005)

Diese spezielle Visualisierungstechnik der Parallelen Koordinaten erlaubt einem Betrachter eine Analyse der Beziehungen zwischen einer „Haupt"-Dimension in der Mitte der Darstellung, und den übrigen Dimensionen am Rand.

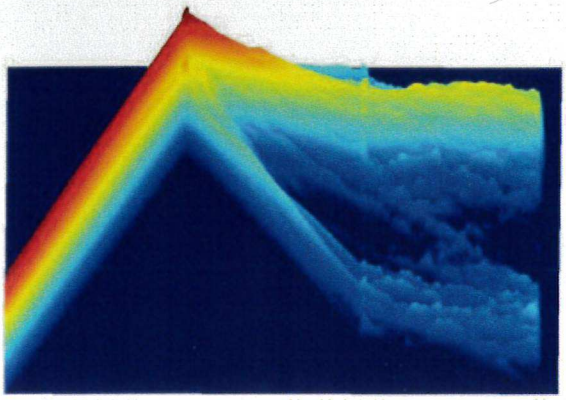


Abbildung 11: Parallele Koordinaten mit zusätzlicher Dichteinformation (Quelle: Streit, 2006)

Vor allem in Fachkreisen der Medizin, Chemie und Biologie wurden in den letzten Jahren mit erweiterten Parallelen Koordinaten Darstellungen experimentiert. Streit et al. (2006) versehen die herkömmliche Darstellungsform mit Dichteinformationen aus Scatterplots. Eine zusätzliche Dimension haben auch Rübel et al. (2008) in ihrer Arbeit im Bereich der Bioinformatik hinzugefügt. Bei diesem Datensatz wurde die räumliche Information eines Gens, welche für eine Analyse von genetischen Mustern unerlässlich ist, in die dritte Dimension der Parallelen Koordinaten überführt.

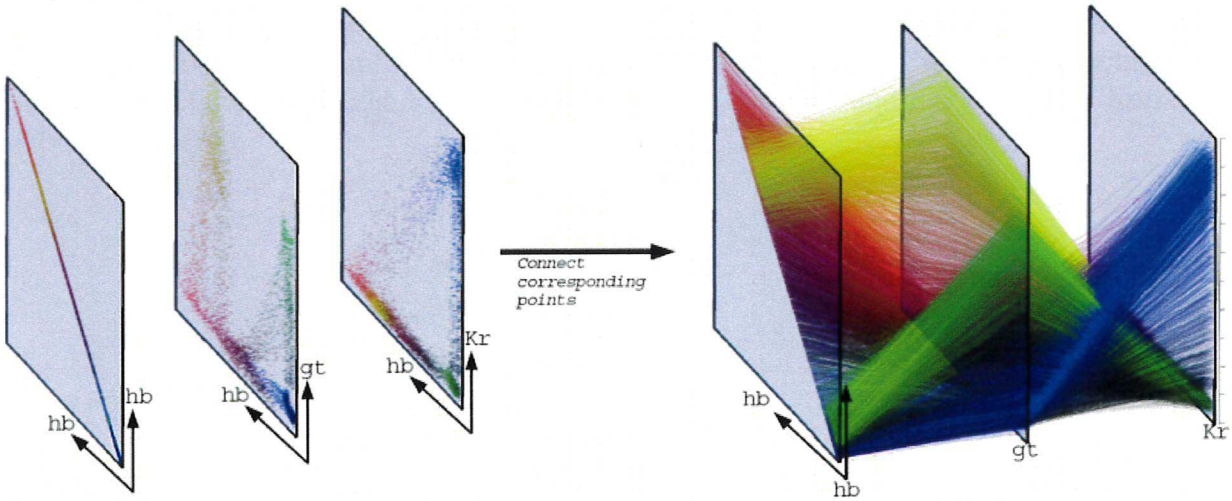


Abbildung 12: Vom 2D Scatterplot (links) zu dreidimensionalen Parallelen Koordinaten (rechts). (Quelle: Rübel et al., 2008)

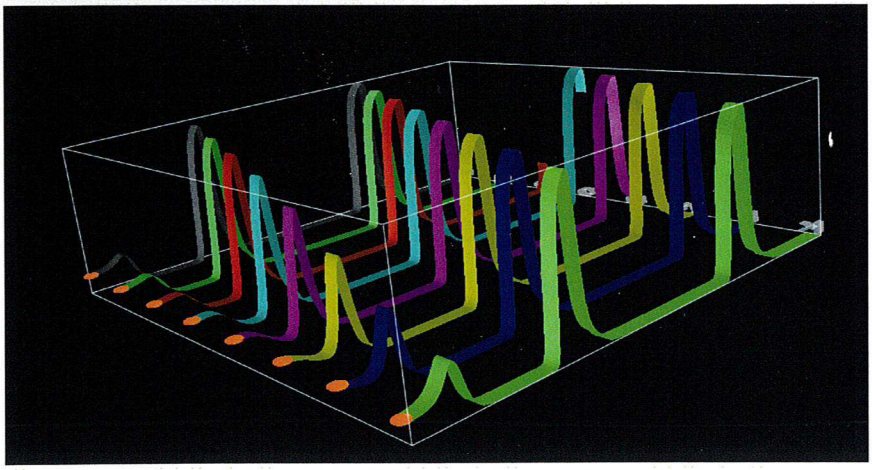


Abbildung 13: erweiterte PCP(Quelle: Technische Universität Wien, 1999)

Die Abbildung 13 zeigt einen weiteren Ansatz von dreidimensionalen Parallelen Koordinaten mit acht Variablen nebeneinander. Eine Forschungsgruppe der Technischen Universität Wien hat, basierend auf der Idee der erweiterten Parallelen Koordinaten von Wegenkittl et al. (1997), eine Visualisierungstechnik für chemische Prozesse in Zellkernen entwickelt. Das ist eine der ganz wenigen Arbeiten, die sich mit der Visualisierung eines zeitlichen Vorganges mittels Parallelen Koordinaten auseinander setzen. Mit dieser Frage beschäftigt sich deshalb die vorliegende Arbeit.

Eine Erweiterung der Parallelen Koordinaten um eine dritte Dimension, welche die Variable Zeit repräsentiert und spezifisch für die Darstellung von zeitabhängigen, multidimensionalen Datensätzen entworfen wurde, wurde bisher nicht oder nur ungenügend untersucht. Um die Parallelen Koordinaten auf diese Weise erweitern zu können, müssen zunächst die theoretischen Grundlagen zur Visualisierung von Zeit erläutert, und ein geeignetes Konzept entworfen werden.

3 Grundlagen und Herleitung

In diesem Kapitel wird die Variable Zeit und deren unterschiedliche Visualisierungsmöglichkeiten besprochen. Durch das Aufzeigen der zeitlichen Primitiven und Strukturen soll später die Forschungsfrage, nach der Besonderheit der Zeit in der Visualisierung, beantwortet werden können. Aufbauend auf diese theoretischen Grundlagen und Überlegungen wird in einem zweiten Teil des Kapitels ein Konzept entworfen, das die Visualisierung von zeitabhängigen, multidimensionalen Daten mittels der Darstellung der Parallelen Koordinaten umsetzen kann.

3.1 Die Visualisierung der Zeitabhängigkeit

Daten, die in einer Dimension des Beobachtungsraumes mit der physikalischen Grösse der Zeit assoziieren, bzw. eine abhängige Variable die Zeit ist, werden gemäss Tominski et al. (2003), als Daten mit Zeitbezug bezeichnet. Die Zeitabhängigkeit eines Datensatzes kann in unterschiedlichen Ausprägungen auftreten:

- **Daten mit statischem Zeitbezug:** Im Datensatz liegt nur ein einzelner Zeitpunkt oder Zeitraum vor.
- **Daten mit quasistatischem Zeitbezug:** Im Datensatz sind mehrere diskrete Zeitangaben vorhanden
- **Daten mit dynamischem Zeitbezug:** Der Datensatz weist eine kontinuierliche Zeitachse auf.

Nach Andrienko u. Andrienko (2006: 25) werden die Parameter Raum und Zeit implizit als unabhängige Variablen betrachtet, während sich die übrigen Attribute eines Datensatzes, dann auf die Variablen Raum und Zeit beziehen. Parameter, die nicht die Zeit oder den Raum beschreiben, in denen eine bestimmte Beobachtung gemacht wird, werden vom Betrachter intuitiv als abhängige Variablen behandelt. Diese Tatsache muss bei einer Visualisierung von Daten mit einem Zeitbezug insofern berücksichtigt werden, da die Variable Zeit je nach Darstellungsmethode als unabhängiger „Behälter" von Ereignissen fungieren kann oder aber als abhängige Variable.

Sobald die Zeit als Schlüsselattribut für die Datenanalyse gilt, müssen spezifische Methoden und Verfahren der Darstellung gefunden werden, um die Zeitkomponente eines Datensatzes richtig interpretieren zu können. Um einen zeitlichen Sachverhalt oder eine Veränderungen zu visualisieren, wird üblicherweise auf dreidimensionale Techniken zurückgegriffen. Die Zeit wird als die dritte Dimension des Raumes dargestellt und verläuft entlang einer Zeitachse. Die Darstellung des Zeitbezugs ist nach Schuman u. Müller (2000: 247) aber nicht unbedingt an

die physikalische Grösse der Zeit gebunden. Mit einer Zeitdimension in einer Visualisierung können nebst dem zeitlichen Charakter der Daten auch räumliche Veränderungen und Variationen in den Datenwerten dargestellt werden. Zeitliche Zusammenhänge in einem Datensatz werden mit Hilfe von topologischen Beziehungen (vor, nach, während, dazwischen etc.) und metrischen Beziehungen (Zeitdauer, Zeitabschnitt) ausgedrückt (Andrienko u. Andrienko, 2006: 348).

3.1.1 zeitliche Primitiven und Strukturen

Für eine Visualisierung des Parameters Zeit kann die Unterteilung und Definition von Frank (1998) verwendet werden. Er gliedert die Zeit in unterschiedliche Typen, die sich in einer orthogonalen Darstellung aufzeigen lassen. Eine Zeitachse wird aus verschiedenen zeitlichen Grundformen aufgebaut: dem Zeitpunkt und dem Zeitintervall.

- Als ein *Zeitpunkt* wird ein Moment verstanden.

- Ein *Zeitintervall* beschreibt eine zeitliche Form mit einer Ausdehnung ist. Das Intervall kann als der Zeitraum zwischen zwei Zeitpunkten oder als ein Zeitpunkt mit einer Zeitdauer aufgefasst werden.

Zeit kann auf zwei verschiedene Arten betrachtet und behandelt werden: als eine linear angeordnete Reihe oder Kreislauf bzw. als in sich wiederholende Zyklen:

- Eine lineare Zeit entspricht unserem natürlichen Empfinden der Zeit als Abfolge von geordneten Zeitprimitiven. Die Zeit verläuft von der Vergangenheit in die Zukunft.

- Eine *zyklische* Zeitachse besteht aus einer endlichen Anordnung von Zeitprimitiven (z.B. Jahreszeiten). In einer zyklischen Zeitreihe folgt auf die Zeitprimitive A eine Zeitprimitive B auf die wieder die Zeitprimitive A folgt usw. (z.B. nach dem Winter kommt der Sommer und dann wieder der Winter). In praktischen Anwendungen bzw. Visualisierungen ist es oft sinnvoll eine zyklische Zeitreihe in eine lineare Zeitreihe zu überführen (Aigner et al., 2007:6).

- *Verzweigende* Zeitachsen werden als graphisches Diagramm in dem zeitliche Primitiven in einer Abfolge angeordnet sind, dargestellt. Durch Auftrennung der Zeitachse an bestimmten Stellen in beliebig viele neue Zeitachsen, kann ein Alternativszenario beschrieben werden. Lineare und zyklische Zeitreihen können als Spezialfälle von verzweigenden Zeitachsen angesehen werden (z.B. lineare Zeitachsen zeigen nur in eine Richtung und besitzen jeweils nur eine einzige Abzweigung pro Zeitprimitive, während die graphische Darstellung einer zyklischen Zeitreihe einen Kreis beschreibt).

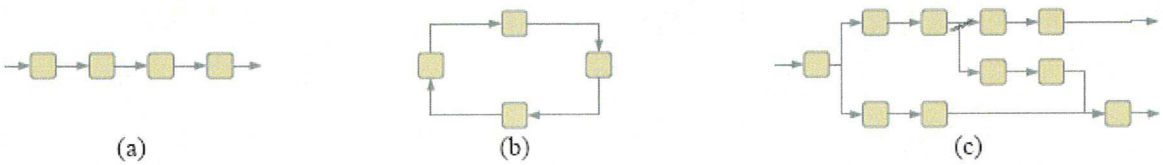


Abbildung 14: (a) lineare Zeitreihe, (b) zyklische Zeitreihe (c) verzweigte Zeitreihe (Quelle: Aigner et al., 2007)

Eine detailliertere umfassende theoretische Übersicht über die Zeit und ihre Strukturen gibt Hajnicz (1996).

3.1.2 Besondere Bedeutung des Parameters Zeit

Die Analyse von zeitlich orientierten Daten ist nicht selten eine der Hauptaufgaben in Forschungsfragen und Visualisierungsszenarios. In den vergangen Jahren entwickelte sich die Visualisierung von Daten mit Zeitbezug zu einem eigenen Forschungsgebiet, dem sich zahlreiche Forschungsgruppen (Aigner et al. (2007), Aigner et al. (2008), Daassi et al. (2006), Demsar et al. (2008), Carlis u. Konstan (1998), Müller u. Schumann (2003), Tominski et al. (2005)) widmen. Die meisten Daten, die heute gemessen und untersucht werden haben eine räumliche und eine zeitliche Komponente. Die Frage zu welchem Zeitpunkt verschiedene Merkmale eine bestimmte Ausprägung haben, ist für viele Fragestellungen relevant. In linear angeordneten Zeitfolgen existieren nur zwei Richtungen: von der Vergangenheit zur Zukunft und zurück. Dieser Umstand gilt auch für eine zyklisch angeordnete Datenfolge, denn auch hier folgen die einzelnen Zyklen linear aufeinander und jeder Zeitzyklus ist in sich wieder eine lineare Abfolge von kleineren Zeiteinheiten. Wenn die Zeit in einem Datensatz als linear betrachtet wird, so können die Zeiteinheiten und Zeitintervalle behandelt werden wie Werte oder Wertintervalle von Attributen und Referenzen beliebiger anderer linear geordneter Datensätze. Die Beschreibung der einzelnen Zeitpunkte oder Zeitabschnitte kann also gleich behandelt werden, wie irgendein anderes numerisches Attribut (Andrienko u. Andrienko, 2006: 346). Da bei der explorativen Datenanalyse aber zu Beginn der Forschungsarbeit nicht immer klar ist, ob es sich bei den Daten um eine lineare Zeitreihe handelt oder nicht, empfiehlt es sich, die Zeit besonders zu berücksichtigen und darzustellen. Hinzu kommt, dass die Wahrnehmung und Interpretation zeitlicher Koordinaten, und die Zuordnung von Ereignissen zu gewissen Zeitpunkten, unserem alltäglichen Umfeld entspricht. Der Parameter „Zeit“ wird vom Menschen im Allgemeinen als eine zusätzliche Dimension des Beobachtungsraumes aufgefasst (Schumann u. Müller, 2000: 57). Dieser Umstand soll bei einer Visualisierung berücksichtigt werden. Um ein intuitiv anwendbares und geeignetes Konzept für die Darstellung von multidimensionalen Daten mit Zeitbezug zu entwerfen, soll die Zeit in der dritten, räumliche Koordinatenachse dargestellt werden.

3.1.3 Ansätze zur Visualisierung der Zeitabhängigkeit in 3D

Über die Visualisierung von zeitlich orientierten Daten in einer dreidimensionalen Darstellung existieren unterschiedliche Meinungen. Einige Experten sind der Meinung, dass zwei Dimensionen ausreichen um eine effiziente und genaue Datenanalyse zu gewährleisten. In vielen Fällen erschwert eine dritte Dimension die Lesbarkeit einer graphischen Darstellung und kann so auch zu einem Verlust von Information führen. Unbestritten ist, dass eine weitere Dimension in einer Visualisierung die Möglichkeit bietet, zusätzliche Information in der Graphik unterzubringen. Einzelne Datentypen (z.B. Fliess- oder Volumendaten) verlangen geradezu eine dreidimensionale Darstellung und oft können die „befürchteten“ Nachteile einer 3D-Darstellung durch Interaktion oder visuelle Hilfestellungen kompensiert werden (Aigner et al., 2007: 9).

In der Literatur sind nur wenige achsenbasierte[5] 3D-Ansätze bekannt, die explizit zur Visualisierung zeitabhängiger Daten entwickelt wurden. In den folgenden Abschnitten werden vier dieser Ansätze kurz vorgestellt:

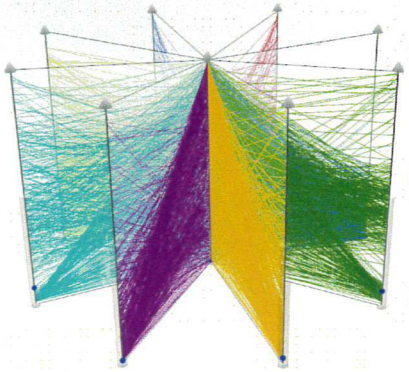


Abbildung 15: links: 3D Zeitrad (Aigner et al., 2008)

Eine Visualisierungsmethode, die entwickelt wurde um periodische Muster in multivariaten Datensätzen zu entdecken, ist das 3D-Zeitrad von Aigner et al. (2008). Um die Zeitachse in der Mitte herum sind mehrere Achsen kreisförmig angeordnet, welche die einzelnen Variablen repräsentieren. Für jeden Zeitpunkt des Datensatzes wird eine Linie zu dem entsprechenden Punkt auf der jeweiligen Merkmalsachse gezogen. Auf diese Weise entsteht ein „Zeitrad" das durch Rotation und Zoomen merkmalsbezogen untersucht werden kann.

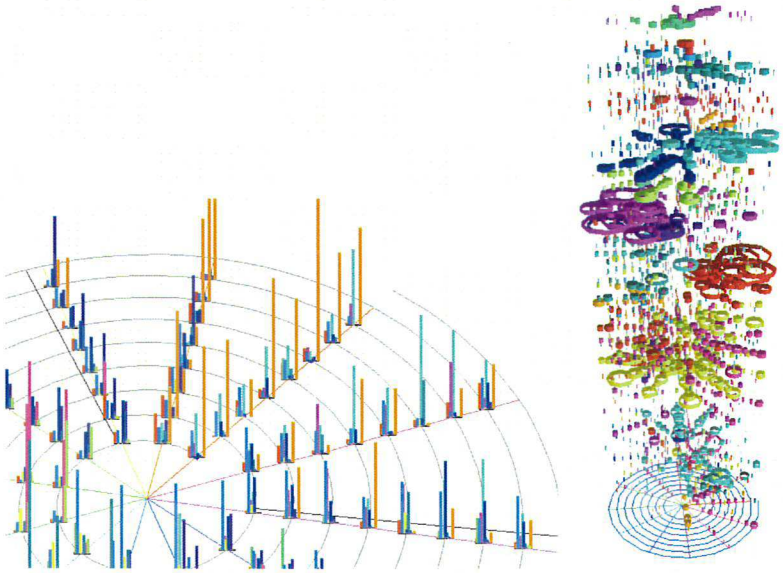


Abbildung 16: spiralförmige Zeitdarstellung für die Konsumation von12 Lebensmitteln während den Monaten zwischen 1980-1989, rechts: spiralförmige Darstellung prozentualer Anteile von 112 Lebensmitteln von 1980-1989 (Quelle: Carlis u. Konstan, 1998)

Die von Carlis und Konstan (1998) entwickelte Zeitspirale in Abb. 16 wurde entwickelt, um mehrere periodische, zeitabhängige Datensätze gleichzeitig darstellen zu können.

