دانشکده مهندسی کامپیوتر

موضوع

مطالعه انواع ساختارهای معنایی و روشهای متفاوت استخراج مفاهیم از داده های تصویری- صوتی

گزارش سمینار کارشناسی ارشد
در رشته مهندسی کامپیوتر

دانشجو: **بهرنگ قاسمی زاده**
استاد راهنما: **دکتر محمدرضا کنگاوری**

سال تحصیلی۱۳۸۳- ۱۳۸۴

تشکر و قدردانی

از استاد ارجمند، جناب آقای دکتر کنگاوری که با راهنمایی های خردمندانه اشان، بنـده را جهت انجام این مطالعه یاری فرمودند کمال تشکر و قدردانی را دارم.

تقدیم به:

پدر و مادر عزیزم

و استاد گرامی جناب آقای دکتر کنگاوری

چکیده

ارائه معنی برای داده ها یکی از زمینه های تحقیقاتی با قدمت طولانی در هوش مصنوعی است. با کاربرد و گسترش روزافزون استفاده از داده های چند رسانه ای به خصوص گسترش تکنولوژی در زمینه بانک های داده ویدئویی و ویدئوهای دیجیتال و پدید آمدن حجم عظیمی از این داده ها جهت پردازش و ارائه به کاربر، وجود ابزارهایی جهت پردازش مفهومی و جستجو و بازیابی مبتنی بر معانی به جای دیگر ویژگیهای سطح پایین، از مباحث تحقیقاتی مهم در سالهای اخیر می باشد.

بزرگترین مشکل در این زمینه شکاف معنایی است و کلیه تلاشها برای حل این مشکل متمرکز شده است. هدف نهایی دستیابی به مدلی است که با استخراج خودکار ویژگیهای سطح پایین تصویر به مفاهیم معنایی سطح بالا برسد. این به معنی پوشش شکاف معنایی است. تا به حال هیچ یک از سیستمها چنین ویژگی را ارئه نداده اند.

جهت رسیدن به این هدف، نیاز است تا با توجه به خصیصه های این نوع داده، هماهنگی دقیق میان روشهای ذخیره و بازیابی و مدل کردن معنایی آن در نظر گرفته شود. در حال حاضر پردازش و بازیابی این داده ها به کمک مفاهیم معنایی محتوی داده ویدئویی با نقاط ضعف فراوانی همراه است. بهبود روشهای ارئه معنی و مدل کردن آن، می تواند راه گشا و کلید حل مسئله در این زمینه باشد. استفاده از ابزارهای ارائه دانش یکپارچه همانند هستان شناسی ها می تواند یک راه حل مناسب برای بازنمایی دانش معنایی باشد. علاوه بر اینکه در استفاده از هر ابزار ارائه دانش نیاز است تا به ویژگیهای مختص داده های ویدئویی، یعنی وابستگی زمانی و فضایی توجهی خاص مبذول گردد.

در مستند پیش رو در باره هر یک از مسائل مهم در ارائه داده های ویدئویی، روش های ذخیره ، بازیابی و شاخص بندی این داده ها بحث شده است. سپس توضیح داده می شود که چگونه حاشیه نویسی ها به عنوان آخرین دستاورد جهت ارائه مدل های معنایی از این داده ها مورد توجه قرار می گیرد.

فهرست مطالب

فهرست شکل ها

فهرست جدول ها

۱ مقدمه

با پیشرفت تکنولوژی و قدرتمند شدن ابزارهای پردازشی و سیستمهای ذخیره و بازیابی اطلاعات، و به موازات آن رشد کاربران کامپیوتری و استفاده روز افزون از داده‌های چند رسانه ای در این مقوله، بحثهای فراوانی جهت افزایش کارایی سیستمهای آرشیو و بازیابی داده های چند رسانه ای و متعاقب آن تصاویر ویدئویی و کاربردهای آن در گرفته است. سیستمهای مدیریت بانک‌های اطلاعاتی متداول، در مدیریت داده‌های ساخت یافته بسیارموفق عمل نموده‌اند اما در ادامه از ارائه یک مدل موفق و کارا برای داده‌های تصویری و سایر انواع داده‌های غیر ساختیافته، توفیق چندانی نداشته اند. شاید بزرگترین مشکل در این زمینه کار آمد نبودن مدلهای شاخص‌بندی، جستجو و بازیابی برای این نوع از داده ها باشد که از طبیعت خاص و حجم بالای آنها حاصل شده است. با توجه به افزایش فزاینده استفاده از این نوع داده‌ها، نیاز به چهارچوب‌های کاری جدید برای ارائه ساده و در عین حال با معنی از این نوع داده‌ها، که تضمین کننده کارآیی در بازیابی و جستجوی آنها در این سیستم باشد، بیش از هر زمان دیگری احساس می‌شود. این سیستمها بایستی پرس و جو بوسیله داده‌های وابسته به زمان ساده، تا پرس و جوهای مبتنی بر محتوی و معنا را پوشش دهند. یک نکته مهم در این زمینه آنست که پرس و جو بصورتی ساده و طبیعی بتواند برای کاربر فرموله شود که جستجوهای مبتنی بر متن نخستین گام در این راه است [5] [2] [1].

با توجه به آنچه گفته شد استاندارد MEPG 7 به عنوان واسط استاندارد برای تشریح محتویات یک داده صوتی- تصویری در سطوح مختلف معنایی معرفی. هدف MPGE7 پیشنهاد ساختار فرادادهای است برای تشریح و حاشیه نویسی محتویات داده های صوتی و تصویری در دامنه‌ای از ویژگیهای سطح پایین سیگنال تا بالاترین سطوح معنایی.[3] اما متاسفانه این استاندارد هیچ گونه ساختار و روند مشخصی را برای چگونگی استخراج مفاهیم از داده‌های صوتی و تصویری ارائه نمی‌نماید. این مسئله امروزه به عنوان یکی از زمینه های تحقیقاتی هوش مصنوعی مطرح می‌باشد.[4]

روشهای متفاوتی جهت بدست آوردن مفاهیم از یک ویدئو وجود دارد. بطور کلی این روشها به سه دسته تقسیم می‌شوند[4]:

مفاهیم مبتنