EDL-Editor: Eine Anwendung zur automatischen Aufbereitung von Vorlesungsvideos

Stephan Kopf, Fleming Lampi, Thomas King, Malte Probst, Wolfgang Effelsberg

Lehrstuhl für Praktische Informatik IV
Universität Mannheim
68159 Mannheim
{kopf, lampi, king, effelsberg}@informatik.uni-mannheim.de
malte.probst@googlemail.com

Abstract: In immer mehr Lehrveranstaltungen werden Vorlesungsmitschnitte den Studierenden als ergänzendes Lehrmaterial zur Verfügung gestellt. Ein wesentlicher Nachteil bei der Erzeugung von Vorlesungsvideos ist der hohe personelle Aufwand, den das Überarbeiten und Schneiden des Rohmaterials verursacht. Dabei sollte das Schneiden der Videos vorlesungsübergreifend erfolgen, da ein Kapitel häufig am folgenden Vorlesungstermin wieder aufgegriffen wird. In diesem Artikel wird die neue Anwendung EDL-Editor (Edit Decision List) vorgestellt, die es ermöglicht, den manuellen Aufwand bei der Erstellung von Vorlesungsvideos zu minimieren. Im Regelfall beschränkt sich die Tätigkeit eines Benutzers auf die Kontrolle der automatisch ermittelten Schnittpositionen in den Videos. Falls der Algorithmus Schnitte an ungeeigneten Stellen vorschlägt und Korrekturbedarf besteht, wird durch die Anwendung gleichzeitig ein effizientes Editieren der Schnittlisten ermöglicht.

1 Einleitung

In immer mehr Lehrveranstaltungen an Universitäten werden nicht nur Vorlesungsfolien und Übungsmaterialien den Studierenden zur Verfügung gestellt, sondern auch die Möglichkeit geboten, auf Vorlesungsvideos zuzugreifen. Studierende nutzen zunehmend die digitalen Aufzeichnungen, um sich Inhalte auch außerhalb der Vorlesungszeiten anzueignen. Zusätzlich begrüßen viele Studierende die Möglichkeit, speziell zur Vorbereitung auf Klausuren, einzelne Themen nochmals intensiv mit Hilfe von Vorlesungsmitschnitten verinnerlichen zu können.

Das regelmäßige Aufzeichnen von Vorlesungen führt während eines Semesters jedoch zu einem erheblichen personellen Aufwand. Sofern die Vorlesungsfolien als Bildinhalte verwendet werden, welche mit dem Ton des Dozenten unterlegt sind, erfolgt die Digitalisierung – abgesehen vom Start und Stopp der Aufzeichnung – automatisch.

Um sowohl thematisch abgeschlossene als auch kurze Lerneinheiten zu erhalten, sollte das Rohmaterial des aufgezeichneten Videos nicht ohne ein Editieren veröffentlicht

werden. Daher wird die Vorlesung in einem zweiten Schritt aufbereitet und geschnitten. Bei der Aufbereitung ist insbesondere eine Einteilung in Kapitel erforderlich, damit die Studierenden schneller auf gewünschte Vorlesungen zugreifen können. Da eine Vorlesung im Allgemeinen nicht eine einzelne thematische Einheit behandelt, muss das Rohmaterial korrekt geschnitten und anschließend passend – d. h. ggf. vorlesungsübergreifend – zusammengefügt werden, so dass der Vorlesungsstoff innerhalb eines aufbereiteten Videos eine semantisch zusammenhängende Einheit bildet. Langfristig betrachtet erzeugt dieser Arbeitsschnitt bisher einen hohen personellen Aufwand. Die weiteren Schritte, wie beispielsweise die Kodierung in unterschiedliche Videoformate oder die Veröffentlichung der Vorlesungsvideos im Web lassen sich vollständig automatisiert realisieren [LKE06].

In diesem Artikel stellen wir ein neu entwickeltes Verfahren zum automatischen Schneiden von Vorlesungsvideos vor. Obwohl bestehende Systeme einzelne Schritte bei der Veröffentlichung von Vorlesungsvideos automatisieren [Ha05], existiert kein System, das für unsere Zwecke ohne größere Anpassungen geeignet ist und insbesondere das automatische Erstellen von Schnittlisten und das Schneiden der Vorlesungsvideos übernimmt. Das Authoring-on-the-Fly-System (AOF) ist ein komplexes System zur Aufzeichnung und Übertragung von Lehrveranstaltungen sowie zur Erzeugung multimedialer Lerneinheiten [OL02]. Um eine Synchronisation der multimedialen Dokumente zu erreichen, wurde ein eigenes Format zur Speicherung entwickelt. Ein ähnlicher Ansatz wurde für das Lecturnity-System gewählt [Lec07], welches die Erzeugung von multimedialen Lernanwendungen anhand von PowerPoint-Präsentationen ermöglicht, wobei Animationen oder eingeblendete Videos nicht unterstützt werden.

