
Skript zur Vorlesung

Spatial, Temporal and Multimedia Databases II

Skript zur Vorlesung: Spatial, Temporal, and Multimedia Databases II
Wintersemester 2011/12, LMU München

© 2011 PD Dr. Matthias Renz

Vorlesungsteam

Vorlesung (Plenum):



Priv.-Doz. Dr. Matthias Renz
Oettingenstr. 67, Zimmer E 1.11
Tel. 089/2180-9331
Sprechstunde: Montag, 14⁰⁰-15⁰⁰

Übungsbetrieb:

Johannes Niedermayer
Übungsgruppenleiter
Oettingenstr. 67, Zimmer E 1.11
Tel. 089/2180-9331

Tutoren:
Verena Link
Gregor Josse

Termine

- Vorlesung: Freitag, **09:30**–12:00 Uhr Raum U 127 (Oettingenstr. 67)
- Übung: Montag, 14-16 Uhr Raum 106 (Amalienstr. 73)
Montag, 16-18 Uhr Raum 209 (Schellingstr. 3 (R))
(*Start des Übungsbetriebes: 14.11.2011*)

Anmeldung für den Übungsbetrieb auf der Homepage

www.dbs.informatik.uni-muenchen.de/cms/Spatial,_Temporal_and_Multimedia_Databases_II

Schein-/Punkteerwerb

- Zulassung: Anmeldung für den Übungsbetrieb (siehe oben)
- Prüfung: Klausur

Kapitel 1

Einführung

Skript zur Vorlesung: Spatial, Temporal, and Multimedia Databases II
Wintersemester 2011/12, LMU München

© 2011 PD Dr. Matthias Renz

Inhalt der Vorlesung

- Methoden zur effizienten Bearbeitung von Ähnlichkeits- und Nachbarschaftsanfragen in „modernen“ STM-Datenbanksystemen
- Objekte mit räumlicher und zeitlicher Information, sowie Multimedia-Daten
- Die in der Vorlesung behandelte Themen:
 - Objekte die sich im Raum bewegen
(spatio-temporal objects, moving objects)
 - Unsichere Objekte
(Unsicherheit bzgl. der Existenz und/oder bzgl. der Attribute)
 - Sensornetzwerkdaten, d.h. Daten die aus einem Sensornetzwerk stammen
 - Multimedia-Objekte

Abgrenzung

Nicht besprochen werden:

- Methoden zur Erfassung und Integrierung der Daten in entsprechende Datenbanksysteme (DBS) / Datenbankmanagementsysteme (DBMS)
- Einbettung und Modellierung der Datentypen nach bestimmten DB-Modellen, wie z.B. das Relationale- / Objekt-Relationale-, sowie Objektorientierte Datenbankmodell
- Umsetzung der Anfragen in gängige Datenbanksystem-Sprachen (DML/DDDL) wie SQL

Relevante Themen aus STMD I

Themen die bereits in STMD I besprochen wurden:

- Grundprinzipien der Feature-basierten Ähnlichkeitssuche
 - Basisalgorithmen für Ähnlichkeits- und Nachbarschafts-Anfragen in multidimensionalen Vektorräumen (statische Punktdaten)
 - Modellierung/Verwaltung/Anfrage für
 - räumlich ausgedehnte Objekte
 - Zeitreihen
- relevant*
- nicht
relevant*

Anmerkung: STMD II erweitert die in STMD I behandelten Themen und baut teilweise darauf auf. Die bereits in STMD I behandelten Methoden/Techniken die zum Verständnis der in STMD II weiterführenden Techniken notwendig sind werden (zumind. kurz) wiederholt.

1.1 Motivierende Beispiele

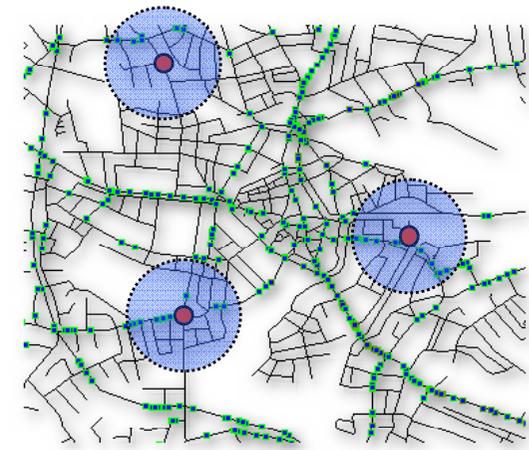
Beispiel 1: Location-Based Services (LBS)

Definition: Standortbezogene Dienste (Location-Based Services) = Dienste, die unter Zuhilfenahme von positions-, zeit- und personenabhängigen Daten dem Endbenutzer selektive Informationen bereitstellen [Wikipedia]

– Gegeben:

- Datenbank von sich bewegenden Objekten (z.B. Autos in einem Strassennetz)
- Menge von ausgezeichneten Positionen (z.B. Orte von Interesse wie z.B. Restaurants, etc.)

– Gesucht: alle Autos die sich in der unmittelbaren Nähe von Tankstellen der Firma „Tankgut“ befinden



– Herausforderung

- Position der Objekte ändert sich ständig
 - Spezielle Methoden zur Konsistenzerhaltung der Datenbank
- Als Distanz wird bei Straßennetzwerken oft die Netzwerkdistanz verwendet
- Behandlung von unsicherer/nicht exakter Information (z.B. Objekte mit unsicheren Positionen)

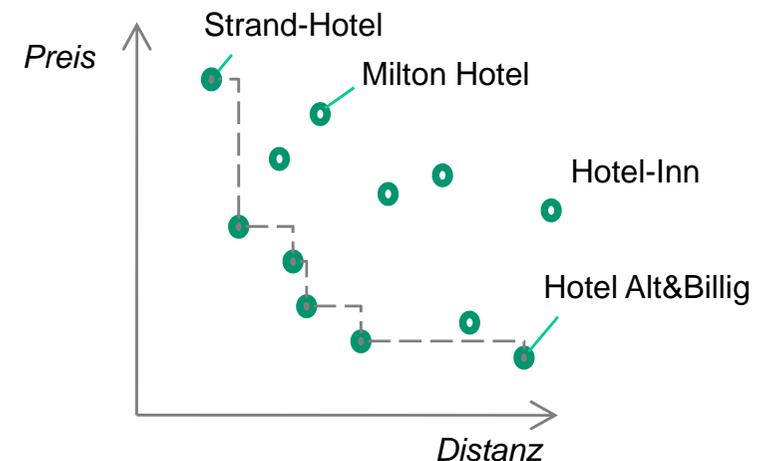
Beispiel 2: Pfadsuche / Navigation

Gegeben: Datenbank mit Verkehrsnetzwerkdaten (Straßenkarte, Positionen von interessanten Orten, Geschwindigkeitslimitierungen, etc.)

- Gesucht: Optimale Route von einem Startpunkt S zu einem Zielpunkt Z, die zuerst an einem Bankautomaten und danach einem Eisladen vorbeigeht.
- Herausforderungen:
 - Große Suchregion (je nach Abstand zwischen S und T)
 - Berücksichtigung mehrerer Optimierungskriterien (Zeit, Distanz)
 - Limitierter Speicherplatz (Einsatz in „Embedded-Systems“ z.B. Navi)
 - Berücksichtigung von dynamischen Attributen (Attributswerte die sich mit der Zeit ändern, Verkehrslage)

Beispiel 3: Empfehlungssysteme (Recommendation Systems)

- Gegeben: Datenbank mit (Multi-Attribut-) Objekten wie z.B. Produkte (Filme, Elektronische Geräte, Bücher, etc.) oder interessante Standorte (Hotels, Restaurants, Sehenswürdigkeiten, etc.)
- Gesucht: Alle Hotels die sich möglichst in der Nähe vom Strand befinden und möglichst preiswert sind?



