

The top-left portion of the slide features a series of thin, light-brown lines that intersect to form various irregular polygons and shapes, creating a complex, abstract geometric pattern.

VISUALISIERUNG GROßER DATENMENGEN (BIG DATA CHALLENGE)

Uwe Renner, Vera Ogunlade und Manfred H. Wolff
Arnold-Sommerfeld-Gesellschaft e.V.



VISUALISIERUNG
GROßER DATENMENGEN

EINLEITUNG
GLIEDERUNG

Gliederung

1. Visuelle Wahrnehmung und Objektivierung
2. Herkunft großer Datenmengen
3. Visualisierungstechniken
4. Datenverarbeitung und Visualisierung
5. Einige Anwendungsgebiete
6. Visualisierung zeitlicher Verläufe
7. Visionen der Visualisierung
8. Literatur

1. Visuelle Wahrnehmung und Objektivierung

Die gewöhnliche **visuelle Wahrnehmung** von Objekten unterliegt der Kritik einer **subjektiven** Betrachtung, d. h. es besteht die Freiheit, Möglichkeit oder gar **Willkür**, dass Betrachter wahrgenommene Daten unterschiedlich interpretieren können und somit die Gefahr für Täuschungen und Irrtümer.

Um hingegen ein objektives Bild bzw. eine wissenschaftliche Bewertung der Daten zu erhalten, bedarf es einer „maschinellen“ Aufarbeitung, welche Mehrdeutigkeiten vermeidet.

Diese **Objektivierung kann durch mathematische Verfahren** erreicht werden, insbesondere durch in Computern ausgeführten Algorithmen.