[5] Eine Definition erfolgt in Kapitel 3.2

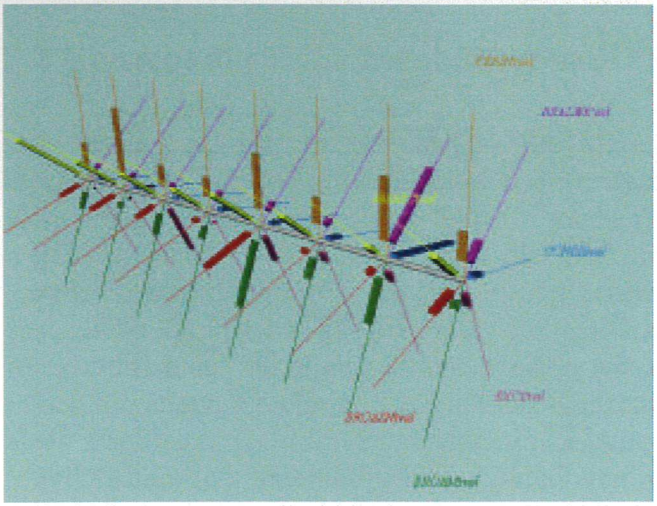


Abbildung 17: Temporal Star Darstellung für Börsenkurse über 8 Wochen (Quelle: Noirhomme-Fraiture, 2002)

Die sternförmige Darstellung eines Datensatzes mit Zeitbezug (vgl. Abbildung 17) wurde entworfen, um die Variation eines symbolischen Objekts im Laufe eines definierten Zeitintervalls mit einzelnen Zeitpunkten, zu visualisieren (Noirhomme-Fraiture, 2002). Die zentrale Achse in der Mitte des Graphs repräsentiert die Zeit. Die Visualisierung kann um ihre Achse gedreht, rotiert und verschoben werden.

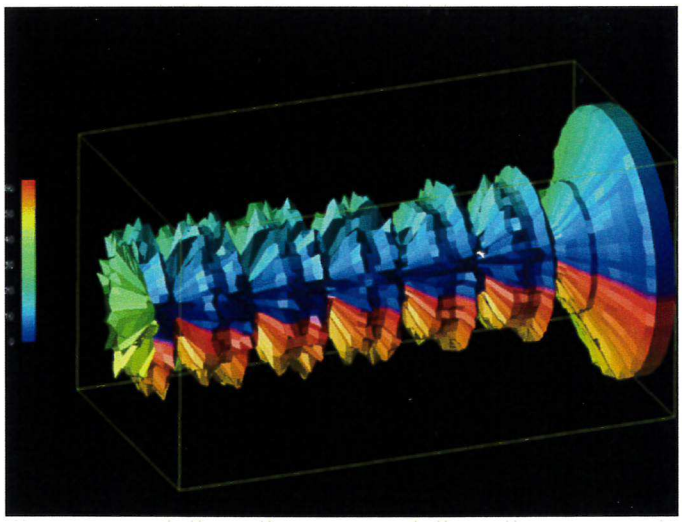


Abbildung 18: Kiviat Tube, (Quelle: Hackstadt u. Malony, 1995)

Die Kiviat Tube (Hackstadt u. Malony, 1995) ist eine dreidimensionale Form des Kiviatgraph und stellt die Prozesse der CPU entlang einer Zeitachse dar. Das Beispiel in Abb. 17 stellt die Auslastung des Prozessors eines Computers zu verschiedenen Zeitpunkten in farbigen, sternförmigen Achsen dar. Die Länge der einzelnen Achsen visualisieren die Stärken. Durch das Hintereinanderstellen der einzelnen Kiviatgraphen entsteht ein geschlossener 3D-Körper, der alle Ausprägungen entlang einer Zeitachse darstellt.

3.2 Konzept der zeitachsenbasierten 3D Parallelen Koordinaten

Die Visualisierungstechniken, welche die Merkmale eines Datensatzes durch eine Achse repräsentieren, werden als *achsenbasierte Techniken* bezeichnet. Die Methode der Parallelen Koordinaten soll nun zu einer solchen achsenbasierten Visualisierungstechnik weiterentwickelt werden. Da sich die herkömmliche, zweidimensionale Methode der Parallelen Koordinaten für die Darstellung zeitlicher Verläufe nur bedingt eignet, wird eine zusätzliche Dimension, in Form einer Zeitachse beigefügt. Mit Hilfe dieser Zeitachse sollen Rückschlüsse auf wichtige zeitliche Muster, wie beispielsweise Periodizitäten, möglich sein. In den folgenden Abschnitten soll auf das Konzept der zeitachsenbasierten 3D- Visualisierung eines Parallelen Koordinaten Diagramm eingegangen werden. Des Weiteren werden die Möglichkeiten der interaktiven Manipulationen, die der Betrachter vornehmen kann, sowie die Wichtigkeit des graphischen Gestaltungsmittel der Farbe, besprochen.

3.2.1 theoretische Grundlage der Idee

Von den in Kapitel 2.5.3 vorgestellten Varianten und Erweiterungen der klassischen Parallelen Koordinaten eignet sich die Methode der *erweiterten Parallelen Koordinaten*[6] , die von Wegenkittl et al. (1997) entwickelt und dokumentiert wurde, für eine zeitachsenbasierte, dreidimensionale Weiterentwicklung. Das von Wegenkittl et al. vorgeschlagene Konzept für dreidimensionale Parallele Koordinaten hat gezeigt, dass das Aufziehen von Flächen zwischen den einzelnen Merkmalsachsen, flexibler und vielfältiger ist, als die polygonalen Linien der herkömmlichen Parallelen Koordinaten (Wegenkittl et al., 1997). Die aufgezogenen Ebenen im dreidimensionalen Raum definieren eine komplexe Oberfläche, die den Datenraum für den Betrachter „topologisch" erfassbar macht. Die erweiterten Parallelen Koordinaten basieren auf den geometrischen Grundsätzen der ursprünglichen Methode von Inselberg (1985).

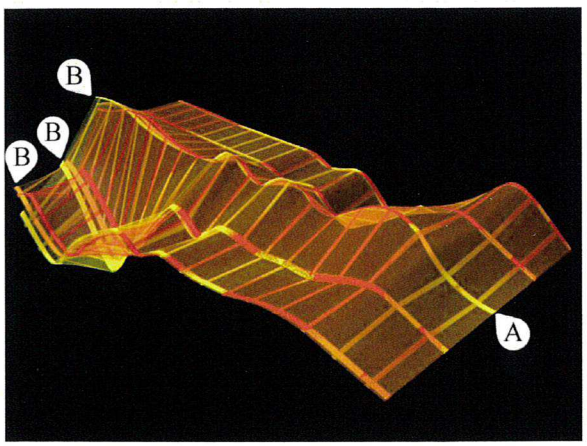


Abbildung 19: Beispiel eines Extruded Parallel Coordinate Plot für einen fünfdimensionalen Datensatz (Quelle: Wegenkittl et al., 1997)

[6] Die erweiterten Parallelen Koordinaten sind in der Fachliteratur als Extruded Parallel Coordinates (EPC) bekannt.

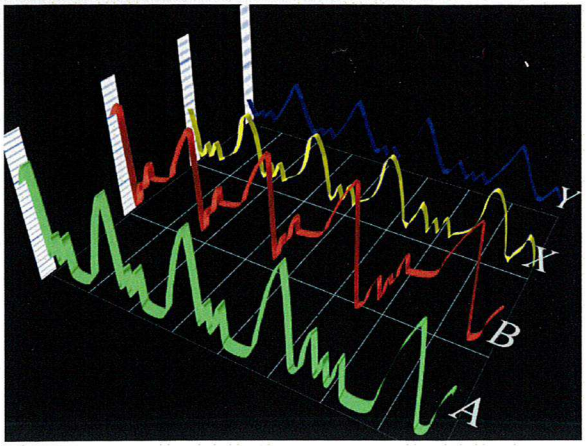


Abbildung 20: erweiterte Parallele Koordinaten für ein vierdimensionales, dynamsiches System (Quelle: Wegenkittl et al., 1997)

3.2.2 geometrische Grundlagen

Wie bereits beschrieben wurde (vgl. Kapitel 2.5), wird ein multidimensionaler Datenpunkt in einem Parallelen Koordinaten Diagramm dargestellt, indem jede Variable auf einer eindimensionalen Merkmalsachse abgebildet wird. Diese zueinander parallelen, nebeneinander liegenden Achsen und die dazugehörigen Punkte darauf, lassen sich zu einer polygonalen Linie verbinden, die einen Beobachtungsfall charakterisiert.

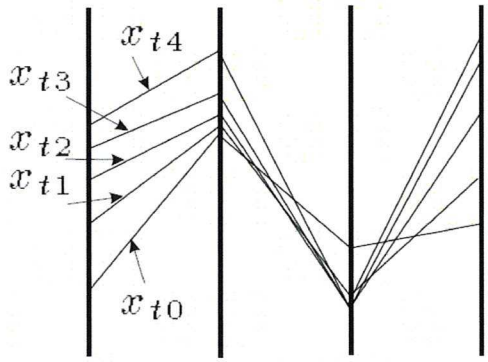


Abbildung 21: Merkmalslinien mit der Variable x zu fünf verschiedenen Zeitpunkten (Quelle: Wegenkittl et al., 1997)

Eine dreidimensionale Darstellung der Parallelen Koordinaten funktioniert nach denselben Prinzipien, mit dem Unterschied, dass anstatt einer Variablen auf einer eindimensionalen Linie, zwei Variablen auf einer zweidimensionalen Ebene abgebildet werden. Die in den Daten enthaltene Information befindet sich so in einem zweidimensionalen Raum. Die auf diese Weise aufgezogene Ebene wird definiert durch vier Punkte: zwei Variablen und zwei unterschiedliche Zeitstände bzw. Zeitpunkte.

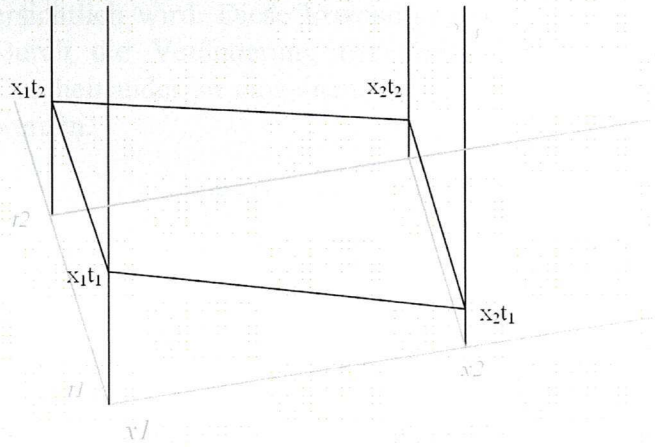


Abbildung 22: aufgezogene Fläche (Quelle: eigene Darstellung)

Durch das Aufziehen vieler solcher Ebenen entsteht eine komplexe, dreidimensionale Oberfläche.

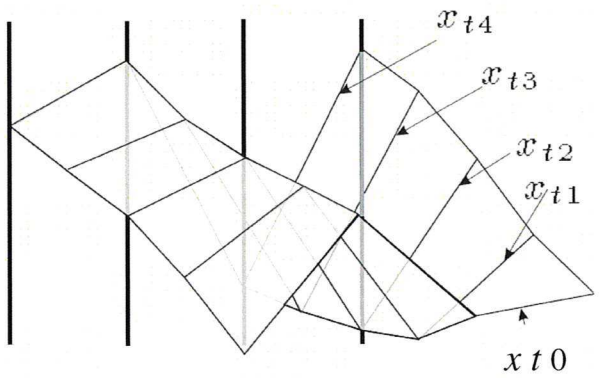


Abbildung 23: dieselben Merkmalslinien wie in Abb. 21 mit dreidimensionalen Parallelen Koordinaten dargestellt (Quelle: Wegenkittl et al., 1997)

Das Grundkonzept, welches Wegenkittl et al. (1997) für dreidimensionale Parallele Koordinaten entwickelt haben, wird ergänzt und weiterentwickelt. Durch ein strukturiertes und übersichtliches Layout (achsenbasierte Visualisierung und Benutzeroberfläche), sowie zusätzlichen Elementen (Attributwerte, Farben) kann die Lesbarkeit der Darstellung für den Betrachter erhöht und die Datenanalyse eine multidimensionalen Datensatzes intuitiver gestaltet werden. Die, in der vorliegenden Arbeit, entwickelte Ergänzung der Methode wird den Anforderungen für eine dreidimensionale Darstellung angepasst. Zudem hält sich die neue Visualisierungsmethode wieder stärker an die herkömmlichen Parallelen Koordinaten, als die von Wegenkittl et al. vorgeschlagene Technik. Dem Betrachter erscheint die Darstellung deshalb vertraut und er kann sich der bereits bewährten Analysewerkzeugen zum Aufdecken von Korrelationen, Ausreissern oder Clustern bedienen. Durch das Hinzufügen einer klar erkennbaren Zeitachse wird der zeitliche Verlauf einzelner Merkmale oder die zeitliche Veränderung eines gesamten Datensatzes visuell erfassbar. Mit der zusätzlichen Dimension erhöht sich die Anzahl der gleichzeitig darstellbaren Merkmale in einem Parallelen Koordinaten Diagramm deutlich. Die Variablen werden auf eine zweidimensionale Ebene abgebildet, wodurch der Datenraum für den Betrachter als „topologische“ Oberfläche

ersichtlich wird. Diese Erweiterung bietet zusätzliche, neue visuelle Analysemöglichkeiten. Durch die Veränderung der Grösse der einzelnen Merkmalsflächen kann das visuelle Durcheinander an einzelnen Stellen im Parallelen Koordinaten Diagramm bewusst reduziert werden.

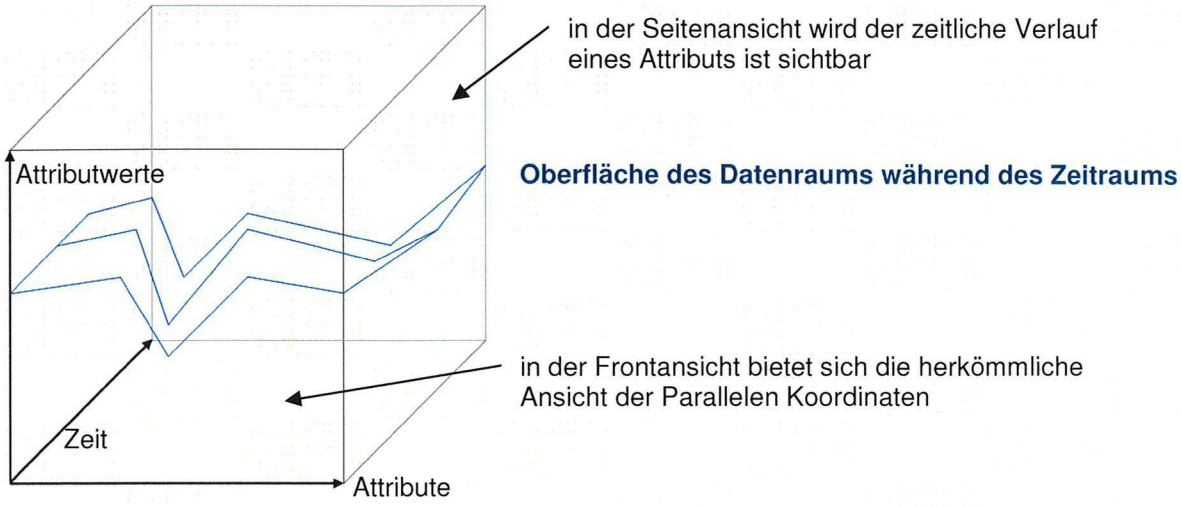


Abbildung 24: Konzept der erweiterten Parallelen Koordinaten für zeitabhängige Daten, (Quelle: eigene Darstellung)

Die Ansicht von vorne (auf die x- und y-Achse) ermöglicht dem Benutzer des erweiterten 3D Parallelen Koordinaten Diagramms die herkömmliche Betrachtungsweise der Parallelen Koordinaten. Die visuelle Datenanalyse erfolgt ebenfalls auf die gewohnte Weise. Abhängigkeiten eines Merkmals in einer Ebene und deren benachbarten Ebenen sind gut erkennbar. Durch die Rotation des gesamten Graphen ist es bei der dreidimensionalen, erweiterten Parallelen Koordinaten nun möglich auf die z- und y-Achse zu blicken. Auf diese Weise sind für den Betrachter die zeitliche Entwicklung eines Merkmals und das zeitliche Verhalten aller Merkmale des Datenraums visuell erfassbar. Vorhandene zeitliche Muster können aufgedeckt und untersucht werden.

3.2.3 Interaktion

Um das Konzept für eine effektive Datenexploration zu gestalten, sind verschiedene Interaktionstechniken unverzichtbar. Die Aufgabe einer Interaktionstechnik ist es, dem Benutzer indirekt und intuitiv die Beweglichkeit im Datenraum zu ermöglichen, die er sich wünscht und die ihm hilft eine effiziente Datenanalyse durchzuführen.
Die Interaktivität wird von Crampton (2002) als „ein System, das seine visuellen Daten nach einer Benutzerabfrage darlegt" definiert. Die interaktiven Möglichkeiten, mit denen sich ein Anwender in einer Visualisierung bewegen kann, werden in verschiedene Grade eingeteilt. Die Beleuchtung, der Standpunkt (von 3D-Anwendungen), zoomen, Massstabsänderungen und Änderungen der Symbole oder Elemente stellen einen niedrigen Grad der Intensität dar, während die Navigation in der Anwendung, das Ein- und Ausschalten von Ebenen, Sortierung und Flüge einem höheren Grad an Interaktivität zugeordnet werden. Der höchste Grad an Interaktivität lässt ein Abfragen in der Datenbank, Filterung von Objekten, Kombinieren von

Datenebenen oder mehreren, parallelen Fenstern zu (Williams, 2005). Für das in der vorliegenden Arbeit entworfene Konzept genügen Interaktionstechniken niederen Grades.

3.2.4 Farbwahl

Um das Konzept einer dreidimensionalen Erweiterung der Parallelen Koordinaten für Daten mit einem Zeitbezug zu vervollständigen, muss das graphische Gestaltungsmittel der Farbe gezielt eingesetzt und verwendet werden. In zahlreichen Studien wurde aufgezeigt, dass Farbe zu einer Reduktion visueller Suchzeit des Betrachters in einer Darstellung führt (vgl. Carter, 1982; Treisman u. Gormican, 1988). Diese Studien, sowie die empirische Studie von Keller u. Grimm (2005), die zum Schluss gekommen ist, dass Farbe auch in 3D-Visualisierungen zu einer verbesserten Lesbarkeit beiträgt, zeigt, dass dieser visuellen Variable in einem erweiterten Parallelen Koordinaten Diagramm Beachtung geschenkt werden muss. Um die optimale Wirkung von Farbe beim Darstellen der Daten zu erzielen muss auf die richtige Kombination und Verwendung geachtet werden. Dabei gelten die Grundsätze:

- Nicht mehr als sieben Farben verwenden (Miller, 1956)
- Komplementärfarben verwenden
- Gleiche Farbe für gleiche Sachverhalte, konsequente Farbgestaltung
- Auslösende Gefühle beachten, Farbbelegung
- Geeignete Farbskala verwenden
- Farbsinnstörungen beachten (z.B. Grünschwäche)

Die Auswahl einer optimalen Farbskala ist nicht abschliessend definierbar und immer noch offener Forschungsgegenstand. Ruth (2003) fasst die verschiedenen Einflussfaktoren auf die Farbwahl zusammen (vgl. Tabelle 1). Die obligatorischen Aspekte müssen bei einer Farbgebung in jedem Fall berücksichtigt werden.