– Herausforderungen:

- Oft negativ-korrelierte Optimierungsvariablen (Attribute) (z.B. Preis und Lage)

=> Anstatt eindeutige Antwort wird eine Liste mit unterschiedlichen Empfehlungen zurückgegeben

- Benutzer haben unterschiedliche Präferenzen:

1) Unterschiedliche Gewichtungen der Attribute

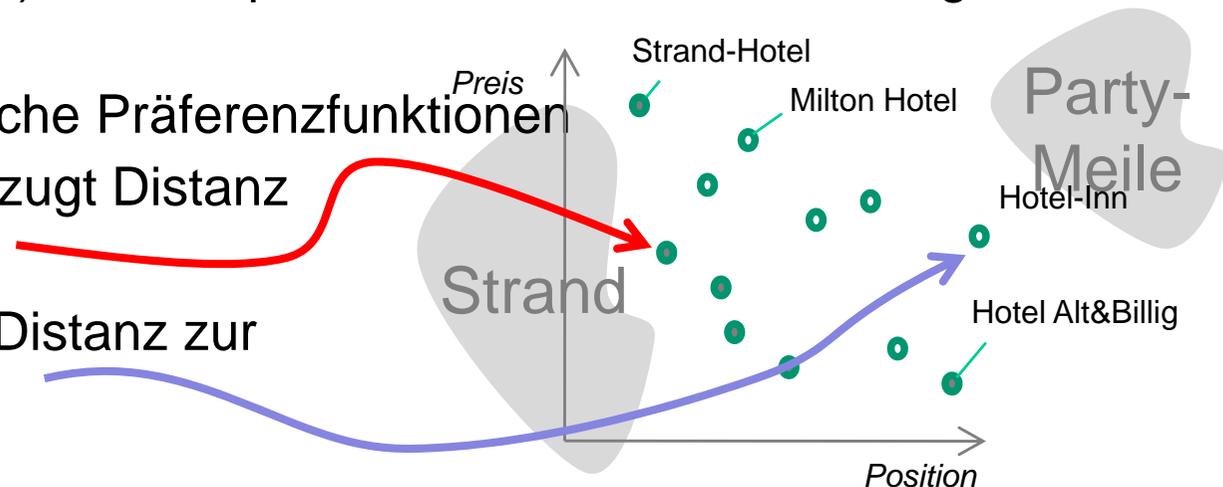
Fritz Steinreich (B1) „Lieber teurer, dafür näher am Strand“

Student Kurt (B2) „Für ein paar Schritte mehr, zahle ich gerne nur die Hälfte“

2) Unterschiedliche Präferenzfunktionen

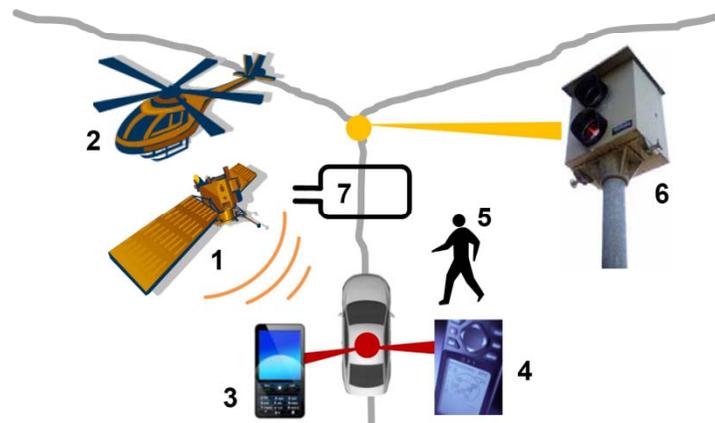
z.B.: B1 bevorzugt Distanz zum Strand

B2 bevorzugt Distanz zur Partymeile



Beispiel 3: Sensorgestützte Erkundungssysteme (Environmental Monitoring)

- Gegeben: Menge von Sensoren die
 - jeweils kontinuierlich (oder in regelmäßigen Zeitabständen) Zustände von Prozessen aufnehmen und
 - über ein kabelloses Netzwerk miteinander verbunden sind um miteinander zu kommunizieren und Daten auszutauschen.



- Gesucht: Kontinuierliche Ausgabe (monitoring) der durchschnittlichen Temperatur und Luftfeuchtigkeit

- Herausforderungen:
 - Hoch dynamische Daten
 - Sehr große Datenfluten
 - Ressourcen-limitierte Sensorgeräte
 - > wenig Speicherplatz
 - > Minimierung der CPU Kosten
 - > Minimierung der Kommunikationskosten

- Weitere Anwendungen
 - Objektverfolgung
 - Verkehrsanalyse
 - Verkehrskontrollsysteme

Problematiken bei der Suche (in großen Datenbeständen) [STMD I]

- Sequentielle Suche („sequential scan“)
 - Vergleich des Anfrageobjekts mit jedem einzelnen Datenbankobjekt
 - Skaliert *linear zur Größe der Datenbank, d.h. 100-mal mehr Objekte*
 - => 100-mal längere Suchzeit
 - => für große Datenbanken dauert Suche „viel zu lange“
- Herausforderungen
 - Beschleunigung der Suche (geschickte Datenorganisation)
 - **Indexierung**
 - Beschleunigung der Einzelvergleiche (geeignete Repräsentationen)
 - **Mehrstufige Anfragebearbeitung**

1.2 Inhalt der Vorlesung

1. Einführung
2. Prinzipien der Anfragebearbeitung in STM-DBS
(inkl. Wiederholung der wichtigsten Konzepte aus STMD I)
 1. Anfragemethoden für Räumlich-Zeitliche Daten
 2. Anfragebearbeitung in unsicheren Datenbanken
 3. Datenverwaltung und Anfragebearbeitung in Sensornetzwerken
 4. Ähnlichkeitssuche und Objektidentifikation in Multimediatdaten

1.3 Warum STMM-Datenbanken [STMD I]

- Spatial-, Temporal-, Multimedia-Objects
 - Irreversibler Trend in der IT
 - Neue Qualität von Informationen
 - Beschreibung von komplexen Strukturen, dynamischen Vorgängen, etc.
 - i.A. geringerer Informationsverlust wenn Verarbeitung im jeweiligen „Medium“ statt Umsetzung in ein anderes Medium

- Warum Datenbank-Technologie für STM?
 - Sehr große Mengen an Daten vorhanden
 - Speicherplatzintensive Daten
 - In vielen STM-Anwendungen ist Mehrbenutzerbetrieb erwünscht
 - Daten sollen (effizient und effektiv) recherchierbar sein

– Standard-DBS

- Konsistenzerhaltender Mehrbenutzerbetrieb
- Physische und logische Datenunabhängigkeit
- Effiziente Anfragebearbeitung durch geeignete Speicherungsstrukturen
- Unterstützung von Transaktionen
 - Concurrency: Isolation gleichzeitiger Updates verschiedener Benutzer
 - Recovery: konsistentes Wiederaufsetzen im Fehlerfall
 - Überwachung der Datenintegrität
- Datensicherheit, Datenschutz

– STM-DBS

- Verschiedene Verwendungen des Begriffs, z.B. für:
 - CD Sammlungen, die Infos mittels Stichwortsuche zugreifbar machen
 - Systeme zur Organisation und Sichtung von Informationen mittels Browser (z.B. WiKi)
 - Video-on-demand-Systeme
 - CAD-Systeme, die DBS nutzen
 - Relationale DBS, die zusätzlich sog. BLOBs (Binary Large Objects) speichern
- In dieser Vorlesung ganz allgemein ein DBS
 - mit hoher Kapazität und Performanz
 - das Spatial-, Temporal-, und Multimedia-Datentypen sowie alphanumerische Datentypen unterstützt
 - das mit großen Datenvolumina umgehen kann

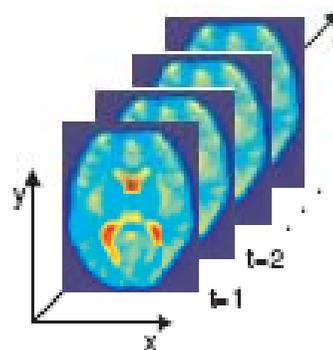
1.4 Anwendungen

1.4.1 Allg. Anwendungen mit räumlich-zeitlichen Daten

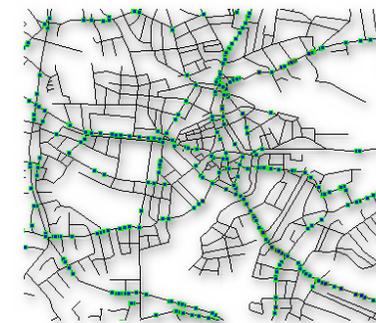
- Prinzipiell alle Anwendungen bei denen Ort- und Zeitinformationen eine Rolle spielen.
- Ort kann auch Position in einem allg. Merkmalsraum (Feature-Raum) sein.



