Einsatz von MPEG-7 für die Entwicklung von akustischen Klassifikationssystemen

Bernhard Rettenbacher[1], Werner Bailer[2], Peter Schallauer[2]

Joanneum Research Forschungsgesellschaft mbH, A-8010 Graz, Österreich
[1]Institut für Angewandte Systemtechnik, [2]Institut für Informationssysteme & Informationsmanagement,
Email: bernhard.rettenbacher@joanneum.at

Einleitung

Um akustische Signale automatisch klassifizieren zu können, werden für die Entwicklung eines Klassifikationssystems Trainingsdaten erstellt und diese manuell verschiedenen Klassen zugeordnet. Man nimmt eine inhaltliche Beschreibung der Daten vor. Diese "Daten über Daten" werden als Metadaten bezeichnet. Für die Beschreibung multimedialer Inhalte existiert eine Technologie, die 2002 von der Moving Picture Experts Group unter dem Namen MPEG-7 standardisiert wurde [1]. MPEG-7 ist von der Kodierung der Medieninhalte unabhängig und erlaubt unterschiedliche Abstraktionsstufen der Beschreibung. Das flexible und erweiterbare Konzept von MPEG-7 soll eine Verwendung in unterschiedlichsten Anwendungsgebieten ermöglichen. MPEG-7 Daten werden wahlweise in einem XML-Dokument oder in einer binären Repräsentation gespeichert.

Wir haben den MPEG-7 Standard auf seine Eignung für die Entwicklung von Systemen zur automatischen Identifikation von realen Schallereignissen untersucht, da sich dieses Metadatenmodell/-format gerade im Bezug auf die Interoperabilität zwischen verschiedenen Systemen als sehr vorteilhaft darstellt. Die Merkmalsextraktion, die temporale Segmentierung und die eingesetzten Klassifikatoren dieser Systeme unterscheiden sich in den Verfahren nicht wesentlich von Anwendungen im Bereich Multimedia. Dagegen stellen die detaillierte Beschreibung der Aufnahmesituation, der Eigenschaften der verwendeten Sensoren und deren Aufstellung, sowie verschiedene Umweltparameter eine spezifische Information dar. Es wird vor allem mit Signalen gearbeitet, die mit umwelt- und systembedingte Störungen wie Übersteuerung, Signal-Ausfällen oder sich temporal ändernden Umgebungsbedingungen (z.B. Wettereinflüsse, Lärm,...) behaftet sein können.

Erfassung der Referenzdaten

Durch ein Datenerfassungssystem werden real auftretende akustische Ereignisse separiert und einem Klassifikationssystem zugeführt. Dieses soll die gewünschten Ereignisse aus dem Datenstrom extrahieren und einer oder mehreren festgelegten Klassen zuordnen. Bei der Entwicklung eines solchen Systems muss diesem zuallererst ein Satz von Trainingsdaten zur Verfügung stehen. Diese Trainingsdaten werden aus aufgezeichneten und manuell annotierten Audiosignalen gewonnen. Orte, Objekte, Personen, Ereignisse, Zustände, die zeitliche Zuordnung und die Relationen der semantischen Entitäten können mit MPEG-7 in unterschiedlichen Abstraktionsgraden unter Verwendung eines kontrollierten Vokabulars beschrieben werden. Dieses Vokabular kann entweder den im Standard definierten oder selbst entworfenen Classification Schemes (CS) entnommen werden.

Signalerfassung

Am Aufzeichnungsort werden verschiedene Sensoren, wie zum Beispiel Mikrofone, aufgestellt. Die Medieninformationen können zentral abgelegt werden und von den kanalspezifischen Beschreibungen referenziert werden. Kanalspezifische Erstellungsinformationen (*CreationInformation DS*) werden in den entsprechenden Beschreibungen direkt angegeben. In der *CreationInformation* können Angaben zu Autor, Ort, und den Erstellungs-Werkzeugen (*CreationTool DS*) gemacht werden. Ein *CreationTool* kann beispielsweise eine Software oder auch ein Sensor sein. Spezifische Eigenschaften, wie die verwendete Richtcharakteristik eines Mikrofons können textuell beschrieben werden; für detaillierte Beschreibungen (zum Beispiel eine Klassifizierung der Sensortypen) ermöglicht es die generelle Erweiterbarkeit von MPEG-7, spezielle Typen abzuleiten.

