

MPEG-7: Beschreibung von Bewegungen

Pantelis Katsukis
pantelis.katsukis@ifi.lmu.de

Universität München
Amalienstrasse 17, 80333 Munich, Germany

Abstract. Diese Arbeit beschreibt die Techniken zur Bewegungsbeschreibung in Zusammenhang mit der MPEG-7 Standardisierung. Es werden die verschiedenen Bewegungsdeskriptoren Motion Activity, Camera Motion, Motion Trajectory und Parametric Motion und deren Organisation im Standard vorgestellt. Zu jedem Deskriptor erfolgt eine detaillierte Beschreibung seiner Syntax und seiner Erfassung. Ferner werden verschiedene Anwendungsszenarien näher gebracht, die mit Hilfe der Bewegungsdeskriptoren realisiert werden können. Dabei wird auf die Videozusammenfassung mit Hilfe des Motion Activity Deskriptors genauer eingegangen.

1 Einleitung

Die Bewegungsmerkmale einer Videosequenz ermöglichen den einfachsten Zugriff auf ihre zeitliche Dimension und haben daher einen hohen Stellenwert in der Videoindexierung. Um das Bewegungsumfeld einer Videosequenz zu beschreiben benötigt man ein großes Maß an Information. Das Ziel des MPEG-7 Standards ist nun die Erfassung der wesentlichen Bewegungscharakteristika mittels präziser und effektiver Deskriptoren [1]. Diese Deskriptoren müssen die MPEG-7 Anforderungen einhalten und sind somit allgemeingültig und nicht speziell für eine besondere Anwendung ausgelegt. Außerdem sind sie kompakt in ihrer Größe und ihr Auszug aus dem Videoinhalt ist einfach. Für jede Art von Medieneigenschaften und Kontextgebrauch gibt es auch nur einen entsprechenden Bewegungsdeskriptor.

Mit diesen Anforderungen sind vier Bewegungsdeskriptoren in MPEG-7 standardisiert: Motion Activity (Bewegungsaktivität), Camera Motion (Kamerabewegung), Motion Trajectory (Bewegungsablauf) und Parametric Motion (parametrische Bewegung).

Anhand dieser Deskriptoren lassen sich mehrere Anwendungen realisieren [2]:

1. Inhaltsbasierte Suche und Abfrage von Videodatenbanken: Videosequenzen können mit Hilfe der Bewegungsdeskriptoren automatisch indiziert werden um sie dann in einer Datenbank abzulegen.
2. Videobrowsing: Die Bewegungsdeskriptoren können z.B. die aktivsten Stellen eines Sportspiels finden.
3. Überwachung: Erkennung eines bestimmten Ereignisses in einem Video. Damit kann man dann z.B. einen Alarm auslösen.
4. Videozusammenfassung: Die Deskriptoren können dazu benutzt werden um ein Video automatisch zusammenzufassen. Dieses wird in Kapitel 4.1 näher betrachtet. tem Manipulation des Videos: Die Videoausgabe kann durch die Deskriptoren gesteuert werden. So können Frames mit einer niedrigen Bewegungsaktivität übersprungen werden oder Frames mit einer hohen Aktivität langsamer abgespielt werden.
5. Videoverknüpfung: In einer Videosequenz können Verknüpfungen von einer Region mit anderen multimedialen Inhalten versehen werden.

Die Arbeit ist wie folgt aufgebaut. Im nächsten Abschnitt werden die verschiedenen Bewegungsdeskriptoren vorgestellt und deren Organisation im Standard erläutert. In Kapitel 3 werden die vier Bewegungsdeskriptoren im Einzelnen detailliert dargestellt und beschrieben. Im darauffolgenden Kapitel werden mögliche Anwendungsgebiete vorgestellt. Zusätzlich wird noch ein Beispiel aus der Industrie ausführlicher betrachtet. Zum Schluss fasst ein kurzes Fazit die Arbeit nochmal zusammen.

2 Überblick der Bewegungsdeskriptoren

Die ausgewählten Bewegungsdeskriptoren im MPEG-7 Standard erfüllen den Komplexitäts- und Funktionalitätsbereich, der in der Einleitung erwähnt wurde und somit können eine Vielzahl an Applikationen unterstützt werden. Die vier Bewegungsdeskriptoren im Standard sind Bewegungsaktivität (Motion Activity), Kamerabewegung (Camera Motion), Bewegungsablauf (Motion Trajectory) und parametrische Bewegung (Parametric Motion). Die Deskriptoren sind wie im Folgenden beschrieben im Standard organisiert [3].

Camera Motion und Motion Activity beschreiben die Bewegungseigenschaften eines Videoabschnitts, während Motion Trajectory die Bewegung einer Region beschreibt. Parametric Motion kann sowohl für globale Bewegungen eines Mosaiks oder für lokale Bewegungen einer Region eingesetzt werden. Ein Mosaik stellt eine visuelle Zusammenfassung von einer Videosequenz dar, indem es mehrere Bilder zu einem Bild transformiert.