Obligatorische Aspekte	Fakultative Aspekte
Wertebereich der Daten	Farbe und Grösse
Visualisierungsaufgabe	Zusammenspiel mehrerer Farben
Visuelles Auflösungsvermögen	Zeitliche Bedingtheit der Wahrnehmung
	Adaptionsmechanismen
	Farbsehschwächen
	Einschränkungen der Farbwiedergabe
	Kulturelle Assoziationen

Tabelle 1: Einflussfaktoren zur Nutzung von Farbskalen (Quelle: eigene Darstellung nach Thomas, 2003)

Die graphische Variable Farbe ist ein sehr wichtiges Mittel zur Visualisierung von Merkmalen, da Farbe vom Menschen intuitiv und ohne kognitive Belastung des Betrachters wahrgenommen und interpretiert werden. Sie kann in der Visualisierung zur Betonung oder Unterscheidung von Elementen, zur Darstellung von Strukturen in der Natur oder realistischen Objekten, zur Repräsentation von symbolischen Assoziationen oder zum Ausdruck von Stimmungen verwendet werden (Schuman u. Müller, 2000: 147). Farbe wird automatisch zusammen mit der Information zu einem Objekt im Langzeitgedächtnis abgespeichert (Hanna u. Remington, 1996). Da der Farbraum an sich multidimensional ist, eignet sich die Variable Farbe für multidimensionale Datenvisualisierungen und ermöglicht einer Darstellung einen zusätzlichen Freiheitsgrad. Für die Visualisierung eines erweiterten Parallelen Koordinaten Diagramms wird die Farbe als Betonung und Unterscheidung von Elementen und Regionen verwendet.

Bei der Farbgebung einzelner Elemente in einer graphischen Darstellung gilt es ebenfalls, die wichtigsten ästhetischen Regeln einzuhalten. Leider stehen diese oft im Widerspruch zu den Richtlinien, die eine gute Farbunterscheidung bzw. Selektion ermöglichen. Bertin (1974: 98) bezeichnet „... Farbzusammenstellungen als helligkeitswert-harmonisch, in denen nur eine Farbe mit unterschiedlichen Helligkeitswerten vorkommt“ und als „... farb-harmonisch, jene Farb-Kompositionen, in denen nur Farben auftreten, die im Farbkreis nebeneinander liegen“. Erst wenn diese „Harmonie-Regeln“ bei einer Visualisierung beachtet werden, empfindet der Betrachter sie auch als ästhetisch. Um ein harmonisches Farbkonzept zu erstellen können also Farben der gleichen Farbstufe oder die Kombination verschiedener Tonstufen der gleichen Farbe verwendet werden.
Für die Farbkodierung stehen verschiedene Farbräume zur Verfügung. Die in der Computergraphik am häufigsten verwendeten Farbmodelle sind der RGB-Farbraum[7] (Klawonn, 2009) und das HSV-Farbmodell[8]. Für den RGB-Farbraum ergeben sich, durch die Kodierung eines Bytes für eine Farbe, 256 Intensitätsstufen für jede der drei Primärfarben. In grösseren Speichermedien sind bis zu 16 Bit pro Farbe kodierbar, wodurch die Darstellung von über 281 Billionen Farben ermöglicht wird. Das HSV-Farbmodell definiert eine Farbe nicht über Grundfarben, sondern mittels drei Parameter. Obwohl dieses Modell aufgrund seiner relativ intuitiven Farbwahrnehmung, gerne für Visualisierungen am Bildschirm verwendet wird, bietet sich aus implementatorischer Sicht für die Entwicklung des hier vorgestellten Prototyps, der RGB-Farbraum an. Das RGB-Farbmodell lässt sich einfach umsetzen und wurde speziell für farbdarstellende Systeme auf dem Bildschirm entwickelt.
Bei der Darstellung von numerischen Werten ist das Skalenniveau der Werte von Bedeutung (Steinmetz, 2008). Um dem Betrachter der Visualisierung die Unterscheidung zwischen den einzelnen Werten zu ermöglichen, sollen Farben gewählt werden, deren Abstände im RGB-Farbraum maximal zueinander sind. Da in der Darstellung des hier entworfenen Konzepts vor allem die Unterscheidung der einzelnen Merkmale im Vordergrund steht, muss eine Farbskala gewählt werden, die maximale Farbunterschiede garantiert. Für die praktische Umsetzung wird eine Farbskala mit zwölf Farben bestimmt (vgl. Abbildung 25).

Abbildung 25: Farbskala mit zwölf Abstufungen (Quelle: eigene Darstellung)

[7] R= Rot, G=Grün, B= Blau
[8] H= Hue (Farbton), S= saturation (Sättigung), V= value (Helligkeit/Intensität)

3.2.5 Navigation

Von *Navigation* wird gesprochen, wenn durch Drehung, Verschiebung oder Skalierung der 3D-Visualisierung, keine Änderung an deren Struktur erfolgt. Die relativen Positionen und Farben aller in der Graphik enthaltenen Geometrien und den dazugehörigen Attributen bleiben erhalten (Holzhüter, 2005: 32). Die Erfassung von Objekteigenschaften durch 3D-Navigation ist für den Menschen ein intuitiver und natürlicher Vorgang. Durch das Drehen und Betrachten eines Gegenstandes zieht der Mensch Rückschlüsse auf die Eigenschaften des Objekts und ordnet es in einen Kontext ein. 3D-Objekte können als Ganzes erfasst werden, ohne alle Seiten eines des Objektes gleichzeitig sehen zu müssen. Die Navigation im Prototyp soll der Benutzer durch das Bewegen und Bedienen eines Eingabegerätes, z.B. der Maus, vornehmen können.

4 Software und Programmiersprache

Um das Konzept einer zeitachsenbasierten 3D-Erweiterung der Parallelen Koordinaten praktisch umzusetzen wird in der vorliegenden Arbeit die Java Softwareerweiterung Java3D verwendet. Das folgende Kapitel erläutert warum diese Programmiersprache verwendet wird und zeigt kurz deren Funktionsweise auf.

4.1 Java 3D

Der Funktionsumfang und die Anwendungsmöglichkeiten von Java3D sind vielfältig. Java3D zeichnet hauptsächlich sich aus durch eine

- hohe Performanz,
- Plattformunabhängikeit,
- grosse Anzahl an vordefinierten 3D-Fähigkeiten,
- Objektorientiertheit und
- ein hoher Abstraktionsgrad aus.

Java3D unterstützt viele der Anforderungen und theoretischen Konzepte die in den vorangegangenen Kapiteln vorgestellt wurden und eignet sich deshalb gut für die Implementierung dreidimensionaler Visualisierungen. Ebenfalls bietet Java3D die Möglichkeit, mit relativ geringem Aufwand Applikationen zu programmieren, die sowohl als Internet Applets, als auch als eigenständige Applikation gestartet werden können.

4.1.1 Entwicklungshintergrund

Java3D wurde von SunMicrosystems entwickelt um deren Programmiersprache Java 3D-fähig zu machen. Ziel war es, ein Programmierinterface zu verwirklichen, das es erlaubt, auf einfache Art und Weise 3D-Anwendungen zu entwickeln, die auf jedem Java2-fähigen Betriebssystem laufen. Eine erste Spezifikation von Java3D wurde 1997 auf der SIGGRAPH97[9] vorgestellt. Seither wurde die Technologie bis zur aktuellen Version 1.5.0 stetig weiterentwickelt und ist keineswegs abgeschlossen. Java3D ist ein optionales, Open-Source Paket welches nicht zum offiziellen Java SDK- oder JRE-Release gehört und deshalb manuell ins Java-System eingebunden werden muss.

Java3D verwendet und vereinigt sowohl die bewährten Konzepte bestehender Low-Level-Rendering-APIs[10], wie beispielsweise OpenGL oder Direct3D, als auch von Graphiksystemen auf höherer Abstraktionsebene wie VRML[11] oder OpenInventor. Diese High-Level-APIs

[9] International Conference on Computer Graphics and Interactive Techniques, Los Angeles
[10] engl: für application programming interface
[11] Virtual Reality Modeling Language (VRML)

machen Java3D zu einer modernen, objektorientierten Anwendung. Wie VRML ist Java3D nicht nur in der Lage 3D-Objekte darzustellen, sondern es können auch Sound, Verhalten, Animationen und unkonventionelle Eingabe- und Ausgabegeräte gesteuert und eingebunden werden (Java, 2010). Die Abbildung 26 stellt die Einbettung der Java3D API in der Hierarchie der Java2 Plattform dar. Eine Java3D Anwendung wird auf der Basis der Java Version 1.2 (oder höher) und einem Low Level Treiber (*DRV* - wie zum Beispiel OpenGL oder DirectX) sowie der Verwendung des *JDK* (Java Development Kit) entwickelt. Durch die *JSDT API* (Java Shared vanced Imaging), *JMF API* (Java Media Framework) und die *Java Sound API* stehen dem Entwickler weitere Möglichkeiten zur Verfügung (Valentin, 2005: 9).

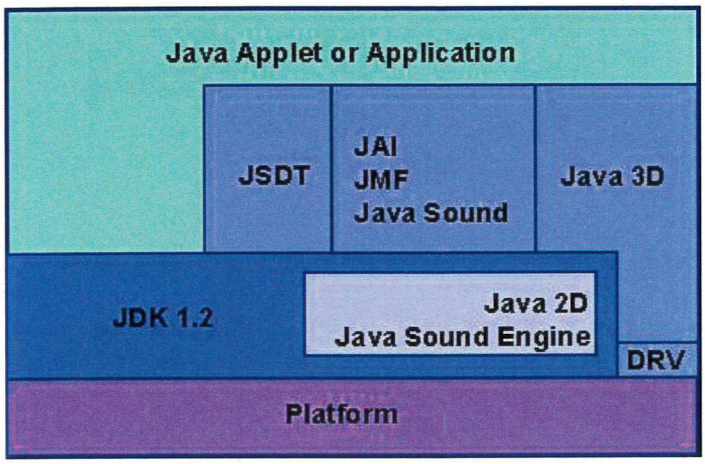


Abbildung 26: Aufbau der Java Plattform (Quelle: Valentin, 2005)

4.1.2 Java3D und seine Komponenten

Java3D besitzt eine umfangreiche Klassenbibliothek, welche verwendet wird um dreidimensionale Graphiken innerhalb von Java-Applikationen und Applets zu unterstützen. Diese Klassen ermöglichen es, komplexe, dreidimensionale Körper zu erzeugen. Sie stellen die Strukturen zur Verfügung um die benötigten Geometrien darzustellen und zu bearbeiten. Durch die umfangreichen, vordefinierten Klassen kann das Konstruieren von virtuellen Welten, Interaktionen und Darstellungen verschiedenster multimedialer Komponenten (Animation, Sound etc.) und komplexen Datenvisualisierungen mit einem relativ geringen Aufwand realisiert werden.

Da Java eine objektorientierte Programmiersprache ist, besteht auch die Java3D API aus diversen Objekten oder Klassen, die in Paketen zusammengefasst sind. Den Kern stellt dabei das `javax.media.j3d` Paket dar. Es liefert alle benötigten Klassen und Methoden, um anspruchsvolle dreidimensionale Inhalte zu erstellen. Das zweite wichtige Paket beinhalten die Utility-Klassen (`com.sun.j3d.utils`), welche die Kernklassen um weitere nützliche Funktionen erweitern. Die Utility-Klassen können in vier Kategorien unterteilt werden: *Content loaders*, *scene graph construction aids*, *geometry classes* und *convenience classes*. Durch diese Utility-Klassen kann der Programmieraufwand deutlich reduziert werden (Fachhochschule Wedel, 2000).

4.1.3 Projektionsfläche

Wie bereits erwähnt wurde, spielen geometrische Transformationen in der dreidimensionalen Computergraphik eine wichtige Rolle. Der, durch die drei räumlichen Koordinatenachsen aufgespannte Raum, wird in der Nomenklatur von Java3D *virtuelles Universum* genannt.

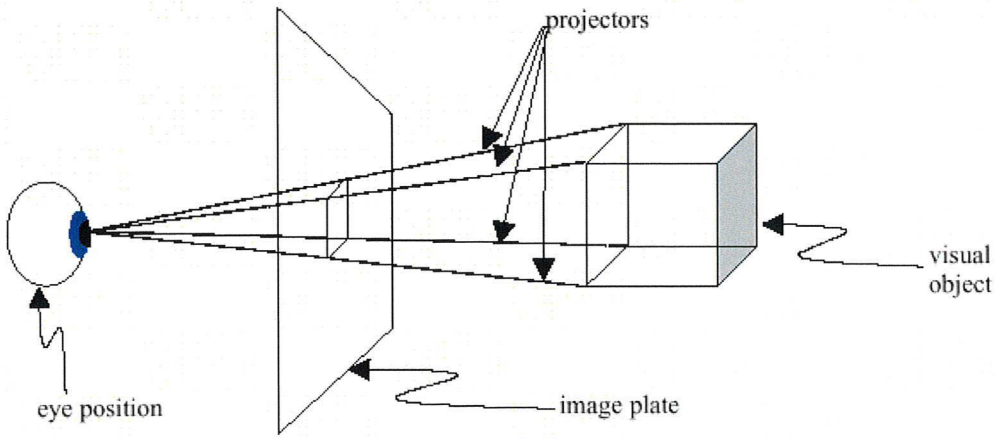


Abbildung 27: Schema der Beziehungen zwischen Betrachter , Projektionsfläche und Szenenkamera in Java3D (Quelle: Heinzel, 2003)

Die Abb. 27 zeigt die Beziehungen zwischen Projektionsfläche, Szenenkamera und den darzustellenden 3D-Objekten (Heinzel, 2003). Vor der Projektionsfläche („image plate") befindet sich die Szenenkamera bzw. der Blickpunkt, von dem ein Betrachter die 3D-Szene sieht. Bei Java3D ist die Szenenkamera standardmässig auf die Position (0,0,2.41) gesetzt. Die visualisierten 3D-Objekte liegen hinter der Projektionsfläche und werden durch eine geeignete Projektion darauf abgebildet. Das Zentrum der Projektionsfläche (0,0,0) stimmt mit dem Zentrum des virtuellen Universums überein. Java3D definiert ein rechtshändiges, kartesisches Koordinatensystem. Vom Zentrum aus betrachtet erscheint dem Betrachter die x-Achse als horizontale Gerade mit positiven Werten nach rechts. Die positive z-Achsenrichtung ist dem Betrachter entgegengerichtet. Die y-Achse ist als vertikale Gerade ersichtlich, deren positive Werte nach oben verlaufen. Die Orientierung des Koordinatensystems muss insofern beachtet werden, da rechtshändige Koordinatensysteme die Eigenschaft besitzen, Raumpunkte eines Polygons entgegen dem Uhrzeigersinn darzustellen.

4.1.4 Das SzenenGraph-Modell

Der Szenegraph ist ein Konzept zur Darstellung eine dreidimensionale Szene und das zentrale Element der Java3D API. Eine Java3D Applikation wird als `VirtualUniverse` bezeichnet. Pro Anwendung darf davon nur eine Instanz existieren. Ein solches virtuelles Universum ist eine, aus vielen einzelnen, dreidimensionalen Objekten erzeugte, virtuelle Welt auf einem hohen Abstraktionsgrad. Diese Objekte des Java3D-Universums werden in einen Szenegraphen eingebunden, in dem alle Informationen über die Objekte abgespeichert werden. Mit Hilfe des Szenegraphen wird der Aufbau der Darstellung definiert. Er besteht aus den Elementen welche die Geometrie, den Sound, die Orientierung, die Beleuchtung, die Positionierung und das Erscheinungsbild der visuellen und Audio-Objekte bestimmen. Der logische Aufbau des virtuellen Universums wird in einer azyklischen, gerichteten

Baumstruktur abgebildet, die im Wesentlichen aus Definitionen von Transformationen und Geometriedaten besteht. Die so strukturierte Sicht der Szene erlaubt eine komfortable Handhabung der Objekte. Abbildung 28 veranschaulicht das Konzept des Szenegraphen und zeigt seine wesentlichen Bestandteile.

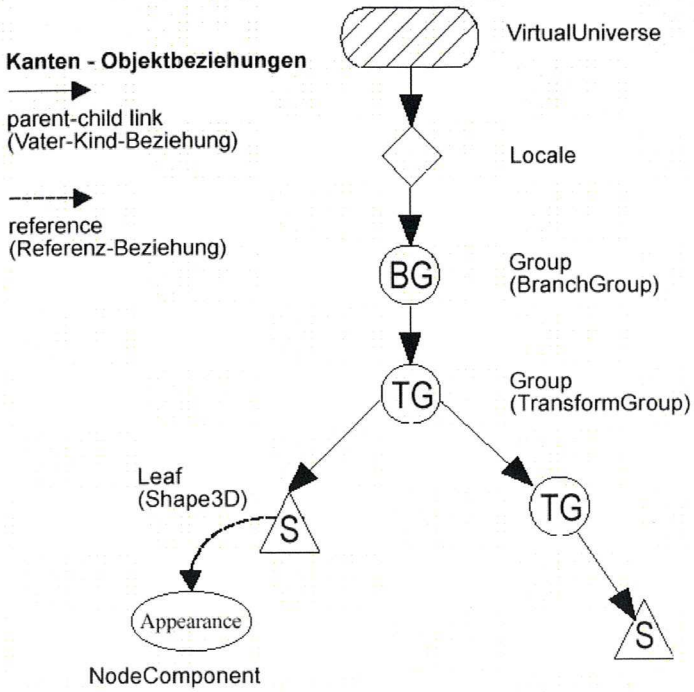


Abbildung 28: Aufbau Szenegraph (Quelle: Java3D_schnell_tutorial.pdf)

Eine graphische Darstellung wird definiert über eine Datenstruktur aus Knoten und Kanten. Ein Knoten ist ein Datenelement, während eine Kante immer eine Beziehung zwischen den Datenelementen beschreibt (vgl. Java, 2003). Die Knoten im Szenegraphen sind die Instanzen der Java3D Klassen und die Kanten repräsentieren die Beziehung zwischen diesen Instanzen. Eine Kante kann auf zwei unterschiedliche Arten eine Beziehung zwischen zwei Knoten herstellen:

- Vater-Kind-Beziehung (parent-child link)
- Referenz Beziehung (reference)

Diese Knotenkomponenten werden beim Darstellen eines visuellen Objektes verwendet um die Geometrie und das Erscheinungsbild zu definieren. Wie in Abb. 28 ersichtlich, befindet sich unterhalb des virtuellen Universums (`VirtualUniverse`) das „`Locale`"-Objekt. Durch dieses Objekt wird der Ursprung bzw. der Nullpunkt des Universums über ein sehr präzises Koordinatensystem definiert. Da die Koordinaten als 256-Bit Festkommazahlen implementiert sind, ist es möglich Objekte aller Grössenordnungen exakt zu berechnen und darzustellen, was die Nutzung der Java3D API für wissenschaftliche Anwendungen so interessant macht. An die `Locales` können Teilbaumgruppenknoten (`BranchGroup Nodes, BG`) angehängt werden. Es werden zwei Arten von Teilbaumgruppen unterschieden. Der Inhalts-Teilbaumgraph (`contentbranchGraph`) und der Ansichts-Teilbaumgraph (`viewbranchGraph`). Der Inhalts-Teilbaumgraph spezifiziert die 3D-Szene mit all ihren visuellen Attributen, den Transformationen, der Geometrie, der Beleuchtung und den

Animationen. Die Parameter der Szene, welche die Position und die Szenenkamera beschreiben, werden vom Ansichts-Teilbaumgraph bestimmt. Ein Subknoten der Teilbaumgruppen wird Transformationsgruppe (`TransformGroup, TG`) genannt, an den wiederum beliebig viele weitere Knoten angehängt werden können. Diese angehängten Objekte sind in Gruppen und Blätter (`Leafs`) unterteilt. Zu den Blattknoten bestehen keine weiteren Untergruppen. Alle grafischen Objekte (Kugel, Würfel, Pyramide etc.) sind eine Ableitung eines Blattknoten. Über ihn werden beispielsweise die Attribute eines 3D-Objektes beschrieben.

4.2 Aufbau und Implementierung des Programms

In diesem Abschnitt soll kurz aufgezeigt werden, wie das Grundgerüst der in dieser Arbeit entwickelten und beschriebenen Visualisierung durch Java3D aufgebaut ist und implementiert wurde. Das Programms entspricht dem Aufbau des Szenegraphen (vgl. Kapitel 4.1.4). Wie die gesamte 3D-Darstellung strukturiert ist und visualisiert wurde, zeigt Abb. 29. Aus einem Beispieldatensatz werden die numerischen Attributwerte aufbereitet und extrahiert. Die Implementierung eines dreidimensionalen Koordinatensystems mit Koordinatenachsen und mehreren parallelen Merkmalsachsen, sowie die Darstellung der Merkmalslinien und Flächen, erfolgt mittels geometrischer Objekte in Java3D. Ein Benutzerinterface vereint die einzelnen 3D-Element und die Navigations- und Interaktionswerkzeuge.

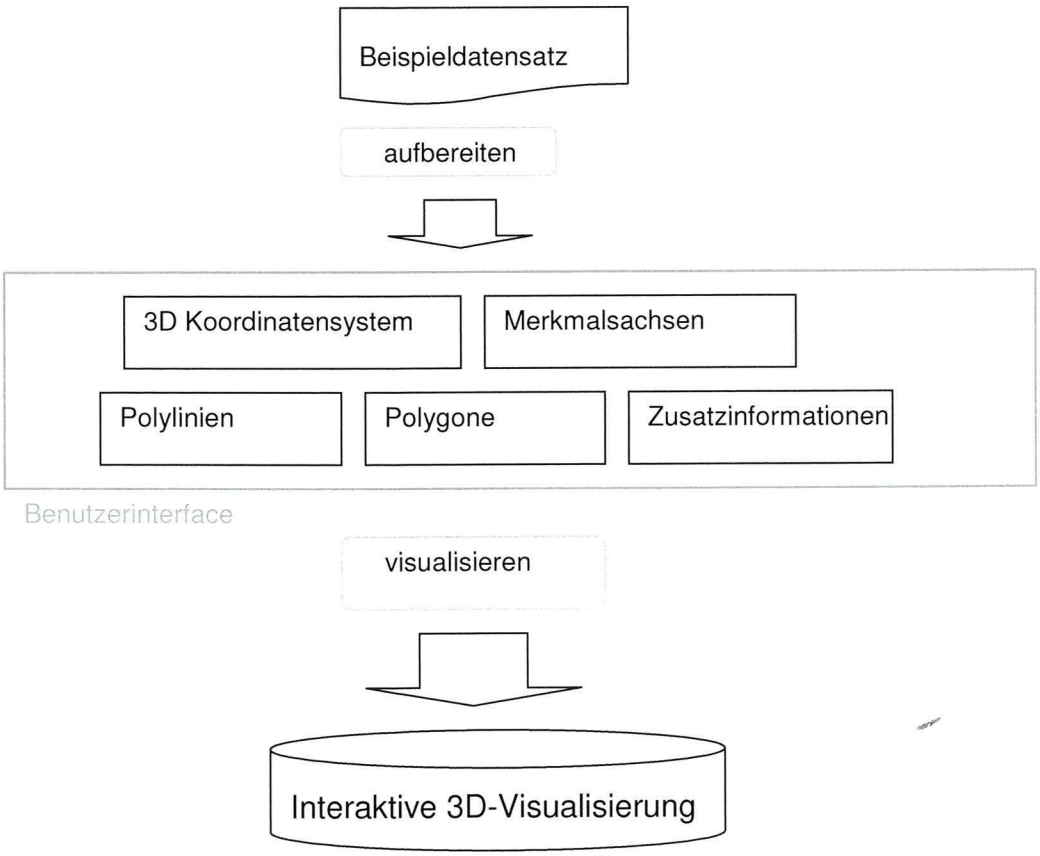


Abbildung 29: Ablauf der Implementierung (Quelle: eigene Darstellung)

5 Visualisierungsdesign des Prototyps

Die im Kapitel 3 erläuterten Grundlagen und das vorgestellte Konzept können nur durch eine praktische Umsetzung überprüft werden. Erst durch die Anwendung und Umsetzung wird die tatsächliche Einsetzbarkeit und Leistung einer visualisierten Zeitachse in dreidimensionalen Parallelen Koordinaten ersichtlich. Ebenfalls können wichtige Einflussfaktoren auf die graphische 3D-Darstellung, wie die Farben der Merkmalsflächen, die Belichtung und Schattierung der Szene, die Navigation in der Graphik oder die Anordnung der Merkmale auf den Achsen erst während einer Implementierung ermittelt werden. In den folgenden Abschnitten wird das Design der 3D Erweiterung für die Parallelen Koordinaten entwickelt und beschrieben. Alle Einflussfaktoren, die für eine angemessene Visualisierung beachtet werden müssen und in den vorangegangenen Kapiteln (vgl. 3.2) bereits besprochen wurden, werden hier noch einmal problembezogen aufgegriffen und hinsichtlich der Implementierung spezifiziert. Die Auswahl geeigneter Techniken oder das Treffen von Entscheidungen für die Darstellung bestimmter Elemente wird begründet und aufgezeigt.

5.1 Kriterien und Rahmenbedingungen

Für ein gelungenes Visualisierungsdesign müssen zahlreiche Aspekte und Kriterien berücksichtigt und eingehalten werden. Die Anforderungen, welche die Qualität einer Visualisierung bestimmen, wurden im Kapitel 2.2.2 erläutert. Diese Anforderungen bilden auch die Grundlage für den Entwurf des Visualisierungsdesigns des Prototyps. Gemäss Lange et al. (2006) muss berücksichtigt werden, dass das Visualisierungsproblem im Wesentlichen durch *Gegenstand*, *Ziel* und *Kontext* der Visualisierung beschrieben wird. Beim Entwurf wurde also immer die Frage nach dem *Was*, *Warum* und *unter welchen Rahmenbedingungen* wird die Darstellung realisiert, gestellt. Zu den Rahmenbedingungen gehören auch die kognitiven Eigenschaften des Benutzers, dessen Erfahrungen und Präferenzen, sowie der allgemeine Anwendungshintergrund (z. B. etablierte Techniken und Werkzeuge) (Lange et al., 2006). Eine weitere Rahmenbedingung stellt die Leistungsfähigkeit der verwendeten Hard- und Software dar.

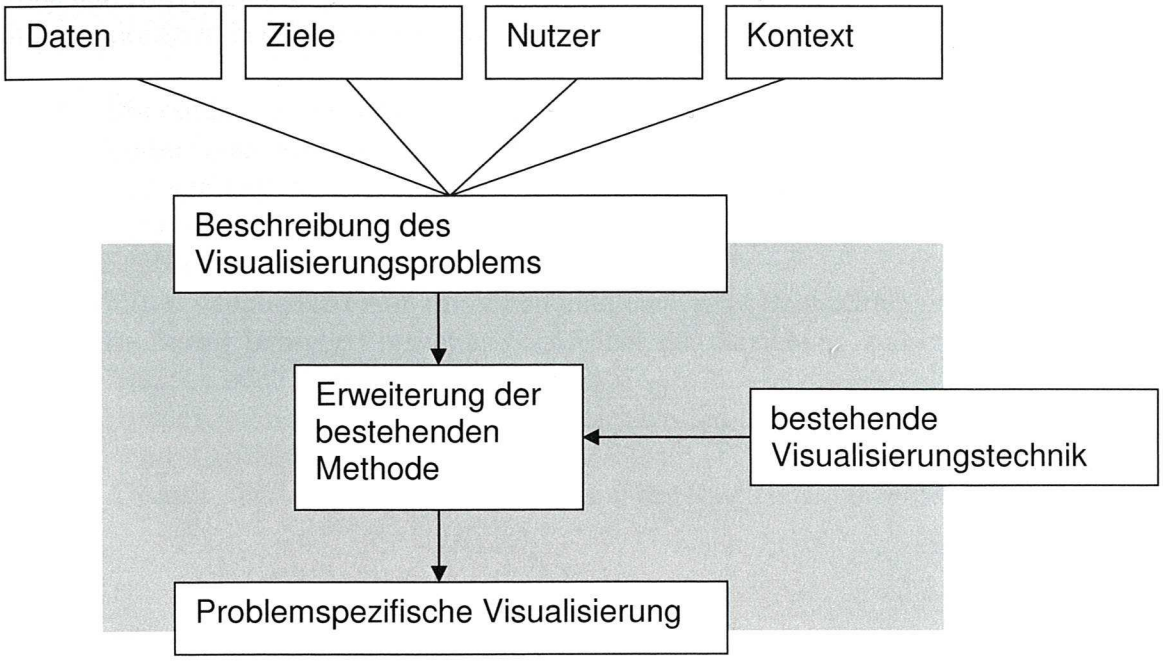


Abbildung 30: Die Auswahlentscheidung für die Entwicklung einer problemspezifischen Visualisierung (Quelle: eigene Darstellung frei nach Lange et al., 2006)

Für die Realisierung und Implementierung einer Visualisierung muss zunächst das bestehende Visualisierungsproblem beschrieben werden (vgl. Abbildung 30). Für das Konzept in der vorliegenden Arbeit, wird berücksichtigt werden, dass es sich um multidimensionale Daten mit einem zeitlichen Attribut handelt, die im Kontext der visuellen Datenanalyse mittels der Visualisierungstechnik der Parallelen Koordinaten, in einer dreidimensionalen Visualisierung dargestellt werden sollen. Diese Parameter und die Eigenschaften der herkömmlichen Visualisierungsmethode fliessen in das Konzept für die Erweiterung mit ein. Auf diese Weise entsteht eine problembezogene, spezifische Visualisierung.

5.2 Interaktion im Prototyp

Wie bereits erwähnt wurde, muss eine gute, interaktive 3D-Visualisierung, dem Benutzer indirekt und intuitiv die Beweglichkeit im Datenraum ermöglichen, die er sich wünscht und die ihm hilft eine effiziente Datenanalyse durchzuführen (vgl. Kapitel 3.2.3). Für eine effektive Datenexploration sind Interaktionstechniken unverzichtbar. Generell kann gesagt werden, dass ein gutes Benutzerinterface nach dem Prinzip *„overview first, zoom and filter, then details-on-demand"* aufgebaut ist (Shneiderman, 1996).

Für den Betrachter einer dreidimensionalen Visualisierungsapplikation existiert eine Vielzahl an interaktiven Techniken, die ihm ermöglichen den Datenraum zu erkunden. Dieser Abschnitt soll einen Überblick darüber ergeben, welche Techniken bei einer dreidimensionalen Erweiterung der Parallelen Koordinaten um die Zeitachse möglich sind und umgesetzt werden können. Shneiderman (1996) führt diesen Grundsatz weiter und fügte

folgende Aufgaben, die ein Interface unterstützen soll, um dem Benutzer den Zugang zu den Informationen in den Daten zu ermöglichen, hinzu:

- **Übersicht**: eine Gesamtübersicht über den Datenraum gewinnen
- **Zoom**: eine Ansicht einer einzelnen Beobachtung oder eines Merkmals vergrössern oder verkleinern
- **Detail-on-demand**: Detailinformation zu einer ausgewählten Gruppe, einer Beobachtung oder einem Merkmal erhalten
- **Filter**: Möglichkeit zum Entfernen uninteressanter Beobachtungsfälle
- **In Bezug bringen** (Beziehung): Ansicht der Beziehung zwischen zwei selektierten Gruppen oder Merkmalen ermöglichen
- **History**: Aktionen des Benutzern speichern um ein Wiederholen oder Zurückspielen zu ermöglichen
- **Auszug**: der Auszug einer Subkollektion der Daten oder einzelner Parameter erlauben

5.3 Navigation

Die interaktiven Navigationswerkzeuge eines Benutzerinterfaces sollen durch eine intuitive Handhabung, den schnellen Erschliessungsprozess der Daten unterstützen. Ein theoretisches Modell zur Beschreibung des Navigationsprozess schlägt Ware (2004: 326) vor. Der *Navigation control loop* erläutert den kognitiven Prozess, der durch eine Navigation des Benutzers durch den Datenraum erfolgt (vgl. Abb. 31).

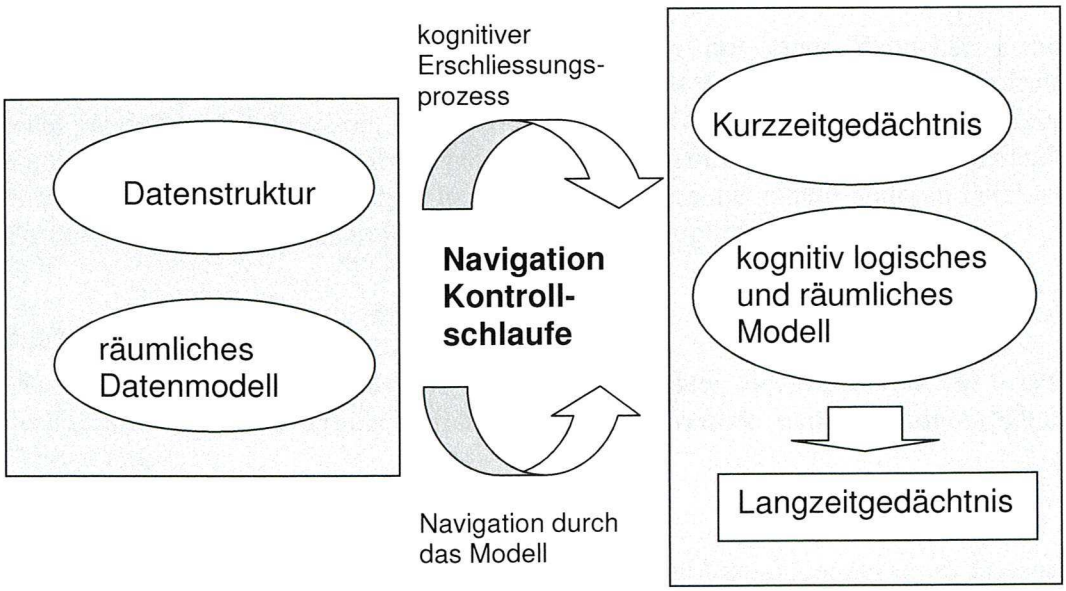


Abbildung 31: Navigation control loop (eigene Darstellung nach Ware, 2004)

Durch die Visualisierung eines Datensatzes kann der Benutzer im Kurzzeitgedächtnis ein kognitives Modell der Datenstruktur bilden. Durch einen fortschreitenden kognitiven Erschliessungsprozess mit den ihm zur Verfügung stehenden Navigationswerkzeugen, kann der Benutzer das Datenmodell mental repräsentieren und schliesslich im Langzeitgedächtnis speichern (vgl. Ware, 2004: 326).

Gemäss Ware (2004) wird eine intuitive Handhabung der Daten möglich, indem die Navigation in der virtuellen Darstellung, den Bewegungsabläufen in der realen Welt angeglichen werden. Ware (2004: 327) unterscheidet vier Klassen solcher Navigationsmethapern:

- **World-in-hand**: Der Benutzer kann die 3D-Welt greifen und um alle Achsen drehen.
- **Eyeball-in-hand**: Der Benutzer blickt von einer festgelegten Position in alle möglichen Richtungen.
- **Walking**: Der Benutzer bewegt sich frei durch den zweidimensionalen Datenraum.
- **Flying**: Der Benutzer bewegt sich frei durch den dreidimensionalen Datenraum.

Um den Benutzer der hier erweiterten Methode vor einem Orientierungsverlust in der dreidimensionalen Szene zu schützen, werden die Navigationsmöglichkeiten eingeschränkt. Der im Rahmen dieser Forschungsarbeit entwickelte Prototyp unterstützt die „*world-in-hand*"-Navigation, sowie die „*eyeball-in-hand*"-Navigation. Auf diese Weise wird dem Benutzer die Möglichkeit geboten, die visualisierte Szene zu greifen und zu drehen. Ebenfalls kann von einem festgelegten Standpunkt aus, jede mögliche Richtung des Datenraumes ersichtlich. In den folgenden Abschnitten werden die zusätzlich umgesetzten Interaktions- und Navigationsmöglichkeiten aufgezeigt und erläutert.

5.3.1 Übersicht

Die Erweiterung der 3D Parallelen Koordinaten mit einer Zeitachse ermöglicht dem Betrachter gleich zu Beginn eine Übersicht über den gesamten visualisierten Datenraum. Mit einer vordefinierten Position, dem Nullpunkt der Java3D Szene, wird gewährleistet, dass die Visualisierung alle Merkmalsflächen und das gesamte Spektrum an Zeitständen darstellt. Auch während der Datenanalyse und dem Betrachten aus einem anderen Blickwinkel ist das Zurücksetzen in eine „Gesamtübersicht" jederzeit möglich.

5.3.2 Zoom

Mit Hilfe der mittleren Maustaste kann der Betrachter die Szene beliebig vergrössern bzw. verkleinern. Dadurch ergeben sich für den Anwender neue Abfragemöglichkeiten und Perspektiven.

5.3.3 Auswählen und markieren

Eine der einfachsten Interaktionsmöglichkeit in einer Visualisierung eines Datensatzes ist das Auswählen und Markieren einer Gruppe von Objekten. Durch das Markieren einer Auswahl kann eine visuelle Analyse und ein Vergleich zu anderen Datenobjekten vorgenommen werden. Das Auswählen und Markieren wird oftmals über die graphischen Variablen der Farbe und der Form realisiert (Wills, 1996).

5.3.4 Translation und Rotation

Java 3D eröffnet dem Entwickler mehrere Möglichkeiten sich mittels Mauseingaben im virtuellen Raum zu bewegen. Für die hier vorgestellte Visualisierung wird die Navigation

deshalb mit Hilfe einer Java3D Klasse implementiert. Das Paket `com.sun.j3d.utils.behaviors.mouse` enthält die Klassen, in welchen die Maus als Input für die Navigation in den Visualisierungen verwendet wird. Es sind Klassen für die Translation (Bewegung parallel zur Projektionsfläche) und die Rotation der gesamten Szene. Der Anwender kann durch die Implementierung dieser Klassen die gesamte Datenvisualisierung beliebig bewegen, verschieben und drehen.

5.4 Designentscheidungen

In diesem Abschnitt werden die Designentscheidungen, welche für die Darstellung und Implementierung der einzelnen Elemente gefällt wurden, beschrieben und begründet.

5.4.1 Merkmalsflächen

In der Erweiterung der Parallelen Koordinaten für Daten mit Zeitbezug, wird die in den Daten enthaltene Information auf zweidimensionale Ebenen transferiert. Eine Ebene im Raum ist durch die Ebenengleichung mit einem Punkt und zwei Richtungsvektoren bestimmbar (vgl. Abbildung 32).

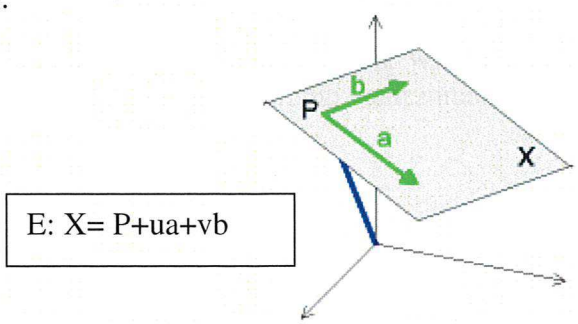


Abbildung 32: Parameterform der Ebenengleichung (Quelle: Gut, 2009)

Auf diese Weise werden die Flächen aufgespannt und im dreidimensionalen Koordinatensystem platziert. Polygone lassen sich mit Java3D implementieren.

Um die Farbtöne einer Oberfläche bzw. einer Fläche zu spezifizieren wird das RGB-Farbmodell verwendet werden. Durch ein Objekt der Klasse `appearance` kann einer Fläche eine Farbe zugeordnet werden. Die Appereance definiert sämtliche Eigenschaften, die das Erscheinungsbild eines 3D-Objektes in Bezug auf sein Material, nicht aber seine geometrische Form, beeinflussen. Mit der Klasse `Color3f` lassen sich über die RGB-Werte mit dem Konstruktor

```
Color3f col = new Color3f(r, g, b);
```

Farbtöne einer Oberfläche spezifizieren. Dabei sind die r, g, b entweder drei float- Werte zwischen Null und Eins, die den Rot-, Grün- und Blauanteil der Farbe angeben, oder drei Integerwerte im Bereich von 0 bis 255, die die Intensitätsstufen für die drei Farben bestimmen. Durch diese drei Werte wird eine Farbe im RGB-Modell im dreidimensionalen Farbraum charakterisiert und korrespondiert mit dem Vektor $(r,g,b) \varepsilon [0,1]^3$. Auf diese Weise lassen sich auch Konvexkombinationen, mit welchen Farbverläufe realisiert werden können, berechnen (Klawonn, 2009:109).

Durch das Hinzufügen von Transparenz kann in einer dreidimensionalen Darstellung das Problem der Überlappung und Verdeckung reduziert oder gar gelöst werden. Der Benutzer kann durch transparente Flächen hindurch sehen und mehrere Elemente zugleich erfassen.
In Java3D wird die Klasse TransparencyAttributes genutzt, um Transparenz zu definieren. Die Art der Transparenz wird mit der Methode setTransparencyMode festgelegt und der Transparenzkoeffizient mit setTransparency als float-Werte zwischen 0 und 1.

```
float transparencyCoefficient = 0.9f;
```

5.4.2 Koordinatenachsen

Wie bereits weiter oben erwähnt wurde, wird die Orientierungsrichtung der Darstellung von der z-Achse beschrieben und ist im Falle von Java3D rechtshändig. Der Ursprung des SimpleUniverse befindet sich im Punkt (0,0,0). Von diesem Nullpunkt aus betrachtet ist die x-Achse eine horizontale Gerade mit positiven Werten nach rechts und die y-Achse eine vertikale Gerade mit positiven Werten nach oben.

Ein möglicher Ansatz um eine geeignete Farbe für die visualisierten Koordinatenachsen zu finden, liefert Grandis (1986). Er ordnet die wichtigsten Farben nach der Reaktionszeit, die der Betrachter braucht um ein Farbsignal wahrzunehmen:

1. gelb
2. weiss
3. rot
4. grün
5. blau

Aufgrund dieser Aufzählung kann gesagt werden, dass möglichst schnell identifizierbare Elemente in einer Visualisierung, in den hellen Farben gelb oder weiss dargestellt werden sollen. Die Farben grün und blau eignen sich hingegen eher für die Visualisierung von Randinformationen (Schumann u. Müller, 2000: 88). Da die y-Achse in der vorgestellten Visualisierung Attributwerte darstellt, die vom Betrachter als Zusatzinformation abgefragt werden können, wurde ihr die Farbe blau zugeordnet. Für die z-Achse, welche je nach Visualisierungsbeispiel die Zeitabhängigkeit der Daten oder die Merkmale selbst beschreibt, wurde die Farbe rot verwendet. Rot lenkt die Aufmerksamkeit des Betrachters auf sich und lässt sich im dreidimensionalen Raum, vor weissem Hintergrund, gut erkennen. Diese beiden Koordinatenachsen wurden mit zwei verschiedenen Farben dargestellt, damit der Benutzer auf den ersten Blick den Unterschied zwischen Zeit- und Attribut-Achse erfassen kann.

Da die Koordinatenachsen keine anderen geometrischen Objekte und Element überdecken, werden sie ohne Transparenz dargestellt.

5.4.3 Text

Durch das Hinzufügen von Text wird die Lesbarkeit von Darstellungen verbessert. Relevante Informationen, wie beispielsweise numerische Werte der Merkmale, Zeitstände, Legenden oder Zusatzinformation, können so an einer Stelle positioniert werden, wo sie der Betrachter benötigt. Beim Hinzufügen von Schrift ist darauf zu achten, dass serifenlose Schriften am Bildschirm besser lesbar sind, als Schriften mit Serifen (Harris, 1999: 398). Auch dünne, fette oder kursive Schriften sollten in einer Visualisierung aufgrund der schwierigen Entzifferbarkeit nicht oder nur sehr sparsam eingesetzt werden. Für die Visualisierung der 3D Parallelen Koordinaten mit Zeitachse wird deshalb der Schrifttyp Arial verwendet.

In einer dreidimensionalen Darstellung kann Text entweder zweidimensional, d.h. flach, oder dreidimensional auftreten. Java3D verwendet Texte u.a. als Inschrift auf Flächen oder als eigene dreidimensionale Buchstabenkörper. Auch zum Beschriften von Variablen in dreidimensionalen Graphiken wird zweidimensionaler Text verwendet. Für die Beschriftung der Visualisierung in der vorliegenden Arbeit eignen sich zweidimensionale Texte aufgrund ihrer guten Lesbarkeit besser als wuchtige, dreidimensionale Buchstabenkörper. Leider hat aber auch die Darstellung von zweidimensionalem Text ihre Nachteile. Sobald der Betrachter die Möglichkeit hat, in der Szene zu navigieren und beispielsweise den Graph zu rotieren um ihn sich von verschiedenen Betrachtungswinkeln anzusehen, treten Probleme auf (vgl. Larson et al., 2000). Zum Einen können beim Herauszoomen die Beschriftungen so klein werden, dass sie nicht mehr lesbar sind, oder die Schrift ist nicht mehr lesbar weil der Betrachter sie von hinten oder der Seite sieht (Klawonn, 2009: 173). Um solche Probleme zu vermeiden müssen auf Beschriftungen geeignete Transformationen angewendet werden. Auf diese Weise kann Text immer zum Betrachter hin orientiert und evt. in der Grösse angepasst werden. Java3D bietet die Klasse Text2D für die Darstellung von zweidimensionalen Texten in einer Szene. Sie bringt den Text in der spezifizierten Farbe als Textur auf ein virtuelles, transparentes Rechteck auf.

```
Text2D text2D1 = new Text2D("1993",new Color3f(0.0f, 0.0f,
1.0f),"Arial", 11, Font.BOLD);
```

5.4.4 Hintergrund

Der Hintergrund der implementierten Erweiterung sollte möglichst ruhig wirken und die Darstellung der Daten hervorheben.

Nach Tufte (1990) müssen sich Hintergrund- und Umgebungsfarben stark von der Farbe des Textes oder der Datendarstellung unterscheiden, um den Lesefluss des Betrachters nicht zu stören. Als Hintergrundfarbe der Darstellungen im implementierten Prototyp wird deshalb ein heller Grauton gewählt. Das Aussehen des Hintergrunds kann für die gesamte Szene global mit dem BackgroundNode über einen RGB-Wert definiert werden.

```
Background hg = new Background();
hg.setColor(new Color3f(1.0f, 1.0f, 1.0f));
```

Für die Darstellung des Hintergrunds wurde keine Transparenz verwendet.

6 Visualisierungsbeispiele

Das Ziel der vorliegenden Arbeit liegt darin, nach neuen Möglichkeiten in der Visualisierung multidimensionaler, zeitabhängiger Daten zu suchen. Die, im Rahmen der Arbeit aufgeworfenen Fragen zur Visualisierung zeitabhängiger Daten durch Parallele Koordinaten, haben nun zu einer dreidimensionalen, zeitachsenbasierten Erweiterung der herkömmlichen Darstellungsmethode geführt. In diesem Kapitel werden die Beispiele, welche mit dem entwickelten Prototyp generiert wurden, vorgestellt und erläutert. Diese Beispiele dienen dazu, die Forschungsfragen, welche zu Beginn der vorliegenden Arbeit formuliert wurden, im Kapitel 7 zu beantworten.

6.1 Benutzerinterface

Die Abbildung 33 zeigt das Benutzerinterface mit dessen Hilfe die Visualisierung eines Datensatzes verändert und gesteuert werden kann. Es wurde darauf geachtet, dass die in Kapitel 2 und 3 formulierten Anforderungen und Konzepte einer Datenvisualisierung eingehalten wurden und entsprechend eingeflossen sind. Das Design des Benutzerinterface ist klar strukturiert und relativ einfach aufgebaut. Das Hauptaugenmerk des Benutzers soll auf dem grosszügig gestalteten Datenfenster in der Mitte des Interfaces, liegen. Hier wird der Datenraum visualisiert.

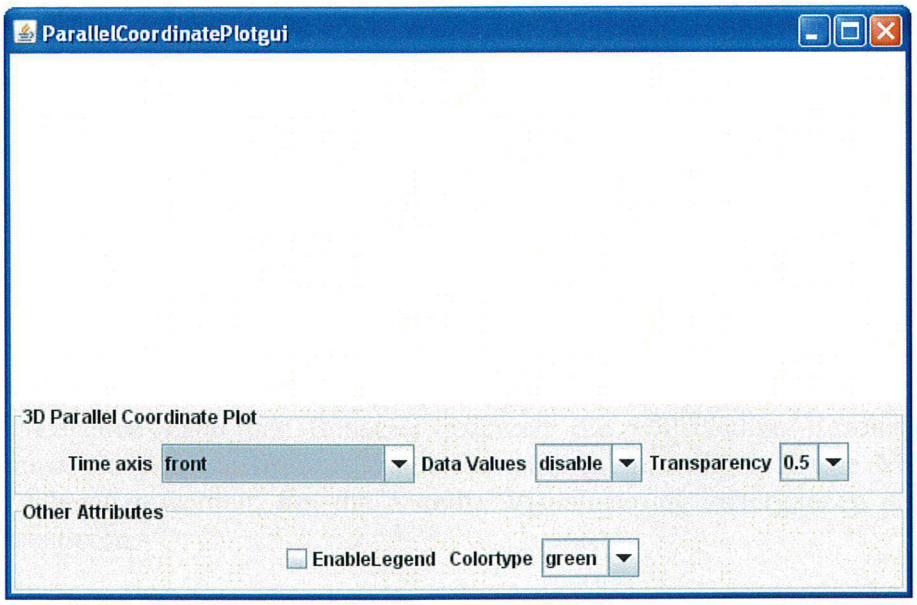


Abbildung 33: Benutzeroberfläche der Applikation

Am unteren Rand des Benutzerinterface befinden sich zwei Taskleisten. Mit Hilfe der darin enthaltenen Funktionen kann der Benutzer verschiedene Einstellungen an der Visualisierung

der Daten vornehmen. Eine Möglichkeit um einen vertieften Einblick in die Struktur der Daten zu gewinnen, soll das Umordnen der Achsen bieten. Mit der Pfeilbox „Time axis“ können die Koordinatenachsen der Darstellung neu angeordnet werden. Durch die Ansicht des Zeitparameters auf der x-Achse erschliesst sich dem Betrachter ein anderes Bild als die Darstellung des Zeitparameters auf der z-Achse.

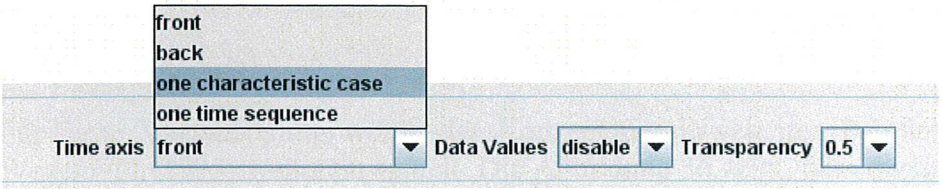


Abbildung 34: Auswahl eines spezifischen Merkmals über die Pfeilbox „Time Axis“

Durch die Pfeilbox „Data Values“ kann der Benutzer die Datenwerte der Merkmale ein- oder ausblenden. Des Weiteren bietet diese Auswahl dem Benutzer die Möglichkeit, die Datenwerte durch Hilfslinien mit dem Datenpunkt auf der entsprechenden Datenebene zu verbinden. Diese Option erleichtert dem Benutzer die Zuordnung bestimmter Attributwerte zu ausgesuchten Positionen im Datenraum.

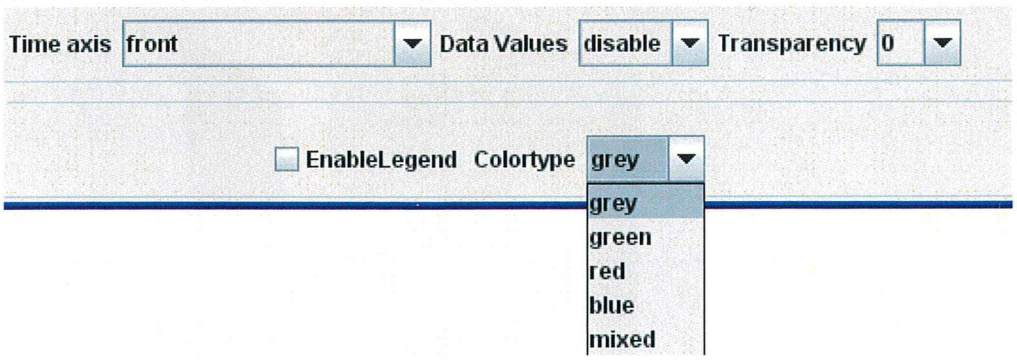


Abbildung 35: Detailansicht GUI des Prototypen

Ebenfalls erhält der Benutzer die Möglichkeit über die graphische Oberfläche die Einstellungen der Transparenz zu verändern. Über die Pfeilbox kann der Transparenzkoeffizient ausgewählt werden. Bei einer hohen Transparenz kann der Betrachter durch die einzelnen Flächen und die gesamte Oberfläche hindurch schauen.

Mit der Checkbox „Enable Legend“ kann eine Legende ein- oder ausgeblendet werden. Aus der Legende kann der Benutzer jederzeit die vollständige Bezeichnung der Merkmale herauslesen oder bestimmte Randinformationen hinzufügen. Diese Möglichkeit kann auch dazu benutzt werden, um die erstellte Visualisierung zu ergänzen und die Ergebnisse zu präsentieren.

Über die Pfeilbox „Colortype“ sind unterschiedliche Farbenkonfigurationen bzw. Farbskalen wählbar. Personen mit Farbsinnstörungen können auf diese Weise auf andere, für sie besser geeignete Farbverläufe, ausweichen. Je nach Fragestellung oder Analyseschwerpunkt kann ein harmonischer Farbverlauf gewählt oder aber trennende, gemischte Farbtöne gewählt werden.

6.2 Beispiele aus den Prototypen

Aufgrund der Überlegungen, die im Kapitel 3 formuliert wurden, konnten vier unterschiedliche Forschungsprototypen, die in den folgenden Abschnitten aufgezeigt werden, implementiert werden.

6.2.1 Beispiel 1

Für diese Visualisierung wurde das Konzept der Erweiterten Parallelen Koordinaten von Wegenkittl (1995) weiterentwickelt. Die Abbildung 36 zeigt, dass durch das Aufziehen einer ebenen Fläche hinter der eigentlichen Merkmalslinie, die Parallelen Koordinaten besser lesbar und plastisch greifbar werden. Diese Flächen können in der Tiefe (d.h. auf der Merkmalsachse) beliebig verkürzt oder verlängert werden um die Lesbarkeit weiter zu verbessern. Durch die drei „Zeitebenen" an der Vorderseite der Darstellung, kann der Wert eines Merkmals zu einem bestimmten Zeitstand, abgelesen werden. Die fünf Merkmale des Datensatzes sind farblich voneinander getrennt und bleiben über den Verlauf der Zeit gleich. Die Zeitstände sind zwecks guter Einsicht auf der x-Achse angeordnet und können vom Betrachter durch unterschiedliche Farbtöne schnell identifiziert werden.

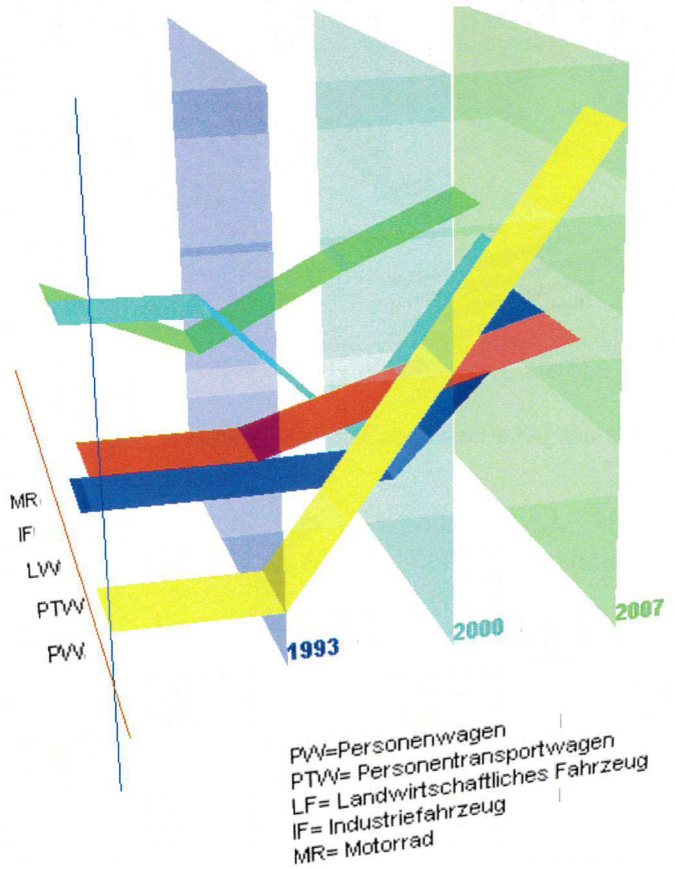


Abbildung 36: Beispiel 1 ParallelCoordinatePlot3d

6.2.2 Beispiel 2

In diesem Visualisierungsbeispiel für eine zeitachsenbasierte Visualisierung der Parallelen Koordinaten, wird jeweils eine Fläche zwischen zwei Merkmalslinien aufgespannt. Die beiden Längskanten einer Merkmalsfläche gehören nun zu zwei unterschiedlichen Merkmalen. Dadurch lassen sich die Relationen zwischen zwei Merkmalen und die Veränderung eines Merkmals über eine bestimmte Zeitdauer gut erkennen. Auf diese Weise sind beispielsweise in den Daten vorhandene Periodizitäten lokalisierbar. Der Vorteil des herkömmlichen Parallelen Koordinaten Diagramms, die deutlich sichtbaren Korrelationen, bleibt ebenfalls erhalten. Die farbliche Betonung der drei Zeitstände (aus dem Beispiel 1) wurde übernommen.

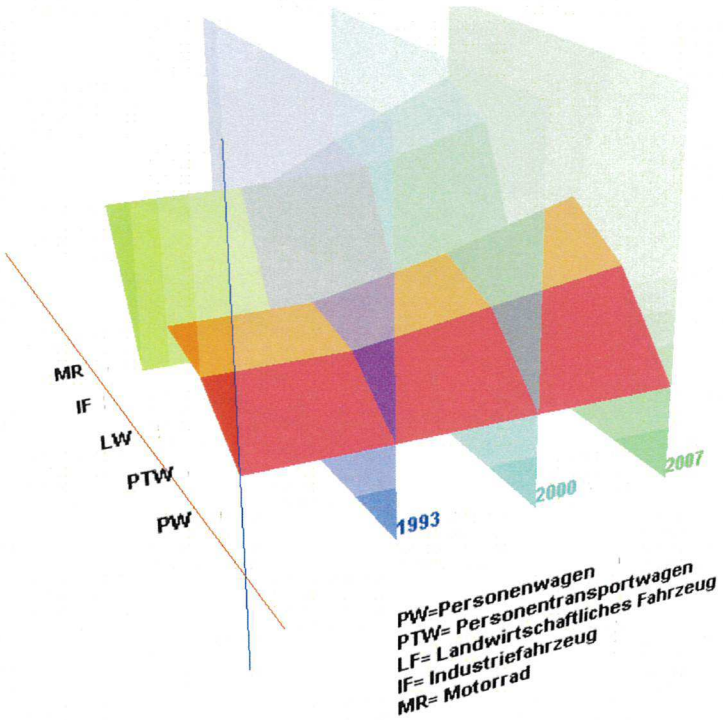


Abbildung 37: Beispiel 2 ParallelCoordinatePlot3d, mit Legende

Durch eine Ansicht von „linksoben“ (vgl. Abbildung 38) wird deutlich, dass für jeden Zeitstand eine farbliche Abfolge der Merkmalsflächen besteht.

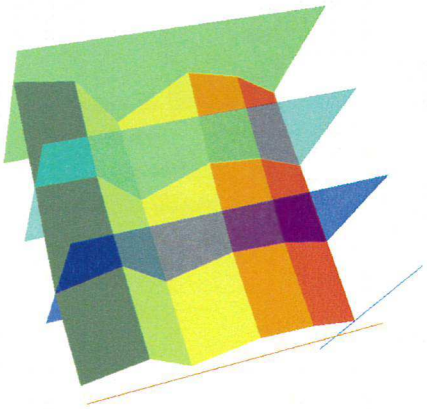


Abbildung 38: Beispiel 2 ParallelCoordinatePlot3d, Ansicht von schräg oben, ohne Legende

6.2.3 Beispiel 3

Das dritte Visualisierungsbeispiel, der in dieser Arbeit entwickelten Erweiterung der Parallelen Koordinaten, enthält die Möglichkeit, sich bestimmte Datenwerte anzeigen zu lassen. Die eingeblendeten Attributwerte werden mittels einer feinen Linie mit dem entsprechenden Datenpunkt im Raum verbunden. Die Minimum- und Maximumwerte eines Merkmals können auf Wunsch des Benutzers ebenfalls eingeblendet werden.