a) Zeitreihen



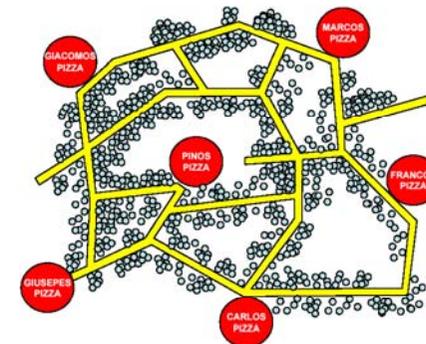
b) CT-Screen-Monitoring



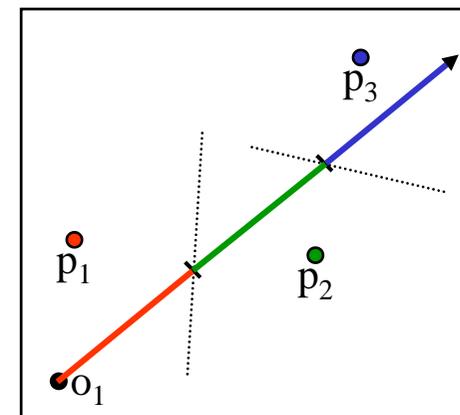
b) Positionen von Autos

1.4.2 Anwendungen mit beweglichen Objekten

- Analyse von Objekte in einem Verkehrsnetz
 - Location-based Services, Verkehrsplanung und -monitoring
 - Objekte, die sich entlang eines Verkehrsnetzes bewegen
 - Autos, Fußgänger, Züge, ...
 - Modellierung
 - Verkehrsnetz \equiv (evtl. gerichteter) Graph
 - Knoten: Kreuzungen, ...
 - Kanten: Verbindungen (Straßen, ...)
 - Objekte auf Kanten oder Knoten platziert
 - Anfragen/Recherche
 - Ähnlichkeit (bzgl. der Lage) zwischen Objekten über Netzwerkdistanz (Dijkstra-Algorithmus)
 - Suche nach räumlich nahen Objekten
 - Suche nach Objekten mit ähnlichen Bewegungspatterns



- Objekte die sich in einem Euklidischen Raum bewegen
 - Raumüberwachung, Chaosforschung, ...
 - Modellierung
 - Räumliche Koordinaten
 - Bewegung als Zeitreihe, zu jedem Zeitpunkt Informationen über
 - » Richtung (linear/nicht-linear)
 - » Geschwindigkeit
 - Anfragen/Recherche
 - Suche nach räumlich nahen Objekten
 - Suche nach Objekten mit ähnlichen Bewegungen
 - Suche nach räumlich benachbarten Objekten mit ähnlichem Bewegungsmuster



- Pfadsuche in Verkehrsnetzwerken
 - Transportationplanning, Navigation, ...
 - Gegeben:
 - Transportationsnetzwerk/Strassennetzwerk mit unterschiedlichen Strassenattributen, wie z.B.:
 - » Zeit
 - » Länge
 - » Anzahl von Ampeln
 - » etc.
 - Anfragen/Recherche
 - Suche nach optimalen Weg

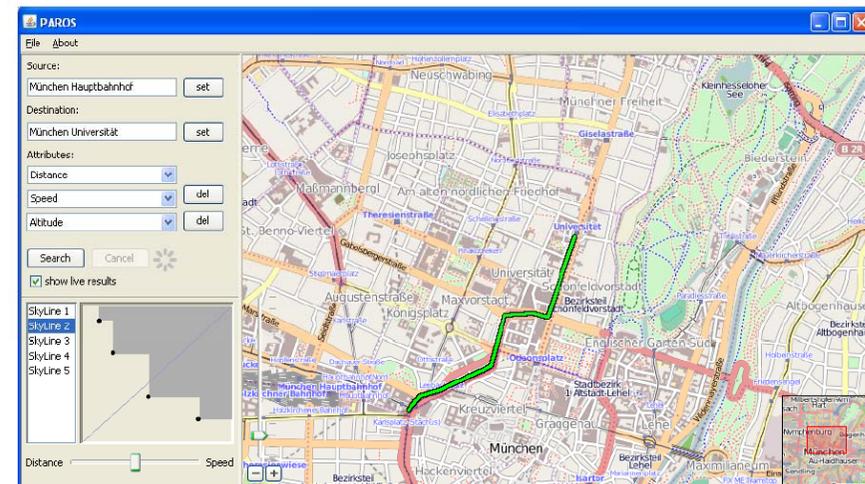
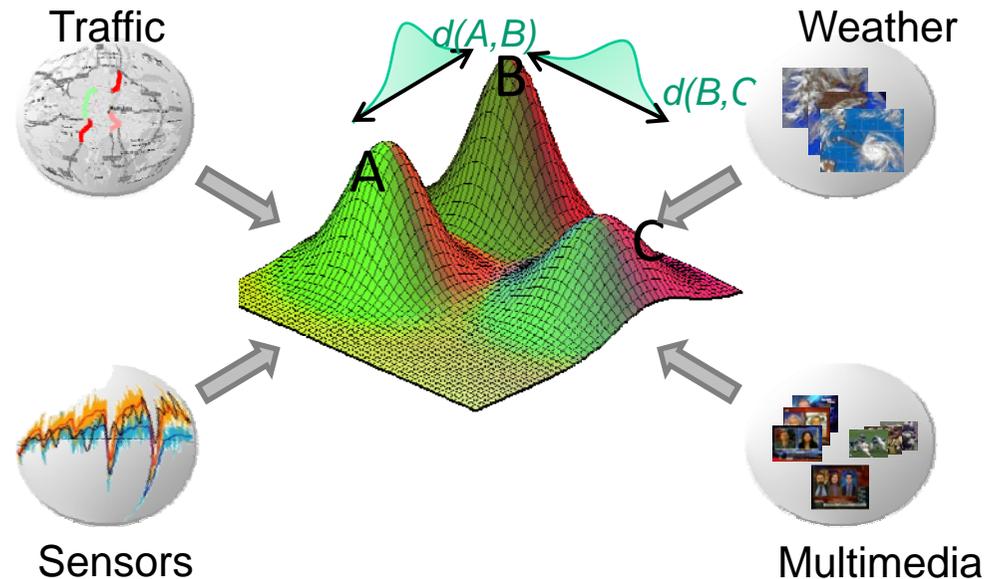


Figure 1: Screen shot of the PAROS interface for 2D Skyline Queries.

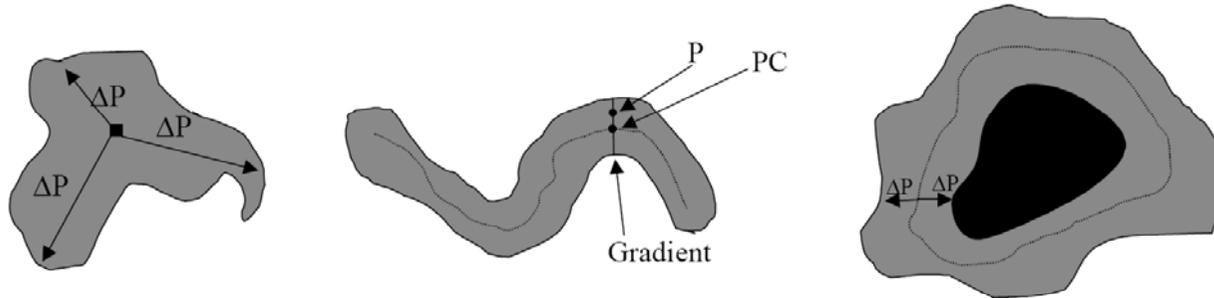
1.4.3 Anwendungen mit unsicheren Objekten

- Wahrscheinlichkeitsbasierte Anfragen auf räumlich unsicheren Objekten



- Unsicherheit der Daten durch
 - Ungenauigkeit der Positionssensoren (GPS, Radar, etc.)
 - Diskrete Abtastung (Snapshots) von sich zeitlich kontinuierlich verändernden Variablen (RFID-tracking Systeme)
 - Vorhersagen (Aktienkurse, Orkane, Positionen von Eisbergen, ..)
 - Maßnahmen zur Bewahrung der Datenschutzrichtlinien (Depersonalisation)