Zusätzlich gibt es Parameter, die das eben erwähnte Mosaik erfasst, nämlich die Warping Parameters.

2.1 Bewegungsbeschreibung für Videosegmente

Die Geschwindigkeit der Bewegungen in einem Videosegment, die oftmals den Grad der Aktivität kennzeichnet, wird von der Motion Activity erfasst. Der Deskriptor enthält Informationen zur globalen Intensität von Bewegungsaktivität anhand einer Skala von 1 bis 5, wobei 1 wenig Bewegung bedeutet und 5 viel. Dieser Deskriptor ist z.B. gut geeignet für das Videobrowsing, da Abschnitte mit viel Bewegungsaktivität auch gewöhnlich mehr Informationen enthalten.

Ein weiterer wichtiger Deskriptor um Bewegung in einem Videosegment zu beschreiben ist Camera Motion. Er beschreibt alle Kamerabewegungstypen, die in einem Segment auftreten. Durch die Kamerabewegung wird der Blickwinkel des Zuschauers nachgeahmt und man kann somit z.B. die Ereignisse in einer Sportart zuordnen (welche Mannschaft einen Angriff startet, wann ein Golfspieler geschlagen hat, oder Ähnliches).

Mit Hilfe von aufeinander folgenden zweidimensionalen parametrischen Transformationen kann globale Bewegung erkannt werden. Diese Transformationen werden in den Warping Parameters erfasst. Damit können verschiedene Frames (Einzelbilder) in Relation zueinander gebracht werden, um den Rundblick des Videosegmentes (Mosaik) zu erstellen. In MPEG-4 ist dieses Verfahren standardisiert worden und man verwendet dafür die Terminologie *sprite*. Die Warping Parameters enthalten die Koeffizienten des parametrischen Bewegungsmodells, die benutzt werden um das Mosaik zu bestimmen.

2.2 Bewegungsbeschreibung für Bewegungsregionen

Die Verschiebung einer Region und zwar die Entwicklung der zentralen Positionierung eines Objektes in Raum und Zeit wird von dem Motion Trajectory Deskriptor erfasst. Hierbei werden Formänderung der Region nicht betrachtet. Er ist ein Deskriptor auf höherer Ebene, da die Position eines Objektes in jedem Moment von einem kennzeichnenden Punkt gegeben ist. Solch eine einfache Implementierung ist relevant, denn Menschen nehmen Objektbewegung sehr leicht wahr. Mit Hilfe der Beschreibung der Bewegungsabläufe von Objekten kann man verschiedene Dinge in einer Videosequenz unterscheiden (z.B. ein Mensch verglichen mit einem Auto) oder verschiedene Handlungstypen erkennen (z.B. ein Ballwechsel beim Tennis am Netz oder an der Grundlinie). Außerdem hat man dadurch die

Möglichkeit die Position, Geschwindigkeit und Beschleunigung eines Objektes zu bestimmen, wodurch man in bestimmten Überwachungsszenarien einen Alarm auslösen kann, wenn ein entsprechendes Ereignis eintritt. Eine detailliertere Beschreibung der Verschiebung einer Region bietet der Parametric Motion Deskriptor. Er benutzt dasselbe Bewegungsmodell und dieselbe Syntax wie die Waring Parameters und es wird durch ihn ermöglicht ähnliche Bewegungen zu erkennen, die durch Rotationen oder Deformationen verändert wurden und die der Motion Trajectory Deskriptor deshalb nicht erfassen konnte.

3 Bewegungskdeskriptoren des MPEG-7 Standards

In diesem Abschnitt werden die vier Bewegungskdeskriptoren des MPEG-7 Standards ausführlich erläutert und beschrieben. Zu jedem Deskriptor werden ihre Syntax, ihre genaue Semantik und die verschiedenen Parameter dargestellt. Außerdem werden die Algorithmen näher ausgeführt, die gebraucht werden um die Deskriptoren zu nutzen.

3.1 Motion Activity

Unter Motion Activity versteht man die Bewegungsaktivität in einem Videosegment. Dieser Deskriptor versucht die menschliche Wahrnehmung von Handlungsintensität oder die Geschwindigkeit eines Videosegments zu fassen. Ein Beispiel für eine besonders handlungsreiche Szene wäre ein Tor in einem Fussballspiel. Im Gegensatz dazu kommen in einem Interview nur wenige Aktionen vor [4]. Man beachte, dass dieses sich unterscheidet von der Kamerabewegung oder der globalen Bewegung, da es die gesamte wahrgenommene Intensität der Bewegungsaktivität einer Szene berücksichtigt [5]. Im Allgemeinen umfasst ein Videoinhalt die Skala von niedriger bis hoher Aktivität. Durch diesen Deskriptor hat man die Möglichkeit die Intensität einer Handlung in einer bestimmten Videosequenz präzise auszudrücken und er deckt die zuvor genannte Skala umfassend ab [3].