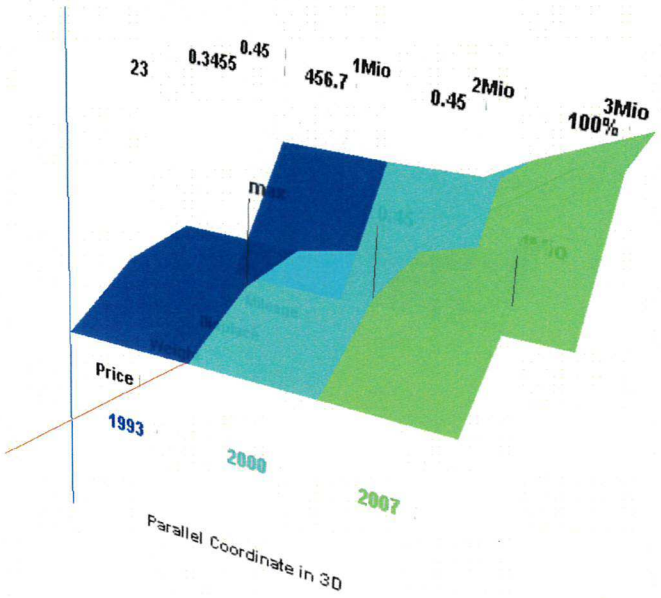


Abbildung 39: Beispiel 3 ParallelCoordinatePlot3d, Ansicht von schräg vorne, mit Extrema und Hilfslinien

In diesem Beispiel wird die Zeit hauptsächlich durch die visuelle Variable der Farbe repräsentiert. Diese Darstellungsart eignet sich speziell für Fragestellungen, die das Verhalten aller Variablen in einem bestimmten Zeitintervall oder an einem Zeitpunkt untersuchen. Bei grossen Datensätzen mit dutzenden Zeitständen wäre auch ein kontinuierlicher Farbverlauf denkbar.

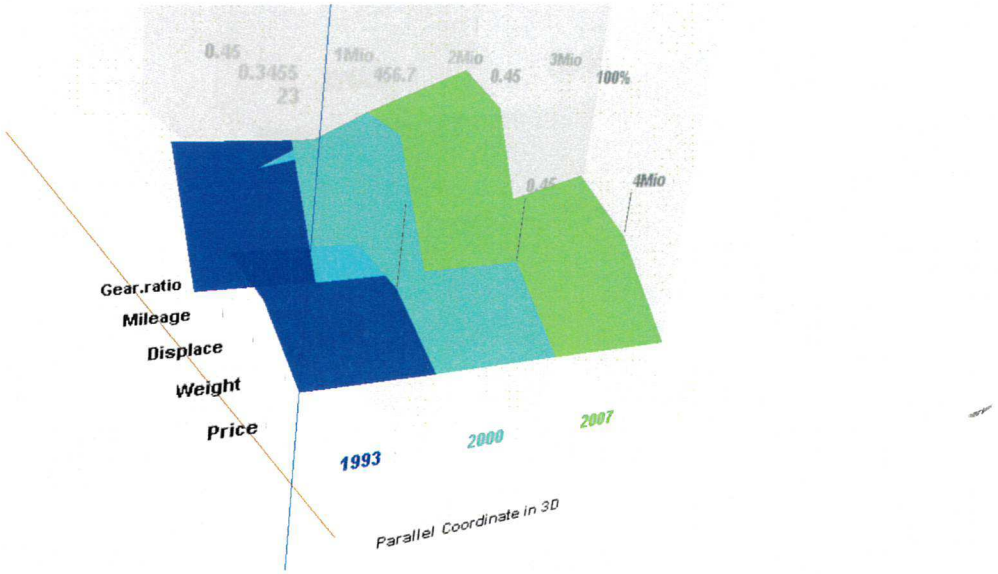


Abbildung 40: Beispiel 3 ParallelCoordinatePlot3d, Ansicht von schräg vorne

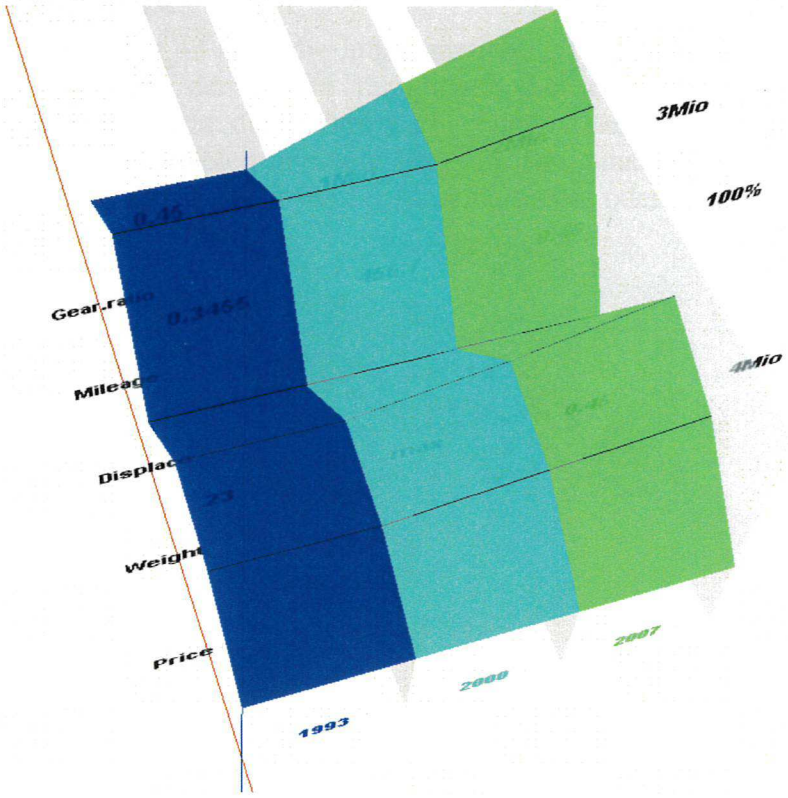


Abbildung 41: Beispiel 3 ParallelCoordinatePlote3d, Ansicht von oben, mit eingeblendeten Merkmalslinien

Durch das Darstellen der Merkmalslinien auf der Oberfläche kann die Entzifferbarkeit der Datenwerte weiter verbessert werden (vgl. Abb. 41 und Abb. 42). Durch das Einstellen der Transparenz einzelner Flächen kann der Betrachter beim interaktiven Drehen der Visualisierung alle Linien auf einen Blick erfassen. Zur besseren Erkennbarkeit können die Datenwerte ebenfalls ein- oder ausgeblendet werden.

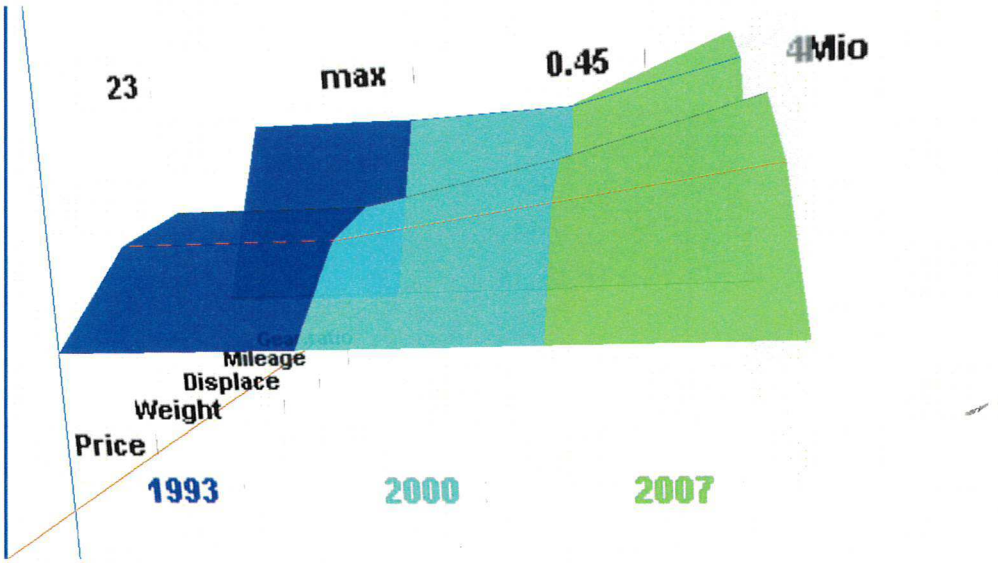


Abbildung 42: Beispiel 3 ParallelCoordinatePlot3d, Frontansicht, mit eingeblendeten Merkmalslinien

6.2.4 Beispiel 4

Die Abbildung 43 zeigt die Implementierung für Daten mit einem zyklischen Zeitbezug. Die Zeit wird auf der dritten räumlichen Achse, der z-Achse, dargestellt. Die einzelnen Zeitintervalle sind durch ihre Farbgebung voneinander unterscheidbar und repräsentieren jeweils alle Variablen in einem Intervall. Die genauen Zahlenwerte können auf Wunsch des Betrachters am oberen Ende der Merkmalsachse oder in der Nähe des Punktes auf der Fläche, eingeblendet werden.

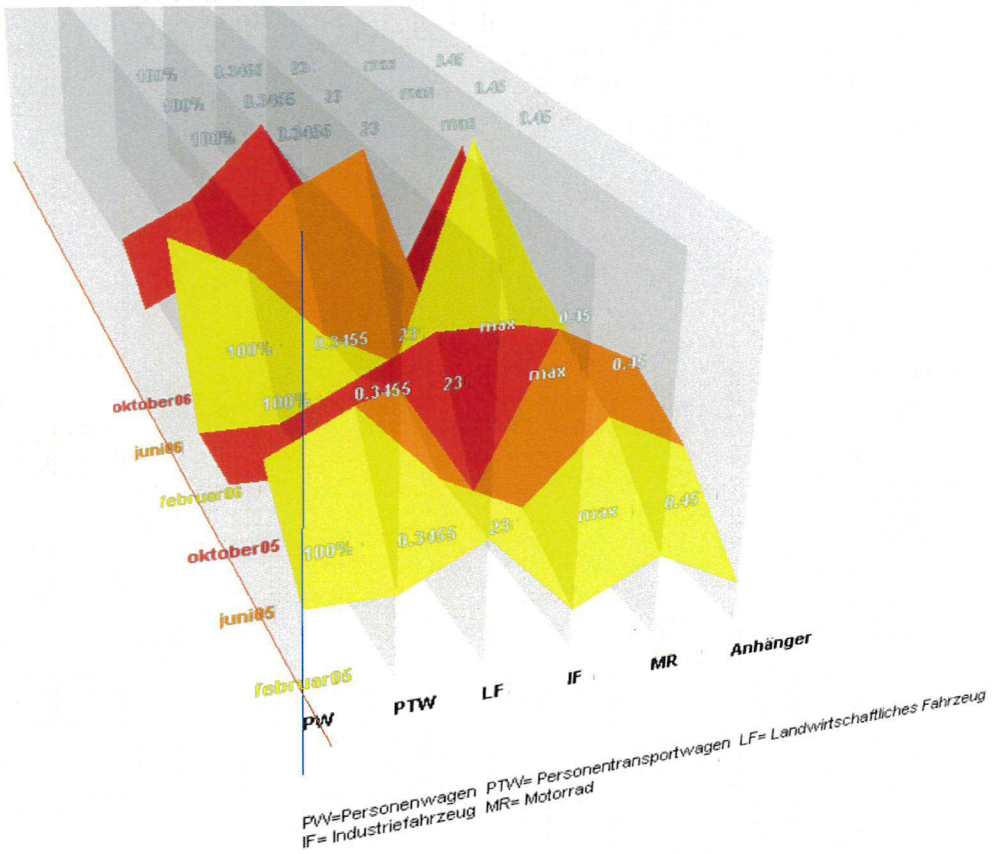


Abbildung 43: Beispiel 4 ParallelCoordinatePlot3d, zyklischer Datensatz

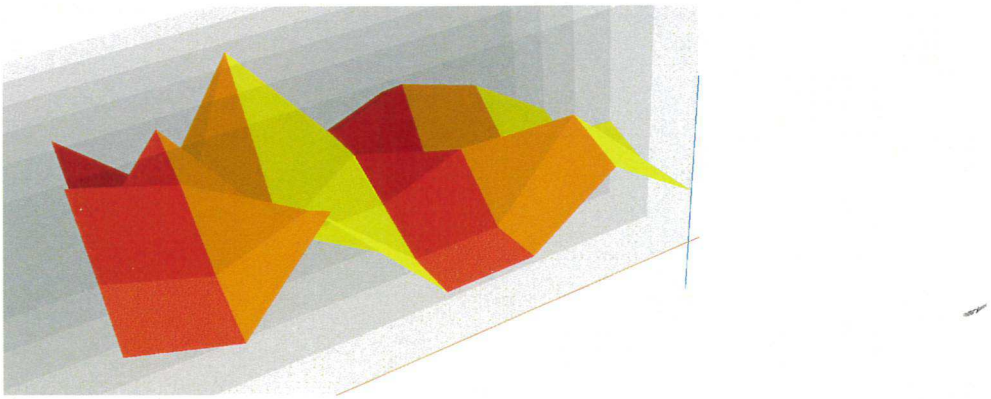


Abbildung 44: Beispiel 4 ParallelCoordinatePlot3d, Seitenansicht, zeitlicher Verlauf und topologische Oberfläche

Durch das Drehen der gesamten Darstellung in eine Seitenansicht (vgl. Abb. 44), wird die Oberfläche des visualisierten Datensatzes sichtbar. An dieser topologischen Oberfläche des Datenraums können nun zeitliche Musters und Verhalten abgelesen werden. Durch das Ein- und Ausblenden einzelner Zeitstände werden zyklische Verhalten oder Periodizitäten ersichtlich. Der Benutzer kann beispielsweise alle gelben Flächen „zusammenschieben" und sieht auf diese Weise das Verhalten der Attribute im Monat Februar. Durch das Ausblenden von Zeitständen kann beispielsweise nur der Monat Februar visualisiert werden.

7 Diskussion

In diesem Kapitel werden die vorgestellten Visualisierungsbeispiele diskutiert und die zu Beginn der vorliegenden Arbeit formulierten Forschungsfragen (vgl. Kapitel 1.3) beantwortet.

7.1. Diskussion der erstellten Beispiele

Alle vier visualisierten Beispiele, welche mit dem entwickelten Prototyp generiert wurden, zeigen den Verlauf der in den Daten enthaltenen, zeitlichen Information. Die Realisierung des in der vorliegenden Arbeit entworfenen Konzepts einer 3D-Erweiterung der herkömmlichen Visualisierungsmethode der Parallelen Koordinaten, wurde vorgenommen und umgesetzt.

Die erstellten Visualisierungsbeispiele werden aufgrund der Erfüllung der Anforderungen an Visualisierungen im Allgemeinen (vgl. Kapitel 2.2.2) und an dreidimensionale Anforderungen (vgl. Kapitel 2.4) beurteilt.

Anforderung	**Umsetzung**
Räumliche Trennung	Alle geometrischen Objekte wurden im visualisierten Datenraum so angeordnet, sich sich gegenseitig so wenig wie möglich überdecken oder überzeichnen.
Räumliche Identifizierbarkeit	Koordinatenachsen und Merkmalsachsen sind eindeutig voneinander unterscheidbar und durch unterschiedliche Farbgebung und Position für den Betrachter identifizierbar.
Unterstützung der Erkennbarkeit	Durch die intuitive Navigation mittels des Eingabegeräts (Maus) wird das Erforschen und Begehen des visualisierten Datenraumes unterstützt. Interaktionen, wie das Ein- und Ausblenden einzelner Elemente oder das Verändern der Transparenz erlauben dem Benutzer die Untersuchung spezifischer Bereiche. Die Gesamtdarstellung aller Merkmale macht die Übersicht über den gesamten Datenraum möglich.

	Der Vergleich zwischen zwei nebeneinander liegenden Merkmalen kann durch die visuelle Beurteilung der Flächenposition vorgenommen werden. Durch das Drehen der gesamten Darstellung kann der Betrachter den zeitlichen Verlauf eines bestimmten Merkmals verfolgen. Das Erkennen von Strukturen und Verläufen wird durch das Visualisieren der topologischen Oberfläche des Datenraums intuitiver und einfacher. Die bestmöglichste Unterscheidung der einzelnen Merkmale und die Verbindung gleicher Zeitstände wird mittels einer geeigneten Farbgebung unterstützt und ermöglicht.

Tabelle 2: Umsetzung der Anforderungen an einen dreidimensionale Visualisierung (Quelle: eigene Darstellung)

Das Beispiel 1 (vgl. Kapitel 6.2.1) visualisiert die dreidimensionalen Parallelen Koordinaten als ebene „Bänder". Die Zeitstände werden auf der x-Achse skaliert, so dass sie für den Betrachter im Vordergrund sind. Dieses Visualisierungsbeispiel eignet sich besonders gut für linear verlaufende Zeitreihen mit einzelnen Beobachtungszeitständen. Durch das „Vertiefen" der Merkmalsflächen kann die Lesbarkeit der Parallelen Koordinaten verbessert werden. Die farbliche Trennung der Merkmale ermöglicht die Identifikation über das Zeitintervall hinweg.

Das Beispiel 2 (vgl. Kapitel 6.2.2) zeigt den visualisierten Datenraum als „topologische" Oberfläche. Diese Darstellung ermöglicht dem Betrachter das intuitive Erfassen des Datenraums und den Vergleich zwischen zwei nebeneinander liegenden Merkmalen. Die Oberfläche, die durch die Visualisierung des Konzepts entsteht, kann nun analysiert werden. Für den Benutzer ergeben sich durch diese Darstellungsart der Parallelen Koordinaten neue Analysemöglichkeiten. Durch die Krümmung und den Winkel zwischen zwei vom Benutzer festgelegten Punkten können Teilflächen begutachtet und ausgewertet werden. Eine mathematische Beschreibung einer solchen Oberfläche ist allerdings keineswegs trivial. Um sie zu untersuchen müssen interaktive Werkzeuge entwickelt werden, die komplexe mathematische Verfahren (Bézierkurven, B-Splines, NURBS etc.) anwenden können.

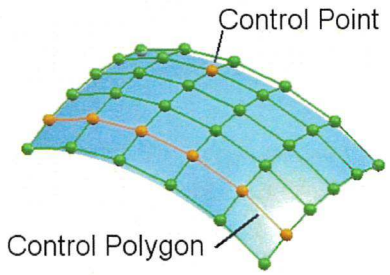


Abbildung 45: Durch eine bestimmte Anzahl von Kontrollpunkten oder Kontrollpolygonen kann eine komplexe Fläche beschrieben werden. (Quelle: CADauno, 2009)

Das Beispiel 3 (vgl. Kapitel 6.2.3) zeigt die Möglichkeit des Darstellens von Rand- und Zusatzinformationen. Der Benutzer kann Attributwerte einblenden und diese mit der Datenoberfläche verbinden. Auf diese Weise kann ein spezifischer, lokaler Datenraum untersucht werden. Die Merkmalsachsen sind vom Merkmalsminimum bis zum Merkmalsmaximum skaliert. Das graphische Gestaltungsmittel der Farbe wird hier eingesetzt um die Zeitstände zu visualisieren. Es werden zeitliche Verläufe und die zeitliche Veränderung aller Merkmale zu einem bestimmten Zeitpunkt ersichtlich. Durch das Einblenden der herkömmlichen Polylinien der Merkmale erschliesst sich dem Betrachter eine Kombination der konventionellen Merkmalslinien der Parallelen Koordinaten, in Kombination mit der Datenoberfläche.