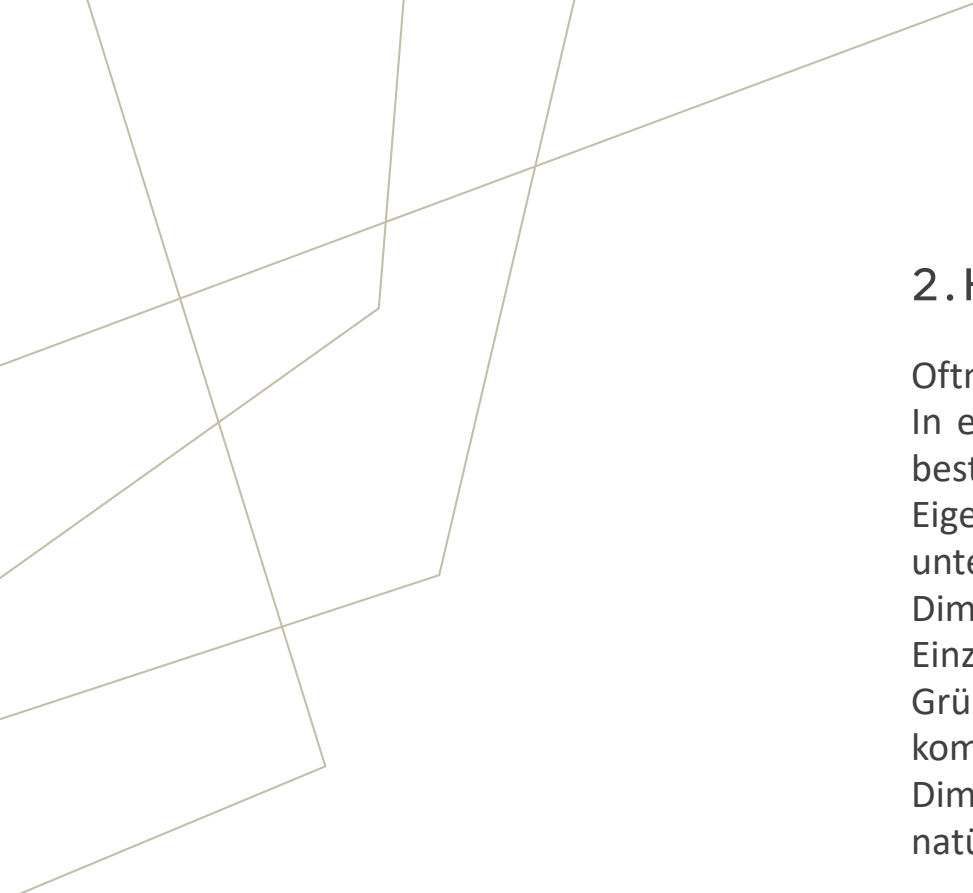
VISUALISIERUNG GROßER DATENMENGEN

2.HERKUNFT

2.Herkunft großer Datenmengen

Mit dem Voranschreiten der Leistungsfähigkeit der Computer und ihrer Vernetzung entstehen durch Simulationen von Modellen in den verschiedenen Wissenschaftsdisziplinen und in der Technik zunehmend mehr Daten, deren Ergebnisse nicht nur numerisch ausgewertet, sondern für das bessere Verstehen von Zusammenhängen und Eigenschaften visualisiert werden müssen.

Die Herausforderungen bei der Visualisierung großer Datenmengen ist auch eine Folge der Fortschritte bei der **Messdatenerfassung**. Insbesondere durch hochauflösende, empfindliche Sensorelemente, Sensorarrays oder Sensornetzwerke in Verbindung mit schnellen Verfahren der Digitalisierung analoger Daten und Speicherung können viele unabhängige Datenkanäle erfasst und zur Verarbeitung in Computern, Computernetzwerken bzw. Clouds zur Weiterverarbeitung bereitgestellt werden. Hierbei geht man von vielen gewonnenen Messdaten aus, die im Idealfall unabhängig sind und somit einen entsprechend der Anzahl der Datenkanäle **dimensionierten** Raum definieren.



VISUALISIERUNG GROßER DATENMENGEN

2.HERKUNFT

2.Herkunft großer Datenmengen

Oftmals beträgt die **Dimensionalität des Raumes mehrere 10er Potenzen**. In einem Spektrometer wird diese z. B. durch die Anzahl der Messkanäle bestimmt, wobei die Kanalwerte bzw. Intensitäten die physikalischen Eigenschaften der Substanzen charakterisieren, wodurch diese unterschieden werden können. In einem Bilderfassungssystem wird die Dimensionalität durch die Anzahl der intensitätsempfindlichen Einzelsensorelemente bestimmt, wobei jeder Spektralkanal (z. B. für Rot, Grün und Blau) die Dimensionalität vervielfacht. Bei Bewegtbildern kommen die einzelnen Zeitpunkte der Bilderfassung als zusätzliche Dimension hinzu, auch wenn diese wegen der Kausalität der Bildfolge bei natürlichen Prozessen dann nicht mehr in der Regel völlig unabhängig sind.

Oftmals reicht die Erfassung eines Messereignisses nicht aus und es sind aufgrund von Schwankungen **Mehrfachmessungen** notwendig, die gleichfalls die Datenmenge zur Erreichung einer hohen Messgenauigkeit vergrößern.

3. Visualisierungstechniken

Die einfachste Art Ergebnisse, Ereignisse oder Entscheidungen zu visualisieren, besteht durch die **binäre Signalisierung** bzw. Alarmierung durch Lampen bzw. LEDs und durch die alphanumerische Anzeige auf LCDs. Mehr Informationen lassen sich als zweidimensionale Abbildungen z. B. auf einem graphischen Display darstellen oder mittels Projektoren präsentieren. Hiermit werden jedoch die Möglichkeiten durch den Gesichtssinn nur teilweise ausgenutzt, denn dieser bietet durch die binokulare Anordnung der Augen die Möglichkeit eines stereoskopischen Sehens. Einen dreidimensionalen Rundumeindruck könnte durch Hologramme verschafft werden. Andere visuelle Hilfsmittel in Gestalt von Brillen bieten Möglichkeiten, die biologischen Möglichkeiten durch zusätzliche Informationen zu erweitern oder virtuelle Datenwelten (z. B. durch Virtual-Reality-Brillen) nicht nur zu erfassen, sondern auch interaktiv zu gestalten.

Zur Darstellung **komplexer Zusammenhänge**, insbesondere von statistischen Daten, so z. B. durch Massenspektrometrie gewonnene Spektren [1] oder Bildinformationen, lassen sich diese zunächst als Punkte abstrahierten Daten in einem Koordinaten-Raum entsprechend ihrer Verwandtschaft oder Ähnlichkeit oftmals als Punktwolken bzw. Clustern zusammengefasst darstellen und gegebenenfalls in unterschiedliche Cluster mittels Verfahren der **multivariaten Statistik** trennen. Somit sind eine objektivierte, automatische Klassifizierung sowie die Benennung mit Marken oder Labels der Ursprungsdaten möglich, auch wenn der Ursprung der Daten bekannt ist.