Bislang wurden schon eine Vielzahl an Methoden verwendet, um solche Bewegungsaktivität eines Videosegmentes zu beschreiben, wie z.B. mit Hilfe einer Kombination von Bild- und Audioeigenschaften [6], mittels einer Berührungsdistanz zwischen aufeinander folgenden Frames [7] oder durch die Verwendung einer zweidimensionalen Matrix, durch die die Anhäufung von quantisierten Pixeldifferenzen zwischen den Frames einer Videosequenz gemessen werden [8]. Der Motion Activity Deskriptor, der

im MPEG-7 Standard benutzt wird, verwendet die Varianz der Werte der Bewegungsvektoren, die schon fertig berechnet sind durch die MPEG-komprimierten Videostreams [9].

Beschreibung des Motion Activity Deskriptors

Damit verschiedene Anwendungen, wie z.B. die Überwachung oder Videobrowsing, effizient durchgeführt werden können, stehen mehrere Parameter der Bewegungsaktivität zur Verfügung. Im Folgenden werden diese Attribute aufgelistet und deren Funktion erläutert [2].

Intensity of activity: Dies ist ein Integer zwischen 1 und 5, der die Bewegungsintensität in einer Szene beschreibt, wobei 1 eine geringe Intensität meint und 5 eine hohe.

Direction of activity: Mit Hilfe dieses Deskriptors kann auch die Hauptrichtung der Bewegung bestimmt werden. Falls es eine Hauptrichtung gibt, so wird diese in einem drei-Bit Integer festgehalten, wobei der Wert einen der acht äquidistanten Richtungen entspricht.

Spatial distribution of activity: Hier wird festgehalten, ob die Aktivität sich über mehrere Regionen verteilt, oder ob nur eine große Aktivitätsregion auftritt in der Sequenz. Eine Szene mit einem Kopf, bei dem sich nur der Mund bewegt, hätte nur eine große aktive Region, wobei eine Aufnahme einer verkehrsreichen Straße viele kleine Aktivitätsregionen hat. Dieser Parameter gibt die Anzahl und die Größe der aktiven Regionen in einem Frame an und wird mittels drei Integerwerten ausgedrückt.

Temporal distribution of activity: Im Gegensatz zu der spatial distribution of activity wird hier nicht die räumliche Aktivitätsverteilung betrachtet, sondern die zeitliche. Dieses Attribut beschreibt, ob die Aktivität während der gesamten Videosequenz gleich ist oder sich im Laufe der Sequenz verändert. Hierfür werden fünf sechs-Bit Integer verwendet.

Besonderheiten des Motion Activity Deskriptors

Die Intensität der Bewegungsaktivität ist ein subjektives Maß für die Geschwindigkeit in einer Videosequenz. Um nun relativ objektive Werte für die Intensität zu erhalten muss man zuerst eine Evaluation zur Messung von Bewegungsaktivität durchführen [5]. Im Falle von MPEG-7 wird der Motion Activity Deskriptor anhand einer Abschätzungsfunktion der subjektiven Bewegungsaktivität festgelegt. Um nun die Bewegungsaktivität in dem Deskriptor berechnen zu können, muss zuerst eine Grundgenauigkeit erstellt und evaluiert werden. In diesem Fall ist die Grundgenauigkeit die durch einen Menschen wahrgenommene Intensität der Bewegungsaktivität in einem bestimmten Videosegment. Dieses kann man als

Maß zur Einteilung von verschiedenen Stufen von Bewegungsintensität benutzen. Der Motion Activity Descriptor im MPEG-7 Standard wurde mit einem Datensatz von 637 Videosegmenten durchgeführt [3]. Diese Videosegmente bestehen aus Aufnahmen, die einige Sekunden lang sind und alle verschiedenen Genres umfassen. Testpersonen haben dann die Ausschnitte ausgewertet und die Intensitätsrate eines jeden Ausschnitts anhand einer Skala von 1 (sehr wenig Intensität) bis 5 (sehr hohe Intensität) bestimmt. Zum Schluss der Bestimmung der Grundgenauigkeit wurde der Durchschnitt der Bewertungen zu jedem Segment berechnet. Es gibt einige kritische Anmerkungen zu dem Grundgenauigkeitsverfahren des MPEG-7 Standards [5]. Zum einen fehlen statistische Daten über die Versuchspersonen und es gibt Segmente, die in Länge und Qualität der Aufnahmeaufteilung variieren.

Die Bewegungsvektoren, die durch MPEG-komprimierten Videostreams zur Verfügung stehen, bilden den einfachsten Ansatz, um die Bewegungscharakteristika in einer Videosequenz zu beschreiben. Da die Größe der Bewegungsvektoren auch ein Anzeichen für die Größe der Bewegung selbst ist, liegt es nahe statistische Eigenschaften der Bewegungsvektoren der Makroblöcke zu benutzen um die Intensität der Bewegungsaktivität abzuschätzen. Die Standardabweichung der Bewegungsvektorgößen treffen nach angemessener Quantisierung recht genau auf die Grundgenauigkeit zu. Die Grenzwerte der verschiedenen Aktivitätswerte erhält man dadurch, indem man die vorher durchgeführte Grundgenauigkeit mit in Betracht zieht. Des Weiteren kann man diese Grenzwerte bezüglich ihrer Framegröße linear normalisieren. Sie funktionieren für alle Arten von Inhalt und Kodierungsumständen aufgrund der Mannigfaltigkeit des Grundgenauigkeitssatzes.