Bei der von uns entwickelten Anwendung soll die Aufzeichnung auf Notebooks mit beliebiger Präsentationssoftware wie beispielsweise PowerPoint, Acrobat Reader oder Open Office Impress möglich sein. Auf dem Präsentationsrechner ist lediglich eine Capture-Anwendung zur Erzeugung eines Videos aus der Audiospur und dem Bildschirminhalt erforderlich. Animationen, Folienübergänge, Videoeinblendungen und Anmerkungen des Dozenten auf den Folien werden erfasst, wobei auch andere Anwendungen wie beispielsweise JAVA-Applets während einer Vorlesung gestartet werden können.

Da insbesondere bei gering strukturierten Vorlesungen einzelne Fehler bei der Erkennung von Schnittpositionen nicht ausgeschlossen werden können, wird im Folgenden die von uns entwickelte und intuitiv zu bedienende Benutzeroberfläche des Programms EDL-Editor (Edit Decision List) vorgestellt, die ein manuelles Bearbeiten und Korrigieren der automatisch identifizierten Schnittpositionen effizient ermöglicht. Schnittpositionen im Rohmaterial können einfach verschoben, gelöscht oder hinzugefügt werden.

Im nächsten Abschnitt werden zunächst Anforderungen an ein Programm zum automatischen Schneiden von Vorlesungsvideos vorgestellt und die Struktur der entwickelten Anwendung erläutert. Abschnitt 3 geht auf die Funktionalitäten und neuen Algorithmen zur Ermittlung von semantischen Inhalten in Videos ein. Die Anwendung EDL-Editor wird in Abschnitt 4 vorgestellt. Auf Erfahrungen, die wir beim automatischen Schneiden von Vorlesungsvideos gewonnen haben, gehen wir in Abschnitt 5 ein. Abschließend werden die gewonnenen Ergebnisse zusammengefasst und ein Ausblick gegeben.

2 Aufbau des Systems EDL-Editor

Schon seit mehreren Jahren werden Vorlesungen des Hauptstudiums an unserem Lehrstuhl aufgezeichnet und den Studierenden als Video zur Verfügung gestellt. Obwohl der Ressourcenbedarf recht hoch ist, bieten Videos deutliche Vorteile gegenüber einer Speicherung der Vorlesung in Form von Einzelbildern. Einerseits werden schriftliche Anmerkungen des Dozenten / der Dozentin auf den Folien im Zeitablauf erfasst, andererseits sind Sprache und Vorlesungsfolien synchron. Ein weiterer ganz wesentlicher Vorteil besteht darin, dass keine spezielle Anwendung zur Wiedergabe der Vorlesungen erforderlich ist, da jeder PC und die meisten mobilen Geräte die Wiedergabe von Videos unterstützen. Da innerhalb der Universität und auch bei vielen Studierenden zu Hause breitbandige Internetverbindungen zur Verfügung stehen, führt der erhöhte Speicherbedarf zu keiner wesentlichen Einschränkung bei der Nutzung der Vorlesungsvideos. Zudem stehen zusätzlich Vorlesungsvideos für eine sehr geringe Bandbreite zur Verfügung, für die nur ISDN-Verbindungen erforderlich sind.

Um eine Vorlesung aufzuzeichnen, muss der Dozent zu Beginn die Aufzeichnung starten und diese am Ende stoppen. Ab dem Startzeitpunkt werden der Bildschirm des Dozenten und der Ton, der über die Lautsprecher der Vorlesungssaals übertragen wird, als Video komprimiert und gespeichert. In früheren Vorlesungen wurde das Video anschließend manuell geschnitten, um beispielsweise den Vor- oder Nachlauf, der keine Vorlesungsinhalte enthält, zu entfernen. Zudem sollen die Videos kapitelweise im Web veröffentlicht werden, so dass beim Start eines neuen Kapitels innerhalb einer Vorlesung ein Schneiden des Rohmaterials sowie ein Zusammenfügen zweier aufeinander folgender Vorlesungen erforderlich sein kann. Im letzten Schritt, der ebenfalls vollständig automatisch abläuft [LKE06], werden die geschnittenen Videos mit unterschiedlichen Profilen kodiert und im Web den Studierenden zugänglich gemacht. Die Profile unterscheiden sich im Wesentlichen in ihren Bitraten, den Bildauflösungen und den verwendeten Videocodecs, um die Anforderungen unterschiedlicher Endgeräte zu erfüllen.

Das manuelle Editieren der aufgezeichneten Videos ist mit einem hohen Zeitaufwand verbunden. Im Folgenden wird unser neues System vorgestellt, das alle Bearbeitungsschritte, die für eine Veröffentlichung von Videos erforderlich sind, automatisch und ohne Benutzerinteraktion durchführen kann. Da Fehler bei der rechnergestützten Aufbereitung von Vorlesungsvideos nie vollständig ausgeschlossen werden können, wird zusätzlich über eine intuitiv zu bedienende Benutzeroberfläche die Möglichkeit gegeben, Korrekturen an den festgelegten Schnittpositionen vorzunehmen.