VISUALISIERUNG GROßER
DATENMENGEN

VISUALISIERUNGSTECHNIKEN



3. Visualisierungstechniken

Eine Visualisierung dieser abstrakten Cluster kann durch Projektionen auf einem graphischen Display erfolgen. Umgekehrt ermöglicht dies bei einer großen Datenbasis auch unbekannte Daten bekannten Clustern zuzuordnen (s. Mustererkennung, Identifizierung).

Werden Daten mit einem geometrischen Kontext ausgewertet, so können diese mit Zusatzinformationen wie Koordinaten oder Abstandsinformationen versehen werden (Tomographie, Kartierung, Mapping). Bei geographischen Bezügen lassen sich die GPS-Positionsdaten oder UTC-Zeitdaten den einzelnen Orten zuordnen. Auch ist es möglich, Zusatzdaten wie etwa Identifikationsmerkmale, die aus einer Mustererkennung einer realen Umgebung gewonnen wurden, mit dieser in einer erweiterten Realität (engl. augmented reality) zu vereinen. Als Bereicherung können zur visuellen Wahrnehmung andere Reize wie akustische Informationen hinzugenommen werden, um bestimmte Eigenschaften zu charakterisieren.

Weitere Möglichkeiten der Visualisierung ergeben sich durch die Darstellung von Daten in abstrakten Landschaften bzw. virtuellen Realitäten mit entsprechenden Hilfsmitteln wie VR-Brillen, die zusätzlich eine Interaktion etwa im Spiel bzw. in einer Simulation ermöglichen und somit Daten besser erfassbar werden lassen.

VISUALISIERUNG GROßER
DATENMENGEN

VISUALISIERUNGSTECHNIKEN



VISUALISIERUNG
GROßER DATENMENGEN

4.DATENVERARBEITUNG
UND VISUALISIERUNG

4. Datenverarbeitung und Visualisierung

Ein **Ziel der Visualisierung besteht darin, verborgene Informationen** in der Menge der Daten sichtbar zu machen und wesentliche Information vom Unwesentlichen wie dem Rauschen zu trennen [1]. Bis zur Visualisierung der interessierenden Messdaten sind viele Bearbeitungsschritte erforderlich und werden erst als Ergebnis der Vorverarbeitung mittels Verfahren der Signal- und Datenverarbeitung sichtbar. Hierzu gehören insbesondere Methoden zur Datenfilterung. Die Gewinnung wesentlicher Bildinformationen erfordert oftmals die Erzeugung von Kontrasten oder die Datenseparation mittels Schwellwerten oder Kantenhervorhebungen oder das Einfärben von Gebieten. Zur Visualisierung lassen sich bestimmte Merkmale zusätzlich markieren.

Manche Informationen, die oft im Verborgenen liegen, werden erst durch **mathematische Transformationen** sichtbar. Man bedient sich derer insbesondere, um Eigenschaften wie Periodizität oder andere Merkmale kenntlich zu machen (FOURIER-, Wavelet- u. a. Transformationen).

Bei den Verfahren führen **numerischen Instabilität sowie die beschränkten Stellengenaugigkeit** der Zahlendarstellung im Computer, der Rundungsfehler und der Fehlerakkumulationen zu weiteren Problemen



VISUALISIERUNG GROßER DATENMENGEN

4. DATENVERARBEITUNG UND VISUALISIERUNG

4. Datenverarbeitung und Visualisierung

Zur eigentlichen Auswertung ist vielfach eine **Datenreduktion** einhergehend mit einer **Verringerung der Datenraumdimension** mittels Projektionen notwendig, so etwa der Hauptkomponentenanalyse. Vielfach bedient man sich der multivariaten Statistik, um Relation der Daten untereinander zu bestimmen, so die gegenseitige Abhängigkeit oder Überbestimmtheit (Redundanz) der Daten durch Dimensionsanalyse oder durch statistischer Test-Verfahren. So lassen sich mittels Korrelationsanalysen gegenseitige Abhängigkeiten herausfinden. Gleichartige oder ähnliche Eigenschaften vereinen sich oftmals zu Clustern (Wolken), die von anderen durch den gegenseitigen Abstand mittels mathematischer Verfahren, Algorithmen der künstlichen Intelligenz basierend auf neuronalen Netzen oder anderen Mustererkennungsverfahren unterschieden und im Weiteren visualisiert werden können.