3.2 Camera Motion

Die Bewegung der Kamera in einer Videosequenz wird vom Camera Motion Descriptor beschrieben. Sie ist von größer Bedeutung, da sie Bezug zu der inhaltlichen Semantik hat. Nimmt man zwei Aufnahmen einer gleichen Szene mit verschiedenen Kamerabewegungen, so führt dies oft zu verschiedenen subjektiven Darstellungen der Szene. Kennt man den zeitlichen Ablauf der Kameraoperationen, kann dadurch zwischen verschiedenen Genres unterschieden werden, wie z.B. Actionfilme, Nachrichtensendungen oder Sportübertragungen.

Beschreibung des Camera Motion Deskriptors

Der Camera Motion Deskriptor unterstützt alle Kameraoperationen, wie

Translationen, Rotationen und Brennweitenänderungen zusammen mit all deren Kombinationen. Es gibt acht Standardkameraoperationen:

- Schwenken (Panning): horizontale Rotation (nach links und rechts)
- Verfolgen (Tracking): querlaufende horizontale Bewegung (nach links und rechts)
- Abkippen (Tilting): vertikale Rotation (nach oben und unten)
- Hoch- bzw. Runterziehen (Booming): querlaufende vertikale Bewegung
- Zoomen (Zooming): Änderung der Brennweite
- Vor- bzw. zurückfahren (Dollying): Translation entlang der Sehachse
- Rollen (Rolling): Rotation entlang der Sehachse (im und gegen den Uhrzeigersinn)
- Keine Operation: fixierte Kamera

Diese Kameraoperationen können für jedes Frame beschrieben werden. Für jedes Framepaar einer Videosequenz lassen sich die Kamerabewegungsparameter qualitativ und quantitativ bestimmen. Solch genaue Informationen können für eine Reihe von ausführlichen Applikationen relevant sein, die ein hohes Maß an Details brauchen. Für die meisten Anwendungen reicht jedoch eine langfristige Beschreibung der Kamerabewegung. Die in einer Videosequenz am häufigsten benutzte Kameraoperation ist meistens schon aussagekräftig genug als die augenblicklichen detaillierten Kameraparameter der Framepaare. Somit ist der Camera Motion Deskriptor so aufgebaut, dass er solche langfristige Darstellungen bewältigen kann, wie auch die niedrigste Ebene der Beschreibung.

Den Kern der Beschreibung macht ein Satz von N Kamerabewegungssegmenten aus [2]. Ist N gleich 1, so erhält man eine Beschreibung auf höchster Ebene. Ist N jedoch gleich der Gesamtanzahl der Frames in der Videosequenz, so bekommt man eine detaillierte Beschreibung, bei der jedes Framepaar betrachtet wird.

Für jeden Kamerabewegungstypen in einer Videosequenz werden der Präsenzanteil (fractional presence) und der Bewegungsanteil (amount of motion) bestimmt. Der Präsenzanteil eines Bewegungstyps in einem Intervall ist definiert als der prozentuale Anteil der Gesamtlänge der Sequenz, die der entsprechende Bewegungstyp hat. Es gibt 15 optionale Präsenzanteilparameter, die für die möglichen Kameraewegungstypen stehen. Von den Bewegungsanteilparametern gibt es 14 optionale, wobei der Wert 0 ist, wenn die Kamera fixiert ist und somit keine Bewegung herrscht.

Zusätzlich zu diesen Parametern gibt es noch Attribute für die Zeiteinteilung und zwar die Startzeit (start time) und Dauer (duration), um ein bestimmtes Segment zeitlich anzusteuern.

Besonderheiten des Camera Motion Deskriptors

Im Camera Motion Deskriptor wird eine Abgleichungsfunktion definiert, damit die Ähnlichkeit zwischen zwei Kamerabewegungsdeskriptoren, die zweien Videosegmenten zugeordnet sind, berechnet werden kann [2]. Mit Hilfe dieser Funktion kann man ein Segment nach einer bestimmten Kameraoperation durchsuchen. Man erhält ein Ergebnis zwischen null und eins, wobei null keine Ähnlichkeit aufweist zwischen den zwei Deskriptoren in den Segmenten und eins eine perfekte Übereinstimmung.

Für die Funktion werden einige Parameter benötigt:

α_I : Gibt an, ob der Präsenzanteil eines bestimmten Kamerabewegungstyps in Betracht gezogen wird ($\alpha_i = 1$, wobei $0 \leq i \leq 15$) oder nicht ($\alpha_i = 0$).

β_I : Gibt an, ob der Bewegungsanteil eines bestimmten Kamerabewegungstypen in Betracht gezogen wird ($\beta_i = 1$, wobei $0 \leq i \leq 14$) oder nicht ($\beta_i = 0$).