Die Anwendung EDL-Editor stellt zwei grundlegende Funktionalitäten zur Verfügung. Um geeignete Schnittpositionen automatisch in einem Video zu identifizieren, erfolgt in einem ersten Schritt die Analyse des Rohmaterials. Weiterhin wird die Arbeit eines Anwenders durch eine grafische Benutzeroberfläche unterstützt, welche die bereitgestellten Funktionen optisch ansprechend dargestellt und eine schnelle Interaktion ermöglicht.

Mehrere zentrale Arbeitsschritte sind bei der automatischen Analyse von Vorlesungsvideos erforderlich. Diese bauen aufeinander auf und können nur in der angegebenen Reihenfolge durchgeführt werden:

- Eine Erkennung von Folienübergängen wird durch die Suche von harten Schnitten in den Vorlesungsvideos realisiert.
- Die Erkennung eines neuen Kapitels erfolgt durch Analyse der Kapitelnummerierung mittels Texterkennung in den Videos.
- Unterbrechungen der regulären Vorlesung wie beispielsweise eine Fragerunde oder das Abspielen externer Dokumente (Audio, Video, Animationen) werden durch eine Änderung des Layouts identifiziert.
- Anhand der Folienänderungen, Kapitelübergängen und der Zuordnung von Unterbrechungen werden Schnittpositionen in den Rohvideos festgelegt. Dabei müssen auch Vorlesungen an aufeinander folgenden Vorlesungsterminen kombiniert werden, sofern diese dasselbe Thema behandeln.
- Der letzte Schritt bei der automatischen Aufbereitung von Vorlesungsvideos umfasst das Schneiden, Zusammenfügen und Exportieren der Videos in vordefinierte Formate. Nach dem Upload der aufbereiteten Videos stehen diese den Studierenden im Web zur Verfügung.

3 Ermittlung semantischer Inhalte in Vorlesungsvideos

Damit ein automatischer Schnitt von Vorlesungsvideos möglich ist, müssen wichtige semantische Informationen innerhalb von Vorlesungsvideos automatisch, d. h. ohne Benutzerinteraktionen, identifiziert werden können. Im Folgenden wird auf die vier zentralen Schritte, die beim automatischen Schneiden von Vorlesungsvideos erforderlich sind, näher eingegangen.

3.1 Erkennung von Schnitten in Vorlesungsvideos

Ein Vorlesungsvideo wird vom Rechner als ein sequentielles, unstrukturiertes Medium interpretiert. Um weitere Analyseschritte innerhalb eines Videos zu ermöglichen, ist zunächst eine Segmentierung des Mediums erforderlich. Bei einer Kameraeinstellung handelt es sich um eine kontinuierliche Aufnahme; die direkten Übergänge zwischen Kameraeinstellungen werden als harte Schnitte bezeichnet. Bei Vorlesungsvideos, in denen die präsentierten Folien mit der Audiospur des Dozenten unterlegt sind, wird eine Kameraeinstellung durch die Dauer der Einblendung einer Folie charakterisiert.

Das menschliche Gehirn kann Übergänge zwischen Kameraeinstellungen ohne große Mühe direkt erkennen. Eine manuelle Segmentierung von Videos ist jedoch mit einem hohen zeitlichen Aufwand verbunden und für ein effizientes Aufbereiten von Vorlesungsvideos ungeeignet.

Eine Vielzahl von Algorithmen zur automatischen Erkennung von Schnitten wurden die letzten Jahre entwickelt [KC00, Ne05]. Die zentrale Idee der automatischen Schnitterkennungsverfahren besteht darin, Unterschiede zwischen aufeinander folgenden Bildern

eines Videos zu bewerten. Dabei liegt die zentrale Annahme zugrunde, dass Unterschiede innerhalb einer Kameraeinstellung relativ gering sind und ein Schnitt zu einer starken Bildänderung führt.

Bildänderungen innerhalb eines Vorlesungsvideos sind auf Folienübergänge, auf eine Unterbrechung der Präsentation oder auf schriftliche Anmerkungen des Dozenten auf den Folien zurückzuführen. Ziel der Schnitterkennung soll es im Folgenden sein, einen Wechsel zu einer anderen Folie oder eine Unterbrechung der Präsentation zu identifizieren. Schriftliche Anmerkungen des Dozenten auf den Folien sollen jedoch nicht als Schnitt identifiziert werden. Um eine mögliche Schnittposition zu erkennen, werden jeweils zwei aufeinander folgende Einzelbilder im Video miteinander verglichen. Übersteigt der Unterschied einen vordefinierten Schwellwert, so wird ein Schnitt zwischen den beiden Bildern angenommen.

Bei der Analyse von Vorlesungsvideos kombinieren wir zwei Verfahren, um die Zuverlässigkeit der Schnitterkennung zu erhöhen. Die Summe der absoluten Differenzen der Pixel zweier Bilder liefert zunächst Kandidaten für mögliche Schnitte. In einem zweiten Schritt werden jeweils zwei Bilder in gleichgroße Regionen unterteilt und Histogrammdifferenzen für die entsprechenden Regionen berechnet. Die Region mit der größten Histogrammdifferenz bleibt dabei unberücksichtigt, da angenommen wird, dass schriftliche Anmerkungen des Dozenten in dieser Bildregion durchgeführt wurden. Durch einen Vergleich der übrigen Histogrammdifferenzen mit einem Schwellwert können die korrekten Schnittpositionen in Vorlesungsvideos äußerst zuverlässig identifiziert werden.