Sowohl die Auswertung, aber auch **die Visualisierung von großen Datenmengen sind oft speicher- und zeitaufwendig**. Zur Erreichung einer zeitnahen Darstellung ist eine schnelle Verarbeitung in den verschiedensten Ebenen der Datenverarbeitung erforderlich, welche durch eine Vielzahl schneller parallel arbeitender Prozessoren erreicht wird, angefangen beim Mainboard-Prozessor, der selbst aus vielen Rechenkernen besteht, bis hin zum Grafikprozessor, der wiederum aus mitunter hunderten Berechnungseinheiten besteht. Zudem lassen sich Prozessoren in Netzwerken zu noch größeren Einheiten verbinden.



VISUALISIERUNG GROßER
DATENMENGEN

ANWENDUNGSGBIETE

5. Einige Anwendungsgebiete

Von großem Interesse und mit zahlreichen Aufgabenstellungen sind die Anwendungen der Visualisierung in der **Medizin** [2]. Ein hier verwendetes Verfahren, um in das Innere des Körpers blicken zu können, ohne diesen zu zerteilen und dennoch ein Gesamteindruck vom Aufbau zu erhalten, ist die **Tomographie, so z. B. die auf RÖNTGEN-Strahlen basierte Computertomographie (CT) oder die in einem starken Magnetfeld auf dem Kernspin basierte Magnetresonanztomographie (MRT)** mittels elektrischer Hochfrequenzfelder. Dabei ist die gemessene Intensität vom Absorptionsvermögen des Materials entlang des Strahlungsweges abhängig. Aus vielen derartiger Messungen lassen sich zunächst zweidimensionale Schnittbilder und in deren Schichtung 3D-Ansichten rekonstruieren. Mittels bildgebender Verfahren, welche die Kenntnisse zur Dichte des Materials verwenden, können äußerlich unsichtbare Objekte visualisiert, vermessen und bewertet werden, s. Abbildung. In ähnlicher Weise kann man durch **Ultraschall bei medizinischen Untersuchungen oder zur Werkstoffdiagnostik, RADAR (engl. radio detection and ranging)** oder vernetzter Widerstandsmessungen in der Geologie **unter die Oberfläche blicken und verborgene Strukturen sichtbar machen.**

5. Einige Anwendungsgebiete



Abb. Einer MRT Rekonstruktion im häutigen Labyrinth des Gleichgewichtsorgans.

VISUALISIERUNG GROßER
DATENMENGEN

ANWENDUNGSGEBIETE

VISUALISIERUNG GROßER
DATENMENGEN

ANWENDUNGSGBIETE

5. Einige Anwendungsgebiete

In der **Biologie** werden zur Bestimmung der Verwandtschaft bzw. Ähnlichkeit verästelte Bäumen bzw. Dendrogramme verwendet, wobei die Information zur Visualisierung der Abhängigkeiten in Graphen oder Netzwerke z. B. aus der molekularen Zusammensetzung der RNA stammen und mittels Massenspektrometrie gewonnen wird und die Verwandtschaftsbestimmung über nicht-EUKLIDISCHE Metriken erfolgt. [1]