Zusätzlich gibt es noch drei Gewichte, um den relativen Einfluss des Präsenzanteils (ω_P), des Bewegungsanteils (ω_B) und der Sequenzdauer (ω_D) festsetzen zu können.

Seien nun $C1$ und $C2$ zwei Camera Motion Deskriptoren, P_i^{C1} bzw. P_i^{C2} die Faktoren für die Präsenzanteile, B_i^{C1} bzw. B_i^{C2} die Faktoren für die Bewegungsanteile eines jeden Kamerabewegungstypen i und schließlich D_i^{C1} bzw. D_i^{C2} die gesamte Länge des zu betrachtenden Videosegmentes. Die Ähnlichkeit zwischen zwei Deskriptoren kann nun wie folgt errechnet werden (Gleichung 1):

$$M(C1, C2) = \frac{\omega_P M_P(C1, C2, \alpha) + \omega_B M_B(C1, C2, \beta) + \omega_D M_D(C1, C2)}{\omega_P + \omega_B + \omega_D} \quad (1)$$

Wobei nun für die einzelnen Größen folgende Gleichungen zur Berechnung der Ähnlichkeit eingesetzt werden (Gleichungen 2, 3 und 4):

$$M_P(C1, C2, \alpha) = \frac{\sum \alpha_i * [1 - |P_i^{C1} - P_i^{C2}| / (P_i^{C1} + P_i^{C2})]}{\sum \alpha_i} \quad (2)$$

$$M_B(C1, C2, \beta) = \frac{\sum \beta_i * [1 - |B_i^{C1} - B_i^{C2}| / (B_i^{C1} + B_i^{C2})]}{\sum \beta_i} \quad (3)$$

$$M_D(C1, C2) = 1 - |D^{C1} - D^{C2}| / (D^{C1} + D^{C2}) \quad (4)$$

3.3 Motion Trajectory

Der Motion Trajectory Deskriptor beschreibt die Bewegungsabläufe von Objekten in einer Videosequenz. Mit diesem Deskriptor kann man die

Verschiebung von Objekten in Raum und Zeit erfassen. Da die Position eines Objektes durch einen stellvertretenden Punkt in jedem Augenblick festgehalten wird, liegt diese Beschreibung auf einer hohen Ebene. Solch eine einfache Implementierung ist von Bedeutung, denn Menschen nehmen Objektbewegung in einem hohen Niveau auf.

Beschreibung des Motion Trajectory Deskriptors

Der Kern der Beschreibung des Motion Trajectory Deskriptors ist ein Satz von Schlüsselbildern (keypoints), die die aufeinanderfolgenden raumzeitlichen Positionen des beschriebenen Objekts darstellen [3]. Sie werden durch ihre Koordinaten in Raum (zwei- oder dreidimensional) und Zeit definiert. In der ersten Stufe der Approximation wird eine lineare Interpolierung verwendet.

Seien f_a und f_b die Positionen an den Zeitpunkten t_a und t_b und v_a die Beschleunigung und Geschwindigkeit eines Objektes zwischen den zwei Zeitpunkten t_a und t_b . Die lineare Interpolierung wird wie folgt berechnet (Gleichungen 5 und 6):

$$f(t) = f_a + v_a(t - t_a) \quad (5)$$

wobei

$$v_a = \frac{f_b - f_a}{t_b - t_a} \quad (6)$$

Zusätzlich kann nun in einem zweiten Schritt ein Interpolierungsparameter a_a hinzugefügt werden, um eine nichtlineare Interpolierung zwischen den Schlüsselbildern durchzuführen. Die neue Formel sieht dann wie folgt aus (Gleichungen 7 und 8):

$$f(t) = f_a + v_a(t - t_a) + \frac{1}{2}a_a(t - t_a)^2 \quad (7)$$

wobei

$$v_a = \frac{f_b - f_a}{t_b - t_a} - \frac{1}{2}a_a(t_b - t_a) \quad (8)$$

Außer diesen eben genannten Ausführungen enthält die Syntax des Motion Trajectory Deskriptors auch noch Informationen über die Maßeinheiten und das Koordinatensystem, die benutzt werden um die Positionen zu beschreiben. Zusätzlich gibt es noch einen booleschen Wert, der wahr ist, wenn die Kamera das zu beschreibende Objekt verfolgt. Das macht Sinn, da das Objekt dann in der Videosequenz still steht und ihr Bewegungsablauf nicht viel Informationen gibt und die Beschreibung deshalb wenig sinnvoll ist.

Besonderheiten des Motion Trajectory Deskriptors

Um ein Objekt in einer Videosequenz beschreiben zu können muss die Position dieses Objektes räumlich und zeitlich bekannt sein. Im MPEG-4 Standard wird die Ortung mit Hilfe von Alpha Kanälen spezifiziert. Und zwar werden zuerst die Zeiten von Schlüsselpunkten festgehalten als aufeinanderfolgende Zeiten der Frames. In diesem Schritt sind das ein Schlüsselpunkt pro Frame. Danach instantiiert man für jede festgehaltene Zeit die Position eines Schlüsselpunktes als die Koordinaten des Zentrums des Objektes und es werden die Interpolierungsparameter berechnet.