Das Visualisierungsbeispiel 4 (vgl. Kapitel 6.2.4) eignet sich für die Darstellung von zeitabhängigen Daten mit einem zyklischen Zeitbezug. Die Zeitachse verläuft entlang der z-Achse, während in diesem Beispiel die Merkmale auf der x-Achse dargestellt werden. Diese Darstellung entspricht dem intuitiven Verständnis des Menschen, Zeit als eine dritte, räumliche Dimension wahrzunehmen. Es hat sich gezeigt, dass die Separierung der einzelnen Merkmale am besten durch vertikale, transparente Flächen erfolgt. Das Einblenden der numerischen Attributwerte in unmittelbarer Nähe der entsprechenden Merkmalsfläche ermöglicht das richtige Zuordnen. Die Seitenansicht (vgl. Abb. 44) zeigt den Verlauf eines zeitlichen Musters über die Zeit. Durch das Ein- und Ausblenden bestimmter Zeitstände kann ein Merkmal über den gesamten Zeitzyklus verfolgt werden.

Für alle der hier vorgestellten Visualisierungsbeispiele gilt folgende Tatsache: Einer der grössten Vorteile der Parallelen Koordinaten, die einfache Aufdeckung von Korrelationen, bleibt in der entwickelten und implementierten, dreidimensionalen Erweiterung der Methode, bestehen. Analog zu den parallelen Linien im herkömmlichen Parallelen Koordinaten Diagramm, zeigen nun zueinander parallele Flächen hohe positive Korrelationen auf. Sich durchschneidende bzw. kreuzende Polygone weisen auf negative Korrelationen hin.

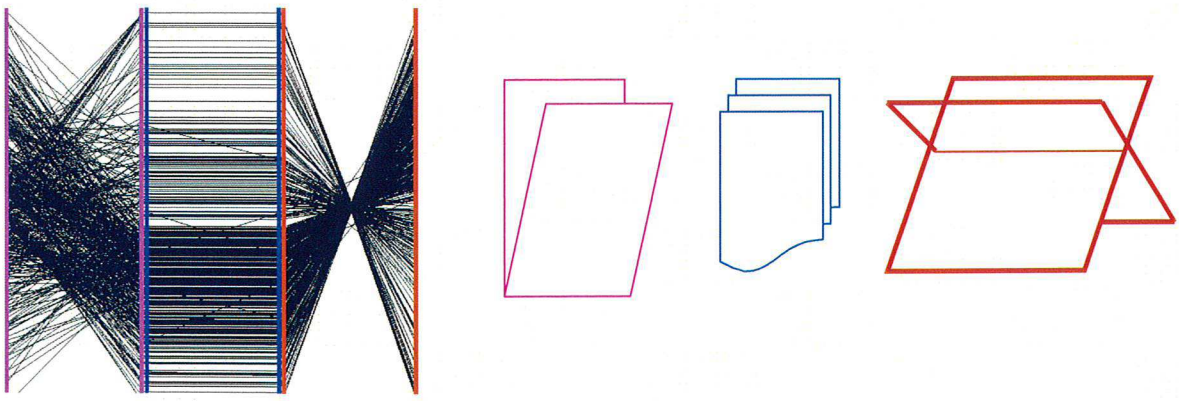


Abbildung 46: keine Korrelation (lila), positive Korrelatin (blau), negative Koorelation (rot) (Quelle: eigene Darstellung nach Electronic Visualization Laboratory, 2001)

Zusätzlich zu den bereits erwähnten und implementierten Analysemöglichkeiten und interaktiven Werkzeugen, bieten sich durch die entwickelte Erweiterung, neue Chancen und Möglichkeiten für die visuelle Analyse dreidimensionaler Parallelen Koordinaten für zeitabhängige Daten: Es wäre denkbar, mittels einer visualisierten Kugel, die der Benutzer auf der Oberfläche des Datenraums platzieren kann, und die den Gesetzen der Schwerkraft folgt, lokale Extremwerte (z.B. ein Minima) oder Steigungen der Fläche zu detektieren. Ebenfalls

kann nun die entstehende Datenoberfläche durch eine gezielt entwickelte Regressionsanalyse analysiert werden. Bei einer multiplen, nichtlinearen Regressionsanalyse ergeben die Lösungsfunktionen gekrümmte Flächen oder Hyperräume. Basierend auf der Arbeit von Demsar et al. (2008) kann mit der statistischen Methode der geographisch gerichteten Regression GWR die zeitliche Dynamik auf der Oberfläche der Darstellung untersucht werden. Mit Hilfe von Regressionsparametern kann das Verhalten und Verändern einer Polygonfläche über mehrere Zeitstände hinaus, untersucht werden. Inwiefern sich die Regressionsanalyse für diese Anwendung eignet muss untersucht werden.

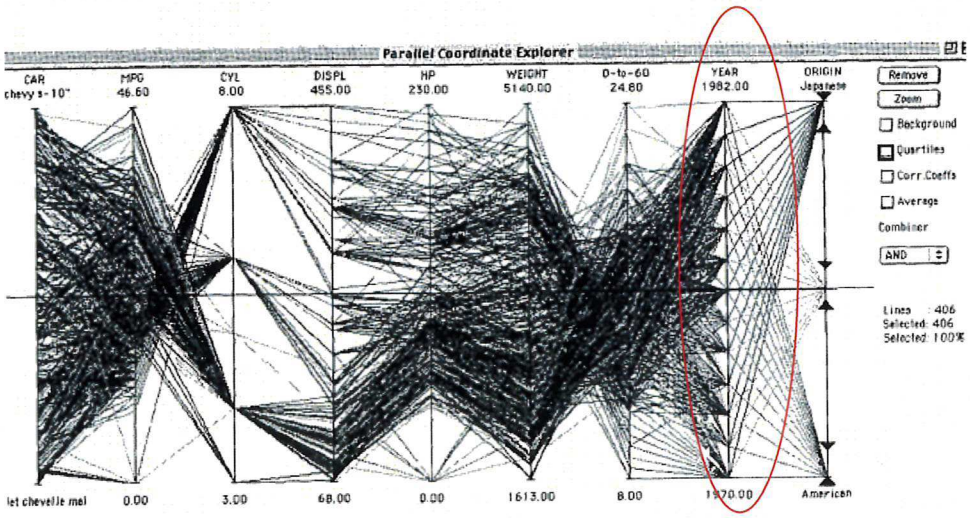


Abbildung 47: herkömmliche Darstellung der Variable Zeit (Variable: YEAR), (Quelle: Siirtola, 2000)

In Abbildung 47 ist die herkömmliche Darstellung der Variable Zeit in einem Parallelen Koordinaten Diagramm ersichtlich. Nur durch das Hervorheben der Merkmalslinien, die durch den interessierenden Bereich auf der Achse mit der Zeit verlaufen, kann herausgelesen werden, wie die anderen Attribute zu diesem Zeitpunkt ausgesehen haben. Die neu entwickelte Darstellungsart der Parallelen Koordinaten (vgl. Abb. 48), welche die Variable Zeit in einer dreidimensionalen Visualisierung mit einer separaten Zeitdimension aufzeigt, ist für den Betrachter einfacher und intuitiver lesbar.

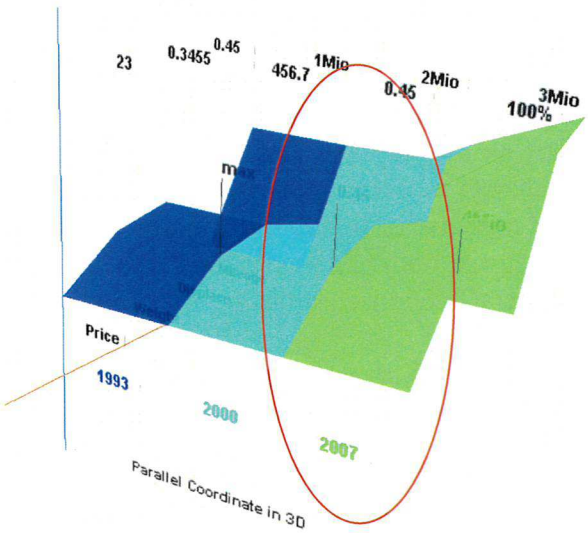


Abbildung 48: neue Darstellungsform mit der Variable Zeit auf der x-Achse

7.2 Die Beantwortung der Fragestellungen

Forschungsfrage 1: ***Ist die Zeit eine „normale" Variable, die analog zu anderen Variablen behandelt werden kann?***

Zeitabhängige Daten finden sich heute in fast allen Bereichen der Forschung oder des täglichen Lebens. Um Fragen wie „wann hatte etwas seine grösste Ausprägung?", „ist da ein zeitlich variierendes Muster?" oder „sind sich zwei Verhalten ähnlich?" beantworten zu können, muss eine geeignete, effiziente Visualisierungsmethode angewendet werden. Im Kapitel 3 wurde aufgezeigt, dass die Zeiteinheiten und die Zeitintervalle einer linearen Zeitfolge in einem Datensatz, analog zu Werten und Wertebereichen von Attributen beliebiger anderen linear geordneter Datensätze behandelt werden können. Da aber speziell in der visuellen Datenanalyse, die oft als explorative Untersuchung am Beginn einer Forschungsarbeit angewandt wird, nicht immer klar ist, ob es sich bei den Daten um eine lineare Zeitreihe handelt oder nicht, empfiehlt es sich, die Zeit als eine aussergewöhnliche Dimension zu betrachten und bei der Visualisierung besonders zu berücksichtigen. Die Variable „Zeit" wird vom Menschen intuitiv als eine zusätzliche Dimension des Beaobachtungsraumes aufgefasst. Diese Tatsache spricht gegen die Gleichbehandlung einer abhängigen Variablen und der Variable Zeit. Sie soll deshalb abhängig von ihrem Zeitbezug (statisch, quasistatisch oder dynamisch) und dem spezifischen Anwendungsziel repräsentiert werden. Nur durch eine angemessene Darstellung können wichtige Datencharakteristika zeitlich richtig zugeordnet werden. Die Forschungsfrage muss folglich mit einem Nein beantwortet werden.

Forschungsfrage 2: ***Welche dreidimensionalen, achsenbasierte Visualisierungstechniken können für eine Darstellung zeitabhängiger, multidimensionaler Daten verwendet werden?***

Im Kapitel 3 wird die Darstellung der Zeitabhängigkeit im Kontext der multidimensionalen Daten ausführlich vorgestellt. Es werden verschiedene dreidimensionale Ansätze vorgestellt, welche die Zeitabhängigkeit eines Datensatzes entlang einer Zeitachse visualisieren. Darstellungen dieser Art entsprechen dem intuitiven Verständnis und der Vorstellung des Menschen, Zeit als eine dritte, in eine Richtung verlaufende, Dimension wahrzunehmen und kognitiv zu erschliessen. Die ebenfalls achsenbasierte Methode der Parallelen Koordinaten erweist sich als vorteilhaft, da ihre Komplexität (hier, die Anzahl der Merkmalsachsen), direkt proportional zu der Anzahl an Attributen ist. Durch die vorgestellten Ansätze der verschiedenen Forschungsgruppe wurde ersichtlich dass „zeitliche" Dreidimensionalität durch ein Hintereinanderstellen einzelner Darstellungen erreicht wird. Diese Erkenntnis, zusammen mit den Vorteilen und charakteristischen Eigenschaften der Parallelen Koordinaten, führte zu dem, in der vorliegenden Arbeit vorgestellten, Konzept der dreidimensionalen Parallelen Koordinaten für multidimensionale Daten mit einem Zeitbezug.

Forschungsfrage 3: ***Wie kann die Visualisierungsmethode der Parallelen Koordinaten um eine „Zeitdimension" erweitert werden?***

Die Methode der Parallelen Koordinaten stellt für multidimensionale Daten eine gute und geeignete Visualisierungstechnik, mit vielen Vorteilen, dar. Sie erlaubt eine verlustfreie und eindeutige Abbildung des n-dimensionalen Raumes in die Ebene. Wie in Kapitel 2.4

aufgezeigt wurde, lässt die herkömmliche Methode nur bedingt eine Analyse des Zeitbezugs in einem Datensatz zu. Die in der vorliegenden Arbeit erarbeitete, dreidimensionale Erweiterung der Parallelen Koordinaten, baut stark auf der Vorarbeit von Wegenkittl (1995) auf. Analog zu seiner Idee basiert das Konzept auf polygonalen Flächen, die durch die Merkmalslinien einzelner Datenelemente aufgespannt werden. Durch das Verbinden dieser Flächen entstehen flächige „Streckenzüge" die in die dritte Dimension hineinreichen. Durch das Abbilden der in den Daten enthaltenen Zeitprimitiven, kann der Zeitbezug prägnant visualisiert werden.

Die allgemeinen Anforderungen (vgl. Kapitel 2.1.2) an eine Visualisierung wurden bei der Entwicklung der vorgeschlagenen Erweiterung berücksichtigt. Eine möglichst unverfälschte Datenwiedergabe, die dem Betrachter eine Vermittlung aller relevanten Aspekte garantiert, wird durch den implementierten Prototyp gewährleistet. Unter Berücksichtigung des Effektivitätskriteriums wurde entschieden, eine dreidimensionale Visualisierung mit Hilfe der Softwareumgebung Java3D zu entwickeln. Wie im Kapitel 4 beschrieben wurde, unterstützt Java3D die theoretischen Konzepte der Fragestellung und eignet sich deshalb gut für die Implementierung dreidimensionaler Visualisierungen. Aufgrund der Möglichkeit von Java3D, mit relativ geringem Aufwand Applikationen zu programmieren, erfüllt der Entwicklungsprozess auch die dritte Anforderung an eine Visualisierung: die Angemessenheit. Das hier entwickelte Konzept zur Erweiterung der Parallelen Koordinaten um eine Zeitdimension, wird ebenfalls den, in Kapitel 2 formulierten Hauptanforderungen, an eine dreidimensionale Visualisierung, gerecht. Die Anforderung der räumlichen Trennung wird durch eine strukturierte, klare Anordnung der geometrischen Elemente umgesetzt. Alle Objekte sind so platziert, dass sie sich so wenig wie möglich gegenseitig überlappen. Durch intuitiv bedienbare Navigationsmöglichkeiten kann der Betrachter alle Elemente des Datenraumes einsehen und untersuchen. Das Kriterium der räumlichen Identifizierbarkeit wird in der neuen Darstellungsform erfüllt. Alle Achsen und deren Funktionen sind voneinander unterscheidbar und für den Benutzer eindeutig identifizierbar. Die Unterstützung der Erkennbarkeit ist durch ein logisch aufgebautes und gut strukturiertes Benutzerinterface gewährleistet. Es ermöglicht dem Benutzer durch verschieden Funktionen und Interaktionen den dreidimensionalen Datenraum zu erkunden und nach eigenen Vorstellungen bzw. Analyseschwerpunkten zu gestalten.

Forschungsfrage 4: ***Welche Vorteile und Möglichkeiten erschliessen sich dem Betrachter eines um die Zeitdimension erweiterten Parallelen Koordinaten Diagramms?***

Durch die in der vorliegenden Arbeit entwickelte und implementierte Erweiterung, wird die Visualisierungsmethode der Parallelen Koordinaten für multidimensionale Datensätze mit einem Zeitbezug, einsetzbar. Die vorgestellten Visualisierungsbeispiele in Kapitel 6 zeigen, dass sich durch die entwickelte Visualisierungsmethode, zeitliche Muster und Besonderheiten darstellen und erkennen lassen. Periodizitäten, Korrelationen, Wachstum oder Abnahme eines Datenmerkmals werden auf den ersten Blick erfassbar. Durch die farbliche Gestaltung einzelner Zeitstände kann ein Merkmal, an einem bestimmten Zeitpunkt oder während eines Zeitintervalls, untersucht werden. Das ist in der herkömmlichen Methode nur möglich, wenn die Merkmalsachse der zeitlichen Variable neben dem zu untersuchenden Merkmal liegt. Diese Möglichkeit der Darstellung von Zeitbezügen erweist sich in allen Beispielen, die mit dem in der vorliegenden Arbeit implementierten Prototyp erstellt worden sind, als vorteilhaft und neu gegenüber der herkömmlichen Methode.

Dem Betrachter der interaktiven, dreidimensionalen Erweiterung der Parallelen Koordinaten, erschliessen sich weiterhin alle Vorteile der traditionellen Visualisierungsmethode, sowie einige neue Analysemöglichkeiten. Die bereits entwickelten interaktive Anwendungen und Werkzeuge für Parallele Koordinaten, wie *brushing and linking, highlighting* oder das Ein- und Ausblenden einzelner Linien, können auch für die polygonalen Merkmalsflächen der neuen Methode verwendet werden. Ebenfalls erwähnt wurden die mathematischen und geometrischen Analysemöglichkeiten der „topologischen" Oberfläche. Durch die Darstellung der Datenoberfläche wird dem Betrachter ein intuitiver und einfacher visueller Zugang zu den Merkmalen ermöglicht.

8 Schlussfolgerung und Fazit

8.1 Erkenntnisse

Die vorliegende Arbeit hat gezeigt, dass es grundsätzlich möglich ist, zeitbezogene, multidimensionale Daten mit Hilfe einer erweiterten, dreidimensionalen Form der Parallelen Koordinaten, effektiv und angemessen zu visualisieren. Der entwickelte Prototyp ist in der Lage, anhand verschiedener Beispielsdatensätze, vier unterschiedliche Visualisierungsbeispiele zu generieren, die dem Benutzer einen neuen Einblick in die Struktur der Daten erlauben. Es wurde ebenfalls gezeigt, dass die herkömmliche Ansicht des Parallelen Koordinaten Diagramms und die damit verbundenen Vorteile der Methode, vollumfänglich erhalten bleiben. Durch die dreidimensionale Ansicht ist es möglich, neue Erkenntnisse über den visualisierten Datensatz zu gewinnen und zeitliche Prozesse oder Veränderungen sichtbar zu machen. Die zusätzliche Dimension, erhöht den Freiheitsgrad der Darstellung, wodurch eine grössere Anzahl an Merkmalen dargestellt werden kann. Dem Betrachter wird mit der entwickelten Methode ein geeignetes Hilfsmittel in die Hand gegeben, mit dem er Ergebnisse einer Datenanalyse visualisieren und kommunizieren kann.

8.2 Verbesserungsmöglichkeiten und Limiten

Das Ziel der vorliegenden Arbeit, die Visualisierungsmethode der Parallelen Koordinaten so zu erweitern, dass die Zeitabhängigkeit multidimensionaler Daten untersucht und dargestellt werden kann, wurde erreicht. Die vorgestellten Beispiele erlauben einen neuen und prägnanten Einblick in die Daten und ermöglichen dem Benutzer Fragestellungen zu zeitlichen Verläufen zu untersuchen und zu beantworten. Der in dieser Arbeit entwickelte Prototyp für die Darstellung von zeitabhängigen Multiparameterdaten ist eine Neuentwicklung. Aus diesem Grund bieten das Benutzerinterface und das implementierte Programm viele Möglichkeiten um Erweiterungen und Verbesserungen anzubringen. Der Prototyp ist so aufgebaut, dass er weiter entwickelt und neuen Erkenntnissen im Bereich der Visualisierung oder der Software, angepasst werden kann. Eine Einbindung in eines der bestehenden, Java basierten Visualisierungstools für multidimensionale Daten (z.B. GGobi[12] oder GeoVISTA Studio[13]), ist ebenfalls denkbar und wünschenswert. Gegenwärtig ist der entwickelte Prototyp auf Experimente ausgelegt, die das Visualisieren einzelner Beispieldatensätze erlauben. Das manuelle Umsortieren der Merkmalsachsen ist derzeit nicht

[12] GGobi ist ein Open-Source-Visualisierungsprogramm für die Analyse von hochdimensionalen Daten. Es enthält dynamische und interaktive Grafiken, Animationen, sowie Visualisierungsmethoden wie Streu- und Säulendiagramme oder Parallele-Koordinaten-Plots. Siehe http://www.ggobi.org für mehr Informationen.
[13] GeoVISTA *Studio* ist eine Opensource Softwareumgebung die Berechnung und Visualisierung von räumlichen Daten. Siehe http://www.geovistastudio.psu.edu/jsp/index.jsp

möglich. Da diese Möglichkeit bei der Visualisierungsmethode der Parallelen Koordinaten wichtig für die Aufdeckung von bestimmten Zusammenhängen zwischen den Merkmalen ist, muss dem Benutzer in einer folgenden Erweiterung die Gelegenheit gegeben werden, die Achsen nach seiner Wahl um- oder anzuordnen. Zugleich ist es zur Zeit nicht möglich einen kompletten Datensatz automatisch einzulesen und zu visualisieren. Bei einer künftigen Erweiterung muss dieser Aspekt berücksichtigt und eine Möglichkeit zum Einlesen von verschiedenen, multidimensionalen Daten gefunden werden. Im Rahmen dieser Arbeit wurde bewusst auf eine Evaluierung der vorgestellten Beispiele verzichtet. Abgesehen von den vorher beschriebenen Weiterentwicklungsmöglichkeiten, wäre eine Evaluation der vorgestellten Beispiele aber wünschenswert. Nur so kann eine genaue Einstufung und Bewertung der visualisierten Beispiele vorgenommen und wissenschaftlich untersucht werden, wie sie qualifizierte Experten oder Laien beurteilen.