Die Photogrammetrie [3 - 5] ermöglicht es, aus vielen überlappenden 2D-Einzelbildaufnahmen ein dreidimensionales Oberflächen- bzw. Reliefbild eines Objektes durch mathematische Verfahren zu rekonstruieren, das man schließlich aus verschiedenen Perspektiven und Vergrößerungen betrachten und beurteilen kann. In ähnlicher Weise ermöglicht das LIDAR-Verfahren (engl. light detection and ranging) nicht nur Objekte, sondern auch Landschaften mittels vieler gepulster Laserstrahlen über Laufzeitmessungen in kurzer Zeit zu erkunden, wobei diese rasterartige abgetastet wird. Da ein solcher Laserstrahl gegenüber photographischen Aufnahmen eine hohe Auflösung besitzt, kann man geringste Unebenheiten erfassen und z. B. durch die Blätter eines Waldes bis zum Boden vordringen. Nach Filterung der Daten hinsichtlich der Laufzeiten, lässt sich speziell die Bodenebene berechnen und diese am Computer so visualisieren, so dass eventuelle Geländeanomalien sichtbar werden. Die LIDAR-Daten können auch dazu dienen, großräumige Computer-Modelle von zerfallenen historischen Architekturen zu erstellen und diese virtuell zu rekonstruieren. Analog zum LIDAR ist es mit modernen Schalldiagnose-Verfahren wie dem Seitensicht-SONAR (engl. sound navigation and ranging) möglich, unter die Wasseroberfläche bis auf den Untergrund blicken und verdächtige Muster ausfindig machen. Derartige Verfahren finden z.B. Anwendung in der **Archäologie** oder Schatzsuche. [6 - 8]

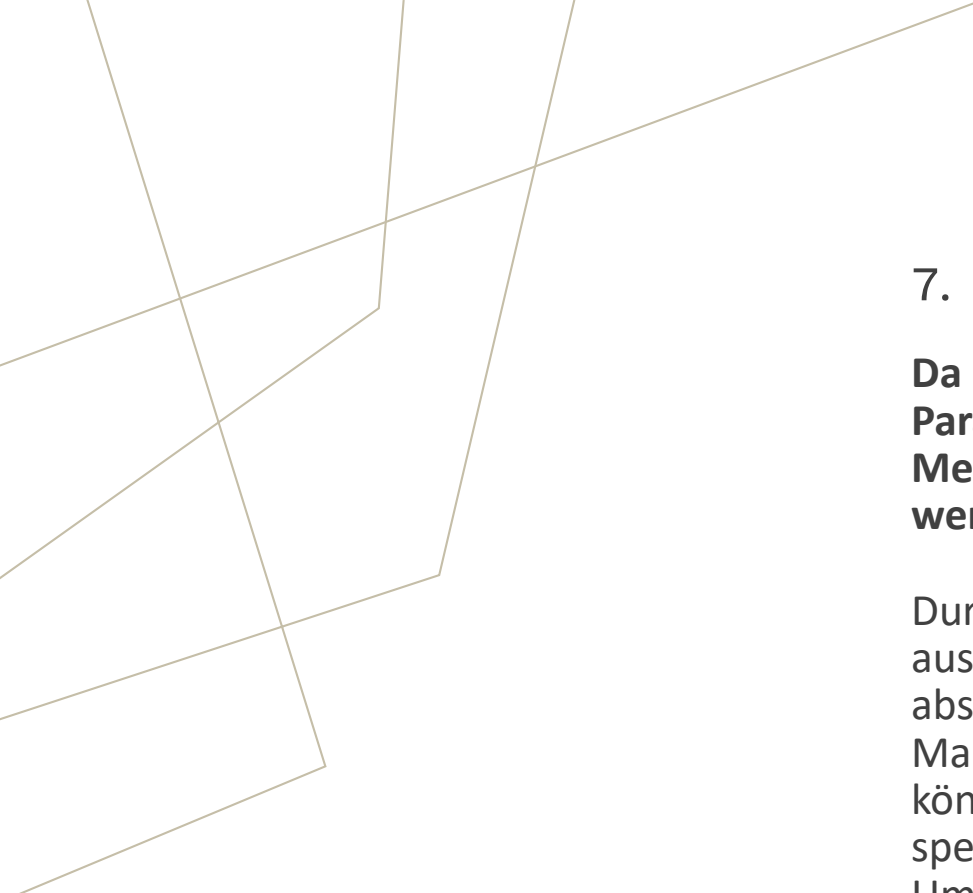


VISUALISIERUNG
GROßER DATENMENGEN

6.V... ZEITLICHER
ABLÄUFE

6. Visualisierung zeitlicher Abläufe

Auch wenn zumeist statische, d. h. zeitlich unveränderliche Daten, zu bestimmten Zeitpunkten von Interesse sind, gibt es viele Situationen in denen sich Bilder zeitlich mehr oder weniger schnell ändern. Dann bietet die Visualisierung zeitlicher Abläufe als Animation bzw. Film, speziell in Zeitraffer oder als Zeitlupe, eine Möglichkeit, Informationen von Prozessen wie Strukturbildung, Wachstum, Morphogenese oder von Bewegungsabläufen in der Biologie (Ganganalyse, Flug) zu erhalten. Hierzu ist es oft nicht nötig, die abgetasteten Einzelbilder auszuwerten, sondern man verfolgt **repräsentative Bewegungsänderungen mittels Marker oder vektorielle Richtungsänderungen.**