Diesen Prozess kann man noch beeinflussen, indem man die Anzahl der Schlüsselpunkte und die Länge der Zeitintervalle ändert. Hat man z.B. in einer Applikation nicht viel Speicher und Prozesszeit zur Verfügung, so kann man eine sequentielle Prozedur einsetzen. Diese funktioniert so, indem man mit der Vereinigung der ersten zwei Intervallen $[t_0, t_1] \cup [t_1, t_2]$ anfängt, die aus den ersten drei Punkten bestehen. Schritt für Schritt wird nun ein neuer Punkt in dem Intervall eingefügt und interpoliert, bis ein festgelegter Grenzwert überschritten wird. Ist der Grenzwert überschritten, so hält man dieses Intervall fest und fängt die Prozedur mit den nächsten drei Punkten wieder an.

Falls nun der Speicherplatz die Haupteinschränkung in einer Applikation ist, so kann man ein rekursives Verfahren anwenden. Dabei hat man am Anfang ein Intervall mit allen Punkten und spaltet dieses dort in zwei Intervalle auf wo der Interpolierungsfehler am größten ist. Dieser Schritt wird rekursiv auf die neuen Intervalle angewendet, bis die angestrebte Beschreibungsgröße erreicht ist oder ein bestimmter Fehlergrenzwert überschritten wird.

3.4 Parametric Motion

Mit Hilfe des Parametric Motion Deskriptors kann die Bewegung und Verformung von Objekten in einer Videosequenz durch zweidimensionale geometrische Transformationen beschrieben werden. Er ist definiert als der Vektor von Parametern, die in Verbindung mit den geometrischen Transformationen stehen. Durch ihn kann man die komplexen Bewegungen von Objekten fassen, wie z.B. eine Rotation.

Beschreibung des Parametric Motion Deskriptors

Das grundlegende Prinzip hinter diesem Deskriptor ist die Beschreibung der Bewegung von Objekten als ein zweidimensionales parametrisches Modell [10]. Im Folgenden stehen die einzelnen Modelle, die im Deskriptor

definiert sind. Seien $v_x(x, y)$ und $v_y(x, y)$ die x und y Verschiebungskomponenten des Pixels mit den Koordinaten (x, y) :

- Translational / Konstant (2 Parameter):

$$\begin{aligned}v_x(x, y) &= a_1 \\v_y(x, y) &= a_2\end{aligned}$$

- Rotation / Skalierung (4 Parameter):

$$\begin{aligned}v_x(x, y) &= a_1 + a_3x + a_4y \\v_y(x, y) &= a_2 - a_4x + a_3y\end{aligned}$$

- Affin (6 Parameter):

$$\begin{aligned}v_x(x, y) &= a_1 + a_3x + a_4y \\v_y(x, y) &= a_2 + a_5x + a_6y\end{aligned}$$

- Planar perspektivisch (8 Parameter):

$$\begin{aligned}v_x(x, y) &= (a_1 + a_3x + a_4y)/(1 + a_7x + a_8y) \\v_y(x, y) &= (a_2 + a_5x + a_6y)/(1 + a_7x + a_8y)\end{aligned}$$

- Quadratisch / Parabelförmig (12 Parameter):

$$\begin{aligned}v_x(x, y) &= a_1 + a_3x + a_4y + a_7yx + a_9x^2 + a_{10}y^2 \\v_y(x, y) &= a_2 + a_5x + a_6y + a_8xy + a_{11}x^2 + a_{12}y^2\end{aligned}$$

Der Deskriptor beschreibt nun, welches Bewegungsmodell von den eben genannten Modellen benutzt wird, welches Zeitintervall und Koordinatensystem verwendet wird und schließlich welchen Wert die Parameter a_i haben.

Im Wesentlichen enthalten affine Modelle Translationen, Rotationen, Skalierungen und Kombinationen von denen. Planare perspektivische Modelle machen es möglich globale Verformungen eines Objektes näher zu beschreiben, während man mit quadratischen Modellen noch komplexere Bewegungen erfassen kann.

Besonderheiten des Parametric Motion Deskriptors

Die Erfassung des parametrischen Bewegungsmodells wird wie ein klassisches Optimierungsproblem gelöst, indem die Variablen die Parameter

des gewählten Bewegungsmodells sind und der Bewegungskompensationsfehler die Funktion ist, die minimiert werden muss [2].

Das Fehlermaß wird als erstes definiert als (Gleichung 9)

$$E = \sum_i [I'(x'_i, y'_i) - I(x_i, y_i)]^2 \quad (9)$$

wobei $I'(x'_i, y'_i)$ das durch Bewegung geänderte Bild ist, welches auf das Referenzbild $I(x_i, y_i)$ angepasst werden muss. Die Summation wird über alle Pixel ausgeführt.