Die erkannten Schnitte werden für jedes analysierte Vorlesungsvideo als Metadaten gespeichert. Folgende Analyseschritte können so effizient auf die Ergebnisse der Schnitterkennung zurückgreifen, so dass auch bei einer Anpassung von Parametern eine erneute Analyse der Schnitte nicht mehr erforderlich ist.

3.2 Einsatz der Texterkennung zur Identifikation von Kapitelübergängen

Nach der Erkennung von Folienübergängen werden in einem zweiten Schritt Textinformationen analysiert und ausgewertet. Dabei wird die Annahme getroffen, dass Kapitelnummern und Foliennummern an fest definierten Bildpositionen innerhalb der Vorlesungsfolien sichtbar sind. Bei den analysierten Vorlesungsvideos sind insbesondere zwei Bildregionen relevant. Die Titelzeile im oberen Bereich des Bildes enthält häufig Kapitelnummern. Weiterhin liefert die Foliennummer, die häufig im unteren Bildbereich eingeblendet ist, die Information, ob ein Folienwechsel in Vorwärtsrichtung oder ob ein Rücksprung auf die vorherige Folie durchgeführt wurde. Die Bildpositionen der Titelzeile und der Foliennummer sind innerhalb der Anwendung frei konfigurierbar und müssen einmalig pro Semester für jede Vorlesung festgelegt werden.

Falls keine Textinformationen in den spezifizierten Bildbereichen erkannt werden, so deutet dies auf den Vor- oder Nachlauf des Rohvideos bzw. auf eine Unterbrechung der Präsentation hin. Speziell in Vorlesungsvideos bleiben Texte über einen längeren Zeitraum sichtbar, so dass es ausreicht, Texterkennungsalgorithmen auf nur einem Bild einer Kameraeinstellung anzuwenden.

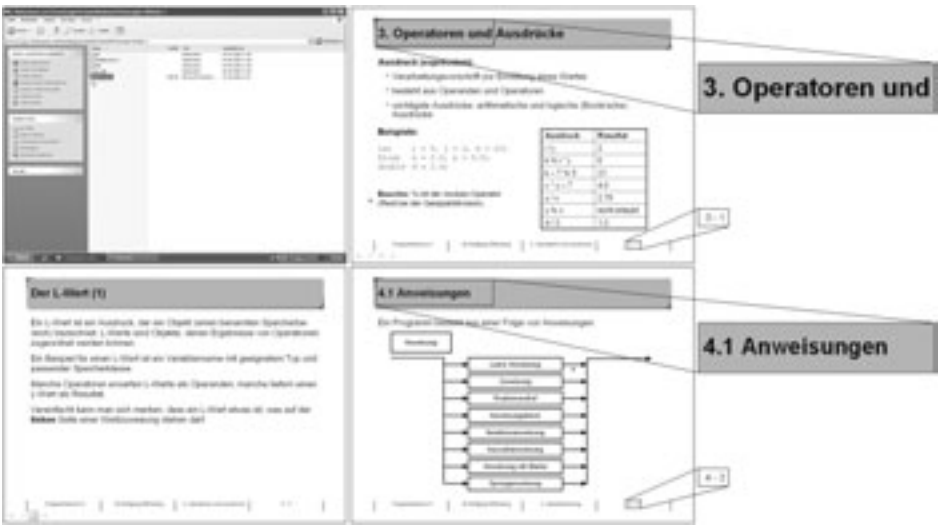

Abbildung 1: Beispiel für vier Kameraeinstellungen innerhalb eines Vorlesungsvideos. Die analysierten Textregionen werden rechts dargestellt.

Abbildung 1 zeigt beispielhaft vier Bilder unterschiedlicher Kameraeinstellungen. Die erste Kameraeinstellung wurde vor Beginn der eigentlichen Vorlesung aufgezeichnet und soll nicht Bestandteil des aufbereiteten Videos sein. In den anderen Bildern werden sowohl Kapitelüberschriften als auch Foliennummerierungen erkannt. Beispielhaft werden im rechten Bereich die Bildregionen, die bei der Texterkennung analysiert werden, verdeutlicht.

Vor der eigentlichen Texterkennung ist eine Segmentierung der einzelnen Buchstaben erforderlich. Speziell bei Vorlesungsvideos liefert die Segmentierung recht zuverlässige Ergebnisse, da ein hoher Kontrast zwischen den Buchstaben und dem Hintergrund besteht. Um einzelne Segmentierungsfehler zu vermeiden, besteht innerhalb der Anwendung die Möglichkeit, die Text- und Hintergrundfarbe manuell zu spezifizieren. Die Festlegung der Farben ist für jede Vorlesungsreihe nur einmal erforderlich.

Wir haben ein neues Segmentierungsverfahren entwickelt, das insbesondere bei geringen Abständen zwischen einzelnen Buchstaben zu sehr zuverlässigen Ergebnissen führt [KHE05]. Dabei werden vor der eigentlichen Segmentierung Trenner zwischen den einzelnen Buchstaben identifiziert, um zu vermeiden, dass zwei Buchstaben kombiniert werden bzw. dass ein Buchstabe unterteilt wird. Zur Bestimmung der Trenner wird innerhalb der Textzeile ein abwärts gerichteter Pfad zwischen zwei Buchstaben gesucht. Von jedem Pixel in der obersten Pixelzeile wird ein Pfad zur untersten Pixelzeile mit den jeweils geringsten Kosten berechnet. Die Kosten des Pfades sind definiert als summierte Pixeldifferenzen zwischen benachbarten Pfadpixeln. Der Pfad mit den geringsten Kosten schneidet nur selten Buchstabenpixel und eignet sich somit gut als Trenner von Buchstaben. Dabei wird der Kürzeste-Pfade-Algorithmus für Graphen von Dijkstra verwendet, um die Trenner zu bestimmen. Jedes Pixel entspricht einem Knoten, der mit drei Nachbarpixeln (links-unten, rechts-unten und unten) verbunden ist. Die Kosten, um von ei-

nem Knoten zum nächsten zu gelangen, sind definiert als absolute Helligkeitsdifferenz dieser beiden Pixel.

Die eigentliche Texterkennung erfolgt durch ein Pattern-Matching-Verfahren [GS90, TJT96]. Dazu werden die einzelnen segmentierten Buchstaben mit bekannten Buchstaben verglichen und das Zeichen mit der größten Übereinstimmung identifiziert. Als Ergebnis der Texterkennung werden ASCII-Zeichen als Metadaten gespeichert und stehen für die weiteren Analyseschritten zur Verfügung.

3.3 Erkennung von Sequenzen

Zur Erkennung von Sequenzen werden redundante Informationen aus den Metadaten entfernt. So ist es beispielsweise für die weitere Bearbeitung eines Vorlesungsvideos nicht erforderlich zu wissen, wie viele Inhaltsfolien auf einen Kapitelanfang folgen. Obwohl die Informationen über Folienanfänge zum Schneiden des Videos nicht benötigt werden, sind sie jedoch für eine schnelle Navigation innerhalb des Videos erforderlich. Zunächst werden iterativ aus der Liste mit allen Kameraeinstellungen gleichartige Einträge entfernt. Innerhalb der analysierten Videos wurden drei Arten von Einträgen definiert: der Anfang eines Kapitels, eine Inhaltsfolie, die jedoch kein neues Kapitel einleitet, sowie unbekannter Inhalt. Unbekannte Vorlesungsinhalte sind beispielsweise eingeblendete Filme oder Animationen. Weiterhin werden regelmäßig interaktive Dienste zur Steigerung der Kommunikation mit den Studierenden während den Vorlesungen eingesetzt [Ko05].

Das Entfernen der doppelten Einträge liefert eine Sequenzliste, anhand derer die endgültigen Schnittpositionen festgelegt werden. Jeder Eintrag wird anhand seiner Vorgängers und Nachfolgers entweder als neues Teilstück identifiziert oder an das vorangegangene Teilstück angehängt. Die Entscheidung erfolgt durch den folgenden regelbasierten Ansatz:

- Falls ein neues Kapitel anhand einer höheren Kapitelnummer im Folientitel gefunden wird, so wird ein neuer Abschnitt festgelegt.
- Unbekannte Inhalte innerhalb des Vorlesungsvideos werden dem davor liegenden Abschnitt zugeordnet.
- Unbekannte Inhalte am Anfang oder Ende eines Videos werden verworfen.

Da eine Vorlesung nicht immer einem starren Schema folgt, war es notwendig eine Mehrzahl an Sonderfällen zu berücksichtigen:

- In der Praxis tritt es wiederholt auf, dass der Dozent / die Dozentin am Ende eines Kapitels auf die nächste Folie wechselt, obwohl das Thema noch nicht vollständig abgeschlossen ist. Häufig erfolgt dann ein Rücksprung auf das vorherige Themengebiet innerhalb weniger Sekunden. Ein kurzes Verweilen auf einer neuen Folie wird nicht als Kapitelanfang erfasst.

- Vor dem eigentlichen Beginn einer Vorlesung wurde wiederholt beobachtet, dass der Dozent / die Dozentin den Foliensatz öffnet und im Schnelldurchlauf bis zur eigentlichen Startfolie wechselt. Falls sehr schnelle Folienwechsel zu Beginn einer Vorlesungsaufzeichnung identifiziert werden, so werden diese entfernt.