- Beispiele von räumlich unsicheren Objekten



a) unsichere Position b) unsicherer Pfad c) unsichere Region

- Modellierung

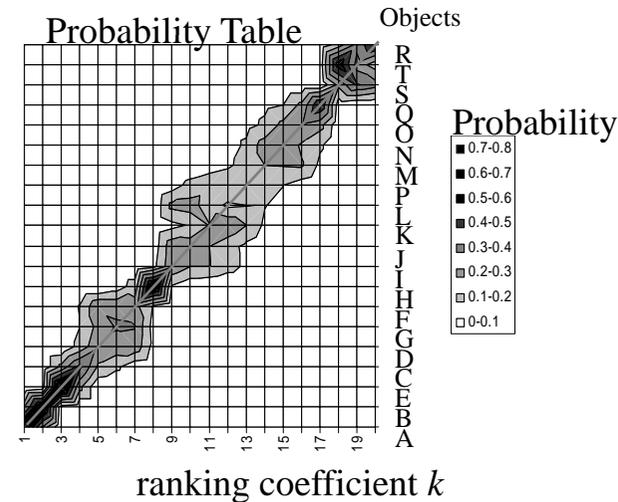
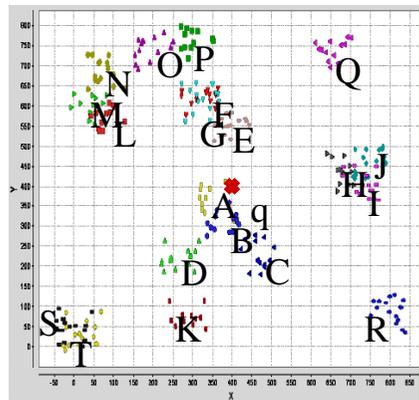
abhängig vom Objekttyp (siehe Skizze oben)

- Punkt (\rightarrow Verteilungsfunktion)
- Trajektorie (\rightarrow Erwartete Trajektorie mit Varianzangaben)
- Region (\rightarrow Wahrscheinlichkeit über die Objektzugehörigkeit von Punkten („fuzzy objects“))

abhängig von der Art der Unsicherheit

- existentielle Unsicherheit (Tupel-Unsicherheit) (\rightarrow Tupel + Wahrscheinlichkeit)
- Attributunsicherheit (Unsicherheit der Ausprägung der Attribute (z.B. Ort))

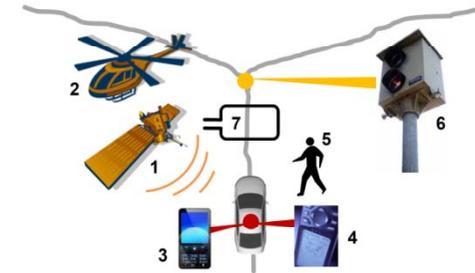
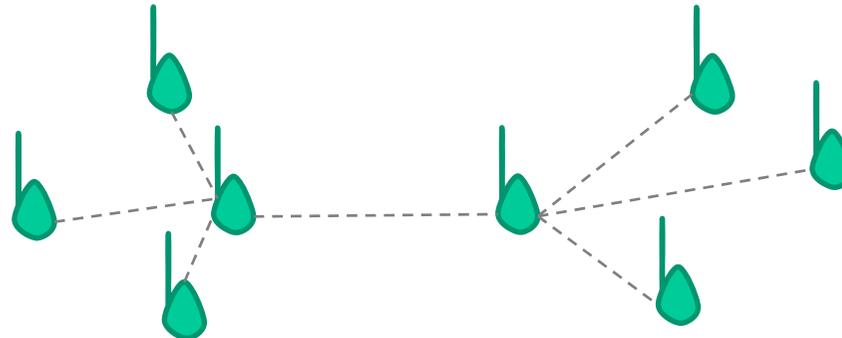
- Anfragen/Recherche
 - Probabilistische Ähnlichkeits-/Nachbarschaftsanfragen
 - Wahrscheinlichkeitsverteilung über Verkehrsdichte
 - Probabilistische Räumliche Anfragen (z.B. Schnittanfragen)
 - Probabilistische Raum-Zeit-Anfragen
 - etc.