Bei der Verwendung mehrerer Sensoren ist es in manchen Fällen notwendig, die relativen Positionierungen der Sensoren untereinander und bezogen auf den Ort der Aufzeichnung zu kennen und diese Angaben dem Datenverarbeitungssystem zur Verfügung zu stellen. Aus diesen Informationen lassen sich zum Beispiel Laufzeitunterschiede errechnen. Es würde sich daher anbieten, ein lokales Koordinatensystem zu definieren, das eine detaillierte Beschreibung des Aufzeichnungsortes ermöglicht. In MPEG-7 sind für Ortsangaben nur Adressangaben und geografische Koordinaten möglich. Dazu können relative Positionen durch Entfernungs- und Richtungsangaben beschrieben werden, die ausschließlich die horizontale Verschiebung vom Ort angeben. Die Einführung lokaler Koordinaten kann also nur durch eigene Erweiterungen realisiert werden. Aus jetziger Sicht sollte dies kein Problem darstellen. Ein allgemein gültiger Ansatz, der in den verschiedensten Anwendungsfeldern einsetzbar ist, muss aber erst gefunden werden.

Fehlerbehandlung und Synchronisation

Nach der Aufzeichnung der Signale müssen etwaige Aufzeichnungsfehler behandelt werden und - bei der Verwendung mehrerer unabhängiger Aufzeichnungssysteme - die Signale synchronisiert werden. Die möglichen Fehlerquellen, sind einerseits durch die Umgebung beeinflusst, andererseits systembedingt. In den Audiosignalen entstehen Abschnitte, die nicht oder nur mit Einschränkungen für eine

Weiterverarbeitung geeignet sind. Auch die Synchronisierung erfordert eine Nachbearbeitung durch Verschiebung auf der Zeitachse. Ziel der Nachbearbeitung ist es, möglichst viele der aufgezeichneten Signale bei der Weiterverarbeitung nutzen zu können. So sollen bei einem kurzfristigen Ausfall eines Kanals die nicht betroffenen Kanäle erhalten bleiben. Auch das fertig entwickelte Klassifikationssystem soll fehlertolerant arbeiten können.

Das Entfernen der unbrauchbaren Teile würde die temporale Struktur der Aufzeichnung zerstören oder führt zu einer Segmentierung der Dateien, daher ist eine Beschreibung der fehlerhaften oder qualitativ minderwertigen Segmente und eine anschließende Fehlerbehandlung im Datenverarbeitungssystem einer destruktiven Bearbeitung der Dateien vorzuziehen. Fehler, wie Übersteuerung oder Signalausfälle, können automatisch detektiert werden, andere benötigen eine manuelle Kennzeichnung. Ein Schema zur Beschreibung der Audioqualität (*AudioQuality DS*) und ein Klassifikationsschema für Audio Defekte (*AudioDefects CS*) bilden die Grundlage für die Annotation "problematischer" Daten.

Wurden bei einer Aufzeichnung mehrere Aufzeichnungsgeräte verwendet, so sind diese Aufzeichnungen nur bedingt zueinander synchron. Synchronisationsmechanismen wie SMPTE-Timecode können die angeschlossenen Geräte zwar mit Synchronisationsinformationen versorgen, die aufgezeichneten Daten müssen aber erst aufbereitet und geschnitten werden. Destruktive Bearbeitung kann hier zum Teil erheblichen Rechenaufwand bedeuten, eventuell sogar verbunden mit einer Neucodierung der Daten. Auch hier würde es sich anbieten, diese Arbeitsschritte virtuell durchzuführen.

In MPEG-7 gibt es grundsätzlich Möglichkeiten, den virtuellen Schnitt zu beschreiben, jedoch lassen die verfügbaren Methoden einen Spielraum für Interpretationen zu. Eine Lösung des Problems kann durch Erweiterung des *MediaInstanceType* um Zeitangaben gefunden werden: Der Ausschnitt aus der der Mediendatei und ein Offset, an dem dieser Ausschnitt auf der Timeline positioniert werden soll.