3.4 Schneiden der Videos

Der automatische Schnitt der Vorlesungsvideos erfolgt mit Hilfe des Freeware-Programms VirtualDub [Le05], das ein Unterteilen und Zusammenfügen von Videos ohne erneute Kodierung und dem damit verbundenen Qualitätsverlust ermöglicht. Zunächst werden die ursprünglichen Videos an den identifizierten Schnittpositionen in Videosegmente unterteilt. Falls Kapitelinhalte vorlesungsübergreifend behandelt werden, ist ein Zusammenfügen von einzelnen Videosegmenten erforderlich. Dazu werden alle Videosegmente in lexikographischer Reihenfolge bearbeitet und später aufgezeichnete Dateien, die dasselbe oder ein niedrigeres Kapitel im Vergleich zum aktuellen Videosegment besitzen, werden an die aktuelle Datei angehängt.

4 Ablauf der automatischen Bearbeitung von Vorlesungsvideos

Neben den Funktionen zur Analyse von Vorlesungsvideos wurde eine grafische Benutzeroberfläche entwickelt, um die automatisch berechneten Daten effizient verändern zu können. Dies ist erforderlich, da die Analysealgorithmen vereinzelt Schnittpositionen falsch festlegen, die Texterkennung vereinzelt Buchstaben falsch erkennt oder spezielle Abläufe in Vorlesungen auftreten können, die bisher nicht berücksichtigt wurden und manuell korrigiert werden sollten. Weiterhin erleichtert die Benutzeroberfläche die Konfiguration der Anwendung, wie beispielsweise die Definition der Schrift- und Hintergrundfarben, der Position der Textregionen, die Quell- und Zielverzeichnisse, die Parameter für die Schnitterkennungsalgorithmen sowie die Pfade für die externen Hilfsprogramme.

Abbildung 2 verdeutlicht den Aufbau der Anwendung. Es können drei Ansichten – Input, Output und Schnittkontrolle –gewählt werden. Unter Input wird der Fortschritt der einzelnen Analyseschritte verdeutlicht, bei denen eine Liste von Quellvideos in geschnittene Teilvideos überführt wird. Mehrere Rohvideos können gleichzeitig ausgewählt werden und der Fortschritt der einzelnen Algorithmen wird für jede Datei angezeigt (vgl. Abbildung 3, unten). Zusätzlich ist es möglich einzelne, alle oder die noch erforderlichen Analyseschritte manuell zu starten. Falls ein Benutzer einen Analyseschritt direkt startet, kann er zusätzlich festlegen, ob vorherige Schritte erneut berechnet werden sollen. Der aktuelle Fortschritt bei der Analyse der aktuellen Datei und der Fortschritt aller ausgewählter Dateien wird zusätzlich angezeigt.

Unter der Ansicht Output sind die Funktionalitäten zum Zusammenfügen von Videosegmenten und der Speicherung der überarbeiteten Videos im Zielverzeichnis zusammengefasst. Weiterhin besteht die Möglichkeit, die geschnittenen Videosegmente mit Hilfe eines eingebetteten Windows Media Players zu betrachten.

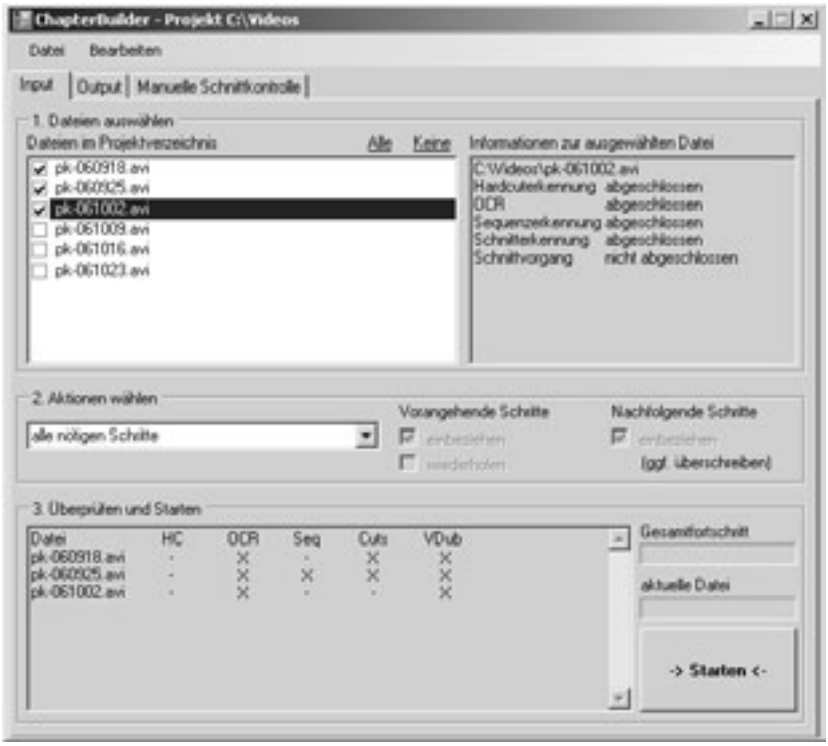

Abbildung 2: Einsatz des EDL-Editors beim automatischen Schneiden von Videos

Eine dritte Ansicht ermöglicht ein nachträgliches Korrigieren der automatisch ermittelten Sequenzlisten. Da nicht sichergestellt werden kann, dass Fehlinterpretationen in einzelnen Videosegmenten auftreten, wurde die Möglichkeit der manuellen Korrektur eingefügt. Jede Schnittinformation wird dabei innerhalb einer Textzeile beschrieben und kann editiert werden. Zudem wurde der Windows Media Player erweitert, so dass eine auf Einzelbildern basierte Navigation im Video möglich ist.