VISUALISIERUNG
GROßER DATENMENGEN

VISIONEN

7. Visionen zur Visualisierung

Da mittels schneller Computersysteme die Auswirkungen von Parameteränderungen sowohl bei der Datenerfassung bzw. Messung als auch bei der Auswertung in Echtzeit visualisiert werden können, ist es möglich, diese gleichzeitig zu optimieren.

Durch Visualisierung können auch abstrakte Daten oder Formen aus Simulationen zugänglich gemacht werden. Hierzu gehören abstrakte mathematische Räume wie mehrdimensionale Mannigfaltigkeiten und Objekte der Differentialgeometrie. So können Bewegungen in der Raum-Zeit wie sie sich durch die spezielle oder der allgemeinen Relativitätstheorie z. B. in der Umgebung eines schwarzen Loches nahe dem SCHWARZSCHILD-Radius als Science-Fiction, hier eine spezielle Form der Visualisierung, beschrieben werden. Aber auch die abstrakte visualisierte Bewegung der Wellenfunktionen im HILBERT-Raum kann ein besseres Verständnis über die zugrunde gelegte physikalische bzw. mathematische Welt ermöglichen.

VISUALISIERUNG GROßER DATENMENGEN

LITERATUR

8.Literatur

[1] Renner, U.: *Beiträge zur Informationsverarbeitung in der Spektroskopie*. In: Synergie – Syntropie – Nichtlineare Systeme, Heft 7: Physik – Statistik – Information. Leipziger Universitätsverlag, 2008. S. 138 ff.

[2] Wolff, M.: *Medizin und Technik – Rechnergestützte Durchführung, Auswertung und Dokumentation von medizinischen Aufgaben und klinischen Studien*. In: Synergie – Syntropie – Nichtlineare Systeme, Heft 3: Softcomputing. Verlag im Wissenschaftszentrum Leipzig, 2000. S. 97 ff.

[3] Henriksen, S. W.: *Mathematical Photogrammetry*. https://www.asprs.org/wp-content/uploads/pers/1965journal/jul/1965_jul_720-726.pdf

[4] Derenyi, E. E.: *Photogrammetry: The Concept*. Department of Geodesy and Geomatics Engineering, University of New Brunswick. Lecture Notes No. 57. September 1996

[5] Lussu, P.: *Ultra close-range digital photogrammetry as a tool to preserve, study, and share skeletal remains*. PHD thesis, Università degli Studi di Cagliari, 2020

[6] Estornell, J.: *Mathematical modelling applied to LiDAR data*. Modelling in Science Education and Learning. Volume 6(3), No. 10, 2013.

[7] Hesse, R.: *Lidar-Anwendung in der Archäologie: Grundlagen, Visualisierung und Interpretation*. https://www.academia.edu/8603039/Lidar_Anwendung_in_der_Arch%C3%A4ologie_Grundlagen_Visualisierung_und_Interpretation

[8] Meyer, M. F. et al: *Die automatisierte Auswertung von LiDAR-Daten in der Archäologie*. 37. Wissenschaftlich-Technische Jahrestagung der DGPF in Würzburg – Publikationen der DGPF, Band 26, 2017