8.3 Offene Fragen

Die Visualisierung grosser Datensätze mit mehreren hunderten Variablen oder Zeitständen werfen neue Probleme bezüglich Speicherplatz, Komplexität, Rechenkapazität, Analyse und Darstellung auf. Es muss weiter untersucht werden, wo die Grenzen für die vorgestellte Darstellungsmethode liegen und wie diese überschritten werden können. Ebenfalls bietet die topologische Oberfläche, welche durch diese Visualisierungsmethode entsteht, weitere Möglichkeiten für weitere mathematische oder geometrische Analysen. Es muss untersucht werden, ob die räumliche Wahrnehmung des Benutzers, durch zusätzliche interaktive, visuelle Verfahren zur quantitativen Analyse räumlicher Strukturen, verbessert werden könnte. Weiterer Forschungsbedarf besteht auch hinsichtlich der Darstellung von zeitlichen Verläufen mittels dem graphische Gestaltungsmittel der Farbe. Es fehlt an Fachliteratur, die untersucht inwiefern der Mensch mit Hilfe von Farbverläufen oder bestimmten Farbskalen, eine zeitliche Veränderung oder eine Richtung in der Zeit intuitiv erfassen kann. Durch eine farbliche Betonung der Zeitachse könnte die Lesbarkeit, des in der vorliegenden Arbeit entwickelten Visualisierungskonzepts, weiter verbessert werden.

8.4 Ausblick

Interaktive, dreidimensionale Visualisierungsmethoden, die einem Betrachter erlauben in den Datenraum einzudringen und ihm spezifische Werkzeuge zur Verfügung stellen um seine Fragestellungen zu beantworten, sind in der visuellen Datenanalyse der Zukunft unverzichtbar. Durch die immer grösser werdende Datenmenge und die immer leistungsfähigere Hardware werden zeitabhängige Multiparameterdaten in Zukunft weiter an Bedeutung gewinnen. Neuste Forschungsarbeiten, wie beispielsweise Siirtola et al. (2009), zeigen, dass die Visualisierungsmethode der Parallelen Koordinaten noch wie vor eine der geeignetsten Methoden für eine umfassende Darstellung von multidimensionalen Datensätzen ist. Das Vorurteil, dass sich Parallele Koordinaten nur für Experten eignen hat sich in einem von Siirtola et al. (2009) durchgeführten Wahrnehmungsexperimente als unhaltbar erwiesen. Projektbezogene, spezifische 3D-Erweiterungen von bewährten, herkömmlichen Methoden machen es möglich, neue Fragestellungen zu stellen und zu beantworten.

Literaturverzeichnis und Onlinequellen

Andrienko, N.; Andrienko, G. (2005): Visual Exploration of the Spatial Distribution of Temporal Behaviors. Proceedings of the ninth International Conference on Information Visualization, IV'05, 6(8): S.799-806

Andrienko, N.; Andrienko, G. (2006): Exploratory Analysis of Spatial and Temporal Data. A Systematic Approach, Heidelberg: Springer

Aigner, W.; Miksch, S.; Müller, W.; Schumann, H.; Tominski, C. (2007): Visualizing time-oriented data. A systematic view, Computers and Graphics, 31(3): S.401-409

Aigner, W.; Miksch, S.; Müller, W.; Schumann, H.; Tominski, C. (2008): Visual Methods for Analyzing Time-Oriented Data. IEEE Transaction on Visualization and Computer Graphics, 14(1): S.47-60

Arnold, C. (2008): Visualisierung im Information Retrieval. Theorie, Anwendung und empirischer Vergleich, Saarbrücken: Dr. Müller Verlag

Bertin, J. (1974): Graphische Semiologie. Diagramme Netze Karten, Berlin: Walter de Gruyter

Bertin, J. (1982): Graphische Darstellungen. Graphische Verarbeitung von Information, Berlin: Walter de Gruyter

Buono, P.; Aris, A.; Plaisant, C.; Khella, A. ; Shneiderman, B. (2005) :Interactive Pattern Search on Time Series. In: Proceedings of Conference on Visualization and Data Analysis, VDA 2005, SPIE: S.175-186

Card, S.K.; Mackinlay, J.D.; Shneiderman, B. (1999): Readings in Information Visualization. Using Vision to Think, San Francisco: CA Morgan Kaufmann

Carlis, J.V.; Konstan, J.A. (1998): Interactive visualization of serial periodic data. Proceedings of the 11th annual ACM Symposium on User Interface Software and Technology, San Francisco, California: S. 29-38

Carter, R.C. (1982): Visual search with Color. Journal of Experimental Psychology, Human Perception and Performance, 8: S.127-136

Chen, C.; Härdle, W.; Unwin, A. (2008): Handbook of Data Visualization. Berlin: Springer

Cook, D.; Swayne, D. (2007): Interactive and Dynamic Graphics for Data Analysis. With R and GGobi, New York: Springer Sience+Business Media

Crampton, J. (2002): Interactivity Types in Geographic Visualization. Cartography and Geographic Information Science, 29(2): S. 85-98

Daassi, C.; Nigay, L.; Fauvet, M.C. (2006): A taxonomy of temporal data visualization techniques. Revue en Sciences du traitement de l'Information, Cépaduès Editions, l5 (2): S. 41-63

Demsar, U.; Fotheringham, A.; Charlton, M. (2008): Exploring the spatio-temporal dynamics of geographical processes with geographically weighted regression and geovisual analytics. Information Visualization 2008, 7: S. 181-197

DiBiase, D.; MacEachren, A.M.; Krygier, J.; Reeves, C.; Brenner, A. (1991): Animated cartographic visualization in earth system science. In: Proceedings of the 15th International Cartographic Conference ICA, Bournemouth, 1: S.223-232

Dodge, M.; McDerby, M.; Turner, M. (eds.) (2008): Geographic Visualization. Concepts, Tools and Applications, West Sussex: John Wiley & Sons Ltd.

Dransch, D.; Rase, W.D. (2000): Software für die Erstellung von kartographischen Animationen. In: Buziek, G.; Dransch, D.; Rase, W.D. (eds.): Dynamische Visualisierung. Grundlagen mit Anwendungsbeispielen aus der Kartographie, Heidelberg: Springer

Edsall, R. (2003): The parallel coordinate plot in action. Design and use for geographical visualization, Computational Statistics & Data Analysis, 43: S. 605-619

Ellis, G.; Dix, A. (2006): Enabling Automatic Clutter Reduction in Parallel Coordinate Plots. IEEE Transactions on Visualization and Computer Graphics, 12(5)

Elsner, F. (1997): Visualisieren von Daten. Techniken und Programme, Rechenzentrum Universität Osnabrück

Fanea, E.; Carpendale, S.; Isenberg, T. (2005): An Interactive 3D Integration of Parallel Coordinates and Star Glyphs. Department of Computer Science University of Calgary, Canada

Frank, A. (1998): Different Types of "Times" in GIS. In: Egenhofer, M.; Golledge, R. (eds): Spatial and Temporal Reasoning in Geographic Information Systems. New York: Oxford University Press

Grandis, L.D. (1986): Theory and Use of Color. Prentice-Hall Inc., Englewood Cliffs, New Jersey

Grinstein, G.; Trutschl, M.; Cvek, U. (2001): High Dimensional Visualizations. In: Data mining conference KDD Workshop 2001

Guo, D.; Chen, J.; MacEachren, A.; Liao, K. (2006): A Visual Inquiry System for Space-Time and Multivariate Patterns (VIS-STAMP). IEEE Transactions on Visualization and Computer Graphics

Hackstadt, S; Malony, A. (1995): Visualizing Parallel Programs and Performance. In: IEEE Computer Graphics and Applications, 15(4): S.12-14

Hajnicz, E. (1996): Time Structures. Formal Description and Algorithmic Representation, Lecture Notes in Computer Science,1047, Berlin: Springer

Hake, G.; Grünreich, D.; Meng, L. (2002): Kartographie. Visualisierung raum-zeitlicher Informationen, 8.Auflage, Berlin: Walter de Gruyter

Hanna, A.; Remington, R. (1996): The representation of color and form in long-term memory. Memory and Cognition, 24: S.322-330

Hao, M.; Dayal, U.; Keim, D. (2008): Visual Analytics Techniques for Large Multi-Attribute Time Series Data. In: Börner, K.; Gröhn, M.; Park, J.; Roberts, J. (eds.): Visualization and Data Analysis 2008, Proceedings of SPIE and IS&T- The Society for Imaging Science and Technology, 6809 680908 (2)

Harris, R. (1999): Information Graphics. A Comprehensive Illustrated Reference, New York: Oxford University Press

Heinzel, W. (2003): Java3D. Unterlagen zum Seminar "Java3D, Hochschule Fulda, CAE Fachbereich für angewandte Informatik

Hochheiser, H.; Shneiderman, B. (2004): Dynamic query tools for time series data sets. Timebox widgets for interactive exploration, Information Visualization 2004, 3: S. 1-18

Holzhüter, C. (2005): Achsenbasierte 3D-Techniken zur Visualisierung zeitabhängiger Daten. Studienarbeit, Institut für Informatik, Universität Rostock

Inselberg, A. (1985): The plane with parallel coordinates. The visual Computer, 1: S. 69-91

Inselberg, A. (1990): Parallel Coordinates: A tool for visualizing multi-dimensional geometry, In: Proceedings of IEEE Conference on Visualization, Los Angeles: S. 361-378

Inselberg, A. (2009): Parallel Coordinates. Visual Multidimensional Geometry and Its Applications, Heidelberg London New York: Springer

Java (2003): 3D Graphik und Animation: Java3D. Vorlesungsunterlagen für das Hauptseminar "Virtuelle Präsenz", WS03/04, Institut für verteilte Systeme, Universität Ulm

Johansson, J.; Ljung, P.; Cooper, M. (2007): Depth Cues and Density in Temporal Parallel Coordinates. In: Museth, K.; Möller, T.; Ynnerman, A. (eds.): Data Visualization. Eurovis 2007, IEEE-VGTC Symposium on Visualization. Aire-la-Ville: Eurographics Association

Johansson, J.; Cooper, M.; Jern, M. (2005): 3-dimensional display for clustered multi-relational parallel coordinates. 9th International Conference on Information Visualization 2005 (London, UK), IEEE Computer Society Press: S.188–193

Johansson, J.; Ljung, P.; Jern, M.; Cooper, M. (2006): Revealing in visualizations of dense in 2D and 3D parallel coordinates. Information Visualization, 5(2): S.125-136

Keim, D. (2002): Datenvisualisierung und Data Mining. In: Datenbank Spektrum, 2: S. 30-39

Keller, T.; Grimm, M. (2005): The Impact of Dimensionality and Color Coding of Information Visualizations on Knowledge Acquisition. In: Tergan, S.; Keller, T. (eds.) (2005): Knowledge and Information Visualization. Searching for Synergies, LNCS 3426, Berlin: Springer-Verlag: S.167-182

Klawonn, F. (2009): Grundkurs Computergrafik mit Java. Die Grundlagen verstehen und einfach umsetzen mit Java3D, Wiesbaden: Vieweg und Teubner

Kojker, K.J (1987): The information capacity of the human fingertip. IEE Transactions on Systems, Man and Cybernetics, 17(1): S.100-102

Lange, S.; Nocke, T.; Schumann, H. (2006): Visualisierungsdesign. Ein systematischer Überblick, In: Simulation und Visualisierung 06: S. 113-128

Larson, K.; van Dantzich, M.; Czerwinski, M.; Robertson, G. (2000): Text in 3D. Some legibility results, CHI '00 extended abstracts on Human factors in computing systems, April 01-06, The Hague, The Netherlands

Lee, Ch. (2008): Visualization of multidimensional database. In: Börner, K.; Gröhn, M.; Park, J.; Roberts, J. (eds.): Visualization and Data Analysis 2008. Proceedings of SPIE and IS&T, The Society for Imaging Science and Technology, 6809 68090M (1)

MacEachern, A.; Taylor, D. (1994): Visualization in modern cartography. Oxford: Elsevier Science Ltd.

Mackinlay, J. (1986): Automating the design of graphical presentations of relational information. Transactions on Graphics, 5(2), New York: ACM

Miller, G.A. (1956): The Magical Number Seven, Plus or Minus two. Some Limits on our Capacity for Processing Information, Psychological Review, 63(3): S.81-97

Mosler, K.; Schmid, F. (2006): Beschreibende Statistik und Wirtschaftsstatistik. Heidelberg: Springer

Moustafa, R.; Wegman, E. (2002): On Some Generalizations of Parallel Coordinate Plot. Center for Computational Statistics, George Mason University, Fairfax

Müller, W.; Schumann, H. (2003): Visualization Methods for Time-Dependent Data. An Overview, In: Chick, S.; Sanchez, P.; Ferrin, D.; Morrice, D. (eds.): Proceedings of the 2003 Simulation Conference

Noirhomme-Fraiture, M. (2002): Visualization of Large Datasets. The Zoom Star –revolution, The Electronic Journal of Symbolic Data Analysis

Peuquet, D. (2002): Representations of Space and Time. New York: The Guilford Press

Robinson, A. (2007): A design framework for exploratory geovisualization in epidemiology. In: Information Visualization, 6: S.197-214

Ruth, Th. (2003): Möglichkeiten und Grenzen der automatischen Farbskalierung unter dem Gesichtspunkt der Vergleichbarkeit von Visualisierungen. Diplomarbeit im Fachbereich Informatik, Universität Rostock

Rübel, O.; Weber,G.H.; Huang, M.Y.; Bethel E.W; Biggin, M.D.; et al. (2008): Integrating Data Clustering and Visualization for the Analysis of 3D Gene Expression Data. IEEE Transactions on Computational Biology and Bioinformatics, LBNL-382E

Schumann, H. (1999): Visualisierung von Multiparameterdaten in Raum und Zeit. Eine Systematisierung, Proceedings "Simulation und Visualisierung '99, Society for Computer Simulation International, San Diego, Erlangen: S.377-395

Schumann, H.; Müller, W. (2000): Visualisierung. Grundlagen und allgemeine Methoden, Berlin Heidelberg: Springer

Shneiderman, B. (1996): The eyes have it. A task by data type taxonomy for information visualizations, in VL '96: Proceedings of the IEEE Symposium on Visual Languages, Washington, DC, USA: S.336

Siirtola, H. (2007): Interactive Visualization of Multidimensional Data. Academic Dissertation, University of Tampere, Departement of Computer Sciences, Tampere

Siirtola, H.; Laivo, T.; Heimonen, T. et al. (2009): Visual Perception of Parallel Coordinate Visualization. In: 13th International Conference Information Visualisation, IEEE Computer Society

Steinmetz, A. (2008): Darstellung multivariater Bilddaten. Diplomarbeit, Universität Bielefeld

Streit, M.; Ecker, R.; Österreicher, K.; Steiner, G.; Bischof, H.; Bangert, C.; Kopp, T.; Rogojanu, R.(2006): 3D Parallel Coordinate Systems. A New Data Visualization Method in the Context of Microscopy-Based Multicolor Tissue Cytometry, International Society for Analytical Cytology, Cytometry Part A, 69: S. 601-611

Theus, M. (2003): Interactive Data Visualization using Mondrian. Journal of Statistical Software, 7(11): S. 1-9.

Tominski, C.(2002): Visualisierungstechniken zur Analyse zeitlicher Abläufe über geographischen Daten. Diplomarbeit, Universität Rostock

Tominski, C.; Schulze-Wollgast, P.; Schumann, H. (2003): Visualisierung zeitlicher Verläufe auf geografischen Karten. Institut für Computergrafik, Fachbereich Informatik, Universität Rostock

Tominski, C.; Schulze-Wollgast, P.; Schumann, H. (2005): 3D Information Visualization for Time Dependent Data on Maps. In: Proceedings of the ninth International Conference of Information Visualisation 2005, 6(8): S. 175 - 181

Treisman, A.; Gormican, S. (1988): Feature analysis in early vision. Evidence from search asymmetries, Psychological Review, 95: S.15-48

Tufte, E. (1990): The Visual Display of Quantitative Information. 2nd edition, Cheshire, Connecticut: Graphic Press

Unwin, A.; Volinsky, C.; Winkler, S. (2003): Parallel Coordinates for exploratory modelling analysis. Computational Statistics & Data Analysis, 43: S. 553-564

Unwin, A.; Theus, M.; Hofmann, H. (2006): Graphics of Large Datasets. Visualizing a Million, New York: Springer Science & Business Media

Ware, C. (2004): Second Edition Information Visualization. Perception for Design, San Francisco: Morgan Kaufmann

Wegenkittl, R.; Löffelmann, H.; Gröller, E. (1997): Visualizing the Behavior of Higher Dimensional Dynamical Systems. Vienna University of Technology, Institute of Computer Graphics

Wegman, E. (1990): Hyperdimensional Data Analysis using Parallel Coordinates. Journal of the American Statistical Association, 85(411): S. 664-675

Williams, J. (2005): Interaktive Wanderkarte des Yosemite Nationalparks. Diplomarbeit am Institut für Kartographie, ETH Zürich

Wills, G. (1996): Selection: 524288 ways to say "this is interesting". In: InfoVis '96: Proceedings of the IEEE Symposium on Information Visualization, IEEE Computer Society: S. 54-61

Onlinequellen

CADauno, (2009): 3D modeling with NURBS surfaces. http://cadauno.sourceforge.net/features.php, Zugriff: 25.01.2010

Chi-Square (2001): Parallel Coordinate Plot. http://chi-square-works.com/panmo/features.html, Zugriff: 25.01.2010

Electronic Visualization Laboratory, (2001): Parallel Coordinates http://www.evl.uic.edu/aej/526/kyoung/Training-parallelcoordinate.html, Zugriff: 20.01.2010

Fachhochschule Wedel (2000): http://www.fh-wedel.de/~si/seminare/ws00/ Ausarbeitung /2.java3d/ java3d3.htm, Zugriff: 25.01.2010

Gut, J. (2009): Kursunterlagen Mathematik. http://members.chello.at/gut.jutta.gerhard/kurs/vektoren3.htm, Zugriff: 23.01.2010

Java (2010): Sun Developer Network. http://java.sun.com/javase/technologies/desktop/java3d/?intcmp=3177, Zugriff: 25.01.2010

Technische Universität Wien (1999): n-DimViewer: A tool for interactive exploration of high dimensional dynamical systems. http://www.cg.tuwien.ac.at/research/vis/dynsys/NDimViewer/, Zugriff: 25.01.2010

Technische Universität Wien (2006): Statistische Analysen mit Parallelen Koordinaten. http://www.cg.tuwien.ac.at/courses/Visualisierung/2005-2006/Beispiel2/Platzer/index.htm, Zugriff: 25. 01.2010

Universität Münster, (2007): Vorlesungsunterlagen zum Kurs Geoinformatik. Kapitel 8. http://ifgivor.uni-muenster.de/vorlesungen/Geoinformatik/kap/kap8/k08_1.htm, Zugriff 25.01.2010

Erklärung der Selbständigkeit

Appenzell, den 25. Januar 2010

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und die verwendeten Quellen und wörtlich oder inhaltlich entnommenen Stellen, als solche kenntlich gemacht habe.

Stefanie Hautle