5 Erfahrungen bei der automatischen Aufbereitung von Vorlesungsvideos

Die Anwendung EDL-Editor wurde mit Hilfe eines Trainingsdatensatzes von sechs aufeinander folgenden Vorlesungen im Fach Computer Networks entwickelt. Die Rohvideos wurden zunächst manuell analysiert und die sinnvollen Schnittpositionen per Hand ermittelt. Anschließend wurde die Programmlogik zur automatischen Festlegung der Sequenzliste spezifiziert. In mehreren Iterationen wurden anschließend die Sonderfälle analysiert und die neuen Verfahren zur automatischen Bestimmung korrekter Schnittpositionen festgelegt.

Nach der Fertigstellung der Anwendung erfolgte ein ausführlicher Test der Funktionalität mit einer unbekannten Folge von Vorlesungsvideos aus dem Wintersemester 2006. Im Vergleich zu den Trainingsdaten wurden Vorlesungen des gleichen Dozenten in einem anderen Studienfach (Multimedia Technology) aufbereitet. Die automatische Analyse einer Stunde Vorlesungsvideos benötigt ungefähr 10 Minuten Rechenzeit auf einem aktuellen PC. Besonders erfolgreich ist zu bewerten, dass nur sehr selten ein manueller Eingriff erforderlich war. Auch beim Zusammensetzen der Teilstücke wurden nur selten Fehler beobachtet.

Bei den Analysealgorithmen liegt der Anteil der korrekt erkannten Kameraeinstellungen bei nahezu 100 Prozent. Da die Bildqualität und Bildauflösung bei Vorlesungsvideos deutlich unter der Qualität von eingescannten Textdokumenten liegt, treten bei der Texterkennung wesentlich höhere Fehlerraten auf. Dennoch ist die korrekte Zuordnung von Kapitelanfängen in den meisten Fällen möglich.

Im Folgenden werden noch einzelne Beobachtungen erläutert und Ursachen für mögliche Fehler vorgestellt, die beim Testen der Anwendung beobachtet wurden. Da EDL-Editor eine Korrektur der automatisch ermittelten semantischen Informationen effizient unterstützt, sind die einzelnen Beobachtungen kein echtes Hindernis für den Einsatz von EDL-Editor.

- Bei einer Änderung der Bildauflösung des Rohvideos sollten die Programmparameter wie beispielsweise die Positionen der Textregionen umgehend angepasst werden, da sonst der Einsatz von EDL-Editor zu ungewünschten Ergebnissen führt.

- Innerhalb des EDL-Editors werden Vorlesungsfolien mit ein- oder zweistufiger Nummerierungsstufe unterstützt. Es wurde bewusst darauf verzichtet, Unterkapitel mit mehr als zwei Nummerierungsstufen zu erkennen, da bei drei Nummerierungsstufen zum Teil sehr kurze geschnittene Ergebnisvideos mit einer Länge von nur wenigen Minuten entstehen.

- Bisher wurde die Auswertung der Vorlesungsaufzeichnungen auf die Videospur beschränkt, da sie in fast immer ausreichend genaue Informationen zum Schnitt der Videos liefert. Eine Analyse der Audiospur erfolgt im bisherigen System noch nicht. Speziell am Anfang oder Ende einer Vorlesung könnte die Audiospur wichtige Hinweise zur Schnittposition liefern, indem beispielsweise die Stimme des Dozenten identifiziert wird.

- Auch bei einem Vorwärts- und Rückwärtssprung zwischen zwei Kapiteln lässt sich anhand der Bildinhalte nicht sicher ableiten, ob der Dozent gerade das vergangene Kapitel wiederholt oder schon auf das nächste Kapitel eingeht. Fehler können jedoch durch die manuelle Schnittkontrolle ohne großen Aufwand korrigiert werden.

- Weiterhin wurde die Annahme getroffen, dass Kapitel mit höheren Nummern nach niedrigeren behandelt werden. Falls der Dozent in der Vorlesung eine andere Reihenfolge wählt, würde das Video fehlerhaft geschnitten und kombiniert.

- Unbekannte Inhalte einer Vorlesung wie z. B. Videoeinblendungen, Animationen oder interaktive Dienste werden nicht immer zuverlässig dem korrekten Videosegment zugeordnet. Falls beispielsweise ein Video ein neues Kapitel einführt, würde das Videosegment fälschlicherweise dem vorherigen Segment zugeordnet. Auch im Vor- oder Nachlauf einer Videoaufzeichnung werden unbekannte Inhalte nicht korrekt zugeordnet. In diesen Fällen ist jedoch durch die manuelle Änderung eine schnelle Korrektur möglich.
- Um eine Verschlechterung der Qualität eines Videos durch erneute Kodierung zu vermeiden, ist ein Schnitt eines Videos nur zu Beginn einer Group of Pictures (GOP) innerhalb des Videostroms zulässig. Abhängig von dem verwendeten Video-Codec ist dadurch eine Verschiebung der korrekten Schnittposition um mehrere Einzelbilder bis zu wenigen Sekunden möglich. Diese Fehler werden beim Betrachten jedoch nicht als störend empfunden.