Um den Fehler E nun zu minimieren wird ein iteratives Verfahren verwendet, das bei jeder Iteration eine entsprechende Suchrichtung im Parameterraum berechnet und danach eine Suche in dieser eben berechneten Richtung durchführt.

Sei \bar{d} die Suchrichtung. Sie wird berechnet, indem die ersten und zweiten partiellen Ableitungen der Fehlerfunktion benutzt werden. Sei nun \bar{g} der Gradient der Fehlerfunktion und H die Hesse-Matrix. \bar{g} verwendet die erste partielle Ableitung der Fehlerfunktion (Gleichung 10) und H die zweite partielle Ableitung (Gleichung 11).

$$g_k = \frac{\partial E}{\partial a_k} = \sum_i 2 * [I'(x'_i, y'_i) - I(x_i, y_i)] * \left(\frac{\partial I'(x'_i, y'_i)}{\partial x'} * \frac{\partial x'}{\partial a_k} + \frac{\partial I'(x'_i, y'_i)}{\partial y'} * \frac{\partial y'}{\partial a_k} \right) \quad (10)$$

$$H_{kl} = \frac{\partial^2(E)}{\partial a_k \partial a_l} = \sum_i \left[\frac{\partial I'(x'_i, y'_i)}{\partial a_k} * \frac{\partial I'(x'_i, y'_i)}{\partial a_l} + [I'(x'_i, y'_i) - I(x_i, y_i)] \times \left(\frac{\partial I'(x'_i, y'_i)}{\partial x'} * \frac{\partial^2 x'}{\partial a_k \partial a_l} + \frac{\partial I'(x'_i, y'_i)}{\partial y'} * \frac{\partial^2 y'}{\partial a_k \partial a_l} \right) \right] \quad (11)$$

In diesen Formeln können die Ableitungen $\partial x'/\partial a_k$ und $\partial y'/\partial a_k$ durch die Bewegungsmodelle einfach berechnet werden, während die Ableitungen $\partial I'(x'_i, y'_i)/\partial x'$ und $\partial I'(x'_i, y'_i)/\partial y'$ so ausgerechnet werden, indem die Gradientenoperatoren auf das veränderte Bild angewendet werden.

Insgesamt gesehen ist die Suchfunktion \bar{d} wie folgt aufgebaut (Gleichung 12):

$$\bar{d} = -(\epsilon \mathbf{E} + H)^{-1} * \bar{g} \quad (12)$$

wobei \mathbf{E} die Einheitsmatrix ist und ϵ so klein wie möglich gewählt wird. Mit Hilfe dieser Formel können nun parametrische Bewegungen in einer Videosequenz erfasst werden.

4 Anwendungsbeispiele

Mit Hilfe der vier so eben aufgeführten Bewegungsdeskriptoren des MPEG-7 Standards lassen sich viele Anwendungen realisieren. Wie schon in der Einleitung erwähnt sind die typischen Anwendungsgebiete die inhaltsbasierte Suche und Abfrage von Videodatenbanken, das Videobrowsing, die Überwachung, die Videozusammenfassung, die Manipulation eines Videos und die Videoverknüpfung [2].

Um z.B. eine Übersicht eines Videos automatisch durchführen zu lassen, kann der Camera Motion Deskriptor verwendet werden [11]. Er erkennt das Benutzen einer neuen Kameraeinstellung zwischen zwei Frames, was ein Anzeichen dafür ist, dass eine neue Szene beginnt. Somit kann dann zwischen den zwei Frames geschnitten werden und der Nutzer erhält eine Übersicht der verschiedenen Szenen. Hat man eine solche Einteilung einmal durchgeführt, so kann man mit verschiedenen Methoden Videobrowsing Szenarien realisieren, wie z.B. für die Navigation durch das eingeteilte Video Baumstrukturen verwenden [12].

Insbesondere in Sportvideos kann der Motion Activity Deskriptor benutzt werden um handlungsreiche Szenen rauszusuchen [4]. Er kann auch dazu benutzt werden um ein Video aufzuteilen in Segmente mit viel bzw. wenig Bewegungsaktivität.

Für Überwachungsszenarien ist der Motion Trajectory Deskriptor besonders gut geeignet, denn er erkennt Objekte, die z.B. ein bestimmtes Gebiet passieren oder Objekte, die sich schneller bewegen als eine bestimmte Geschwindigkeit, was z.B. für die Straßenüberwachung in Frage kommen würde. Auch Videoverknüpfung lässt sich mit Hilfe des Motion Trajectory Deskriptors realisieren. Objekte, die mit diesen Deskriptor erfasst werden, können nämlich mit anderen multimedialen Inhalten verknüpft werden. Im Folgenden wird die Videozusammenfassung mit Hilfe des Motion Activity Deskriptors noch genauer erläutert.

4.1 Videozusammenfassung mittels Motion Activity Deskriptors

Eine Arbeit der Mitsubishi Electric Research Laboratories beschäftigt sich damit, dass der Motion Activity Deskriptor dazu benutzt werden kann um Videos zusammenzufassen [4]. Es wird zuerst die Hypothese aufgestellt, dass hohe oder niedrige Bewegungsintensität ein Indikator dafür ist wie sehr eine Szene sich ändert. Dadurch ist die Bewegungsintensität also ein Maß für die Zusammenfassbarkeit einer Videosequenz.