VISUALISIERUNG
GROßER DATENMENGEN

ZUSAMMENFASSUNG

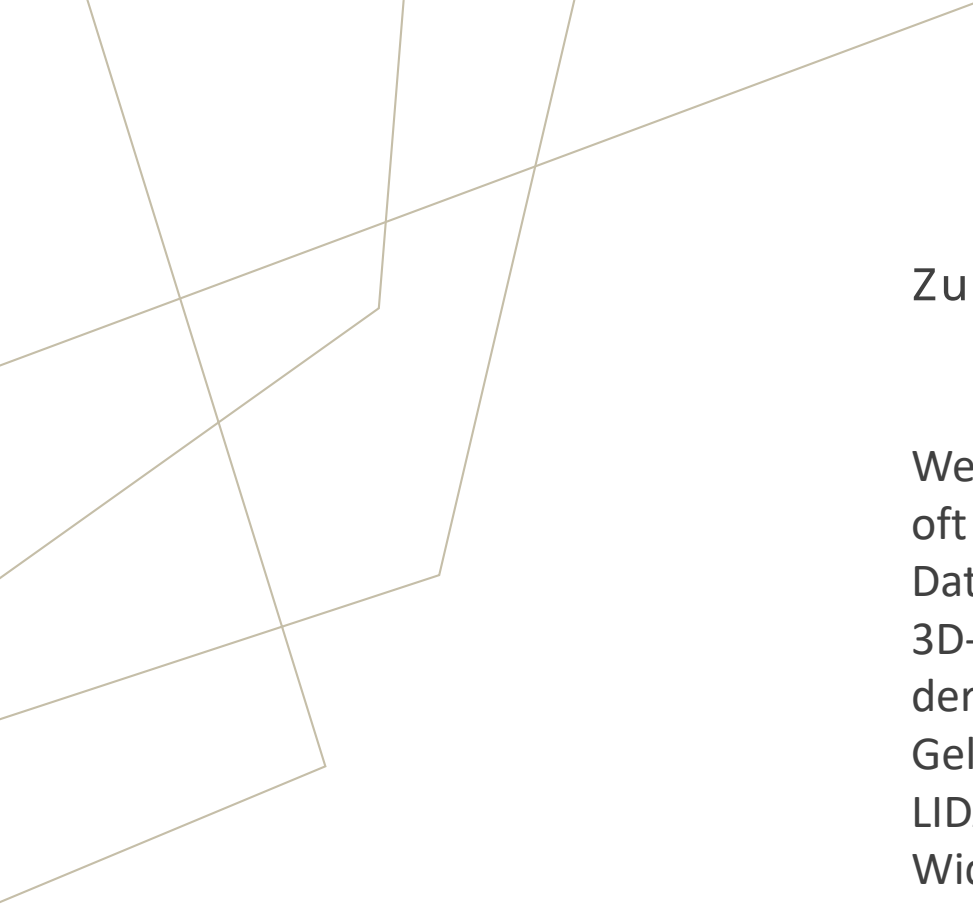
Zusammenfassung

Im Mittelpunkt dieses Überblicks steht das Sichtbarmachen, d. h. die Visualisierung, von großen Datenmengen und weniger die subjektive Ansicht bzw. Wahrnehmung durch einen Beobachter.

Insbesondere werden Methoden betrachtet, mit denen aus großen Datenmengen Informationen gewonnen werden, die für die (menschlichen) Sinne oftmals unsichtbar sind und z. B. ihren Ursprung in modernen Messverfahren und den Fortschritten der physikalischen Sensorik haben.

Deren Visualisierung wird erst durch optimierte mathematische Modelle und numerische Algorithmen, vielfach durch parallele Rechenverfahren in Verbindung mit entsprechender Computer-Hardware ermöglicht.

Große Datenmengen entstehen auch als Ergebnis von Computersimulationen, deren Grundlage Modelle der Natur in oftmals abstrakten Räumen sind und zum besseren Verständnis nicht nur räumlich, sondern manchmal auch zeitlich visualisiert werden müssen.



VISUALISIERUNG
GRÖßER DATENMENGEN

ZUSAMMENFASSUNG

Zusammenfassung

Wegen der vielfältigen Anwendungen und der verwendeten, oft ähnlichen Methoden ist die Visualisierung großer Datenmengen ein interdisziplinäres Thema. Beispiele sind die 3D-Rekonstruktion tomographischer Aufnahmen (CT, MRT) in der Medizin, Anatomie oder Biologie, die Erkennung von Gelände-anomalien und Strukturen in der Archäologie mittels LIDAR, Seitensicht-SONAR oder geologischer Widerstandsmessung von Bodenreliefs sowie die Photogrammetrie.



VIELEN DANK

info@renner-uwe.de

manfredh.wolff@t-online.de

www.asg-ev.org