6 Fazit und Ausblick

Obwohl es sich beim Schneiden von Vorlesungsvideos um einen komplexen mehrstufigen Prozess handelt, wird das zentrale Ziel, den manuellen Aufwand zur Aufbereitung und Veröffentlichung von Vorlesungsvideos signifikant zu reduzieren, mit Hilfe des vorgestellten Systems EDL-Editor erreicht. Die von uns entwickelte Anwendung extrahiert in mehreren Schritten Informationen aus den Vorlesungsvideos. Nach der Erkennung von Kameraeinstellungen werden Kapitelgrenzen mit Hilfe von Texterkennungsalgorithmen identifiziert und Regeln abgeleitet, um geeignete Schnittpositionen zu spezifizieren. Falls Fehler bei der automatischen Berechnung auftreten, stellt die Benutzeroberfläche Möglichkeiten zu einer einfachen Korrektur und Anpassung der Schnittpositionen zur Verfügung.

Der manuelle Aufwand bei der Erstellung von Vorlesungsvideos lässt sich durch das vorgestellte System erheblich reduzieren. In den meisten Fällen ist lediglich eine kurze Überprüfung der automatisch geschnittenen Vorlesungen erforderlich. Nur in Ausnahmefällen ist eine manuelle Korrektur erforderlich, die mit Hilfe des Benutzerinterface sehr effizient durchgeführt werden kann. Um eine genauere Klassifikation der Vor- und Nachlaufs der Vorlesung zu ermöglichen, ist eine Weiterentwicklung von EDL-Editor geplant, die eine Auswertung charakteristischer Merkmale der Audiospur ermöglicht.

Mit den heute existierenden Verfahren zur Analyse von Videos ist ein vollständiges Verständnis eines Vorlesungsvideos nicht möglich. Auch für einen Menschen existieren häufig mehrere vergleichbare Möglichkeiten zum Schneiden von Vorlesungsvideos, so dass es die „perfekte“ Lösung nicht gibt. Wenn ein Benutzer Änderungswünsche an der automatisch getroffenen Auswahl an Schnittpositionen wünscht, kann er diese komfortabel und effizient durchführen.

Literaturverzeichnis

[GS90] Govindan, V. K. und A. P. Shivaprasad: Character recognition - a review. In: Pattern Recognition, Bd. 23 (7), S. 671–683, July 1990.

[Ha05] Hartle, M., H. Bär, Ch. Trompler und R. Rößling: Perspectives for Lecture Videos. 11th International Euro-Par Conference, 901-908, Lisbon, Portugal, 2005.

[KC00] Koprinska, I. und S. Carrato: Temporal video segmentation: A survey, Signal Processing: Image Communication, Vol. 16, Issue 5, Pages 477-500, January 2001.

[KHE05] Kopf, S., T. Haenselmann und W. Effelsberg: Enhancing Curvature Scale Space Features for Robust Shape Classification, Proc. of IEEE International Conference on Multimedia and Expo (ICME), Amsterdam, The Netherlands, July 2005.

[Ko05] Kopf, S., N. Scheele, L. Winschel und W. Effelsberg: Improving Activity and Motivation of Students with Innovative Teaching and Learning Technologies, Methods and Technologies for Learning, Palermo, Italy, April 2005.

[LKE06] Lampi, F., S. Kopf und W. Effelsberg: Mediale Aufbereitung von Lehrveranstaltungen und ihre automatische Veröffentlichung - Ein Erfahrungsbericht, Proc. of DeLFI 2006 – Die 4. e-Learning Fachtagung Informatik der Gesellschaft für Informatik, 11-14, Darmstadt, Germany, September 2006.

[Lec07] LECTURNITY – das führende Rapid Authoring Tool, http://www.im-c.de/Produkte/Rapid-Authoring-Tool/, last checked: 06/2007.

[Le05] Lee, A.: VirtualDub scripting language reference, v0.7, http://www.virtualdub.org/docs/vdscript.txt, 2005, last checked: 03/2007.

[Ne05] Nesvadba, J. F. Ernst, J. Perhavc, J. Benois-Pineau und L. Primaux: Comparison of shot boundary detectors, IEEE International Conference on Multimedia and Expo (ICME), July 2005.

[OL02] Ottmann, T. und T. Lauer: Means and Methods in Automatic Courseware Production: Experience and Technical Challenges. In Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 553-560, Chesapeake, VA, 2002.

[TJT96] Trier, Ø., A. Jain und T. Taxt: Feature extraction methods for character recognition – a survey. In: Pattern Recognition, Bd. 29 (4), S. 641–662, 1996.