Verarbeitung

Wie bereits oben erwähnt, extrahiert das Klassifikationssystem aus den Audiosignalen verschiedene Merkmale, bestimmt temporale Segmente und führt die gefundenen Merkmalsvektoren einem Klassifikationssystem zu. Für das Training eines Klassifikators ist eine manuelle Segmentierung und Klassifizierung notwendig. MPEG-7 unterstützt diesen Vorgang durch Methoden zur zeitlichen und räumlichen Strukturierung, durch textuelle und semantische Annotations-Elemente und durch Klassenhierarchien. Die Ergebnisse der Klassifikation können selbstverständlich wiederum mit MPEG-7 ausgedrückt werden. Auch Audiovisuelle Merkmale können in der MPEG-7-Beschreibung abgelegt werden. Es kann sich hierbei sowohl um MPEG-7 Audio- oder Video-Deskriptoren als auch um selbst definierte Deskriptoren oder DS handeln.

Ergebnis

Zur Untersuchung des MPEG-7-Standards wurden von uns mehrere Szenarios bereits durchgeführter Projekte und möglicher zukünftiger Messaufgaben untersucht. Die MPEG-7 Metadaten wurden exemplarisch generiert, die notwendigen Erweiterungen zum Teil theoretisch ausgearbeitet, zum Teil konkret implementiert. Wir betrachteten weiters die Voraussetzungen für eine Interoperabilität verschiedener MPEG-7-basierter Anwendungen. Die generischen Strukturierungsmöglichkeiten von MPEG-7 ermöglichen es, idente Inhalte auf unterschiedliche Arten in der Beschreibung zu strukturieren und zu organisieren, was zu Probleme im Bezug auf die Interoperabilität führt. Durch die Einführung von Profilen, die Teilmengen der MPEG-7 Tools enthalten, kann die Komplexität der Beschreibungen verringert und die Nutzung von MPEG-7 konkretisiert werden. Die im Standard definierten Profile decken jedoch unsere Anforderungen nicht vollständig ab. Keines der Profile enthält MPEG-7 Part 3: Video und MPEG-7 Part 4: Audio. Das unter [2] frei verfügbare "Detailed Audiovisual Profile" [3] schließt diese Teile ein und definiert die Verwendung von MPEG-7 eindeutig.

Fazit

MPEG-7 stellt auch für die Entwicklung von Systemen zur automatischen Klassifikation von realen Schallereignissen ein leistungsstarkes Werkzeug dar, das im gesamten Arbeitsprozess, von der Aufzeichnung bis zur Ausgabe der Klassifikationsergebnisse, einsetzbar ist. Für die Anforderungen, die über den klassischen Multimediabereich hinausgehen, ist eine Erweiterung verschiedener Typen erforderlich. Im Allgemeinen kann der Aufwand für die Erweiterung relativ gering gehalten werden, da die neuen Typen durch den Ableitungsmechanismus von XML-Schema [4] erzeugt werden können. Selbst wenn andere MPEG-7-fähigen Systeme diese Erweiterungen nicht kennen, ist es ihnen doch möglich, zumindest die im Standard definierten Inhalte zu verarbeiten - ein gewichtiger Vorteil gegenüber anderen audiovisuellen Metadaten-Standards.

Wir setzen MPEG-7 bereits erfolgreich für Multimedia-Applikationen wie Video-Analyse, -Suche und Medienbeobachtung ein und werden in Zukunft auch im Bereich der Erfassung und Klassifizierung von akustischen Signalen die Metadatenbeschreibung mit MPEG-7 durchführen.

Literatur

[1] ISO/IEC, Multimedia Content Description Interface, ISO/IEC 15938:2001.

[2] Detailed Audiovisual Profile,
URL: `http://mpeg-7.joanneum.at`

[3] W. Bailer and P. Schallauer, "The Detailed Audiovisual Profile: Enabling Interoperability between MPEG-7 based Systems", Proceedings of International MultiMedia Modelling Conference (MMM 2006), IEEE Press, Beijing, CN, Jan. 2006.

[4] W3C XML Working Group, XML Schema Definition Language
URL: `http://www.w3.org/XML/Schema`