a) Unsichere Punktobjekte b) Probabilistisches Distanzranking-Ergebnis

1.4.4 Anwendungen mit Sensornetzwerkdaten

– Suche in Sensornetzwerken



- Sensordaten: Über die Zeit hinweg Beobachtungen von Ereignissen
 - Einfache Attribute: Temperatur, Luftfeuchtigkeit, etc.
 - Positionen von sich bewegenden Objekten: RFID, Mobile Devices, Radar, etc.
- Anfragen/Recherche
 - Kontinuierliche Ausgabe von
 - Top-k-Anfragen,
 - Aggregationen (Mittelwert, Median, Summe)

- Ziele und Herausforderungen
 - Minimierung der Übertragungskosten
 - Anfragebearbeitung möglichst in Echtzeit (Realtime)
 - Limitierte Ressourcen:
 - » Speicherplatz
 - » Übertragungreichweite
 - » CPU Belastung
- Allgemeine Problemklassen (Strategien):
 - Optimierung der Netzwerktopologie (Adhoc-Netzwerke), d.h. Wahl des Weges zur (peer-to-peer) Datenübertragung über das Netzwerk
 - » möglichst kurze Übertragungswege
 - » möglichst wenig Einzel-Übertragungen
 - In-Network Anfragen:
 - » Teile der Anfrage werden bereits in den Netzknoten verarbeitet
 - » Extraktion und Weiterleitung von relevanter Information (Filtering)

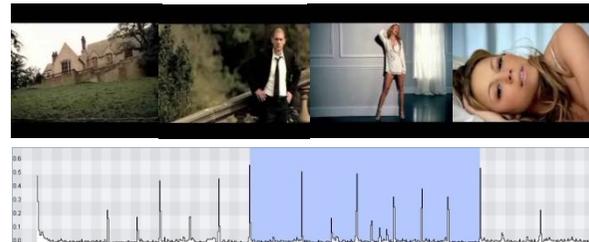
1.4.5 Anwendungen mit Multimediadaten

– Suche in Multimediadaten



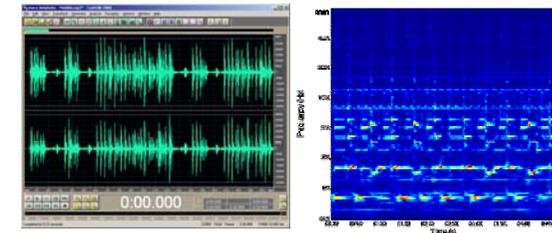
a) Bilddaten

- Ähnlichkeitssuche
- Objekterkennung
- Image Parsing
- Content-Based Image Retrieval
- ...



b) Videodaten

- Shot Detection
- Motion Tracking
- Video Similarity
- Video Indexing
- Objekterkennung
- ...



c) Audiodaten

- Title Identification
- Genre Klassifikation
- Query by Humming
- Speech Recognition
- Metadaten Extraktion
- ...

1.5 Recherche in STMM-DBS (Überblick)

- Unterschiede zu traditionellen DBS
 - In Standard-DBS spezifiziert Benutzer Bedingungen, die Ergebnisse erfüllen müssen (bestimmte Attributswerte); deklarative Anfragen in SQL
 - In STMM-DBS sind Anfragen nach bestimmten Attributswerten eher die Ausnahme
 - Typisch: Recherche auf Basis von Ähnlichkeit/Nachbarschaft
 - Spezifikation einer Anfrage durch
 - Konkretes Anfrageobjekt, das durch den Benutzer zur Verfügung gestellt wird (z.B. durch URL, Datei, ...)
 - Vereinfachte Approximation eines Anfrageobjektes (Skizze, Summen, ...)

Literatur zur Vorlesung

- Die Vorlesung basiert im Wesentlichen auf aktuellen Forschungsergebnissen
- Meist sind die in dieser Vorlesung besprochenen Konzepte bisher nur in den Originalpublikationen besprochen
- Daher orientiert sich diese Vorlesung leider nicht an einschlägigen Lehrbüchern
- Falls weiterführende Literatur zu einzelnen Aspekten der Vorlesung existiert, wird darauf an entsprechender Stelle hingewiesen.