In der Arbeit wurde nun zuerst diese Hypothese nachgewiesen. Dazu

wurde ein Experiment durchgeführt, bei dem ein Nachrichtenprogramm in seine verschiedenen Aufnahmen eingeteilt wird und deren Bewegungsaktivität berechnet wird. Dabei kam heraus, dass die Zusammenfassbarkeit der Aufnahmen schlechter wurde wenn die Bewegungsaktivität zunahm. Daraus kann man schließen, dass sich Aufnahmen mit niedriger Bewegungsaktivität für die Zusammenfassung eignen.

Die nächste Idee war nun bei einem Video die weniger interessanten Teile zu überspringen indem deren Geschwindigkeit beim Abspielen erhöht wurde. Somit wird das Video noch komplett beibehalten und nur die Geschwindigkeit beim Abspielen ändert sich. Bei einem Fussballspiel z.B. überspringt man die Stellen, an denen das Spiel unterbrochen ist.

Dieses Beispiel zeigt, dass sich auch die Industrie mit den neuen Möglichkeiten der visuellen Bewegungsbeschreibung im MPEG-7 Standard befasst und versucht das hohe multimediale Datenvolumen in unserer heutigen Zeit automatisch zusammenzufassen.

5 Zusammenfassung

Wie in dieser Arbeit gezeigt wurde, liefert der MPEG-7 Standard eine Sammlung an Bewegungsdeskriptoren, die viele Bewegungsaspekte in einer Videosequenz erfassen können. Für diese Erfassung steht auch viel Präzisionsspielraum zur Verfügung: So kann man eine Beschreibung auf Kosten von Speicherplatz detailliert durchführen lassen oder knapp und dafür schnell. Es gibt insgesamt vier Bewegungsdeskriptoren, die in zwei globale und zwei regionale Deskriptoren aufgeteilt werden. Auf der einen Seite sind der Motion Activity Deskriptor, der die wahrgenommene Handlungsgeschwindigkeit beschreibt, und der Camera Motion Deskriptor, der die verschiedenen Kameraoperationen erkennt, die zwei globalen Deskriptoren. Auf der anderen Seite beschreiben der Motion Trajectory Deskriptor, der die raumzeitlichen Veränderungen eines Objektes darstellt, und der Parametric Motion Deskriptor, der die Verformungen von Objekten erfasst, Bewegungsregionen in einem Videosegment.

Es gibt viele Anwendungsmöglichkeiten, die mit Hilfe der vier Bewegungsdeskriptoren realisiert werden können, wie z.B. Videobrowsing oder Videozusammenfassung. Werden sie mit den anderen Deskriptoren des MPEG-7 Standards, wie z.B. Farb oder Textur Deskriptoren, kombiniert, so können die Anwendungen noch zusätzlich verbessert werden.

References

1. N. Day, J. Martinez, "Introduction to MPEG-7", ISO/IEC JTC1/SC29/WG11 N3751, 2000
2. S. Jeannin, A. Divakaran, B. Mory, "Motion Descriptors", in Introduction to MPEG-7 , 261-280
3. S. Jeannin, A. Divakaran, "MPEG-7 Visual Motion Descriptors", in IEEE Transactions on circuits and systems for video technology, vol. 11, no. 6, 2001
4. A. Divakaran, K. Peker, R. Radhakrishnan, Z. Xiong, R. Cabasson, "Video Summarization using MPEG-7 Motion Activity and Audio Descriptors", TR-2003-34, 2003
5. K. Peker, A. Divakaran, "Framework for measurement of the intensity of motion activity of video segments", TR2003-64, 2003
6. S. Pfeifer, R. Lienhart, S. Fischer, W. Effelsberg, "Abstracting digital movies automatically", in J. Visual Comm. Image Representation, vol. 7, no. 4, pp. 345 - 353, 1996
7. N. Vasconcelos, A. Lippman, "Towards semantically meaningful feature spaces for the characterization of video content", Proc. of ICIP97, 1997
8. J. Oh, V. Gavirneni, "Automatic Generation of Motion Activity Descriptors in Video Sequences", University of Texas at Arlington, Department of Computer Science and Engineering
9. "MPEG-7 Visual part of the XM 4.0", ISO/IEC MPEG99/W3068, Maui, USA, 1999
10. T. Zaharia, F. Preteux, "Parametric Motion Models for Video Content Description within the MPEG-7 Framework", ARTEMIS Project Unit
11. J. Llach, P. Salembier, "Analysis of Video Sequences: Table of Contents and Index Creation", Universitat Politècnica de Catalunya, Campus Nord
12. H. Eidenberger, "A Video Browsing Application Based on Visual MPEG-7 Descriptors and Self-Organising Maps", in International Journal of Fuzzy Systems, Vol. 6, No. 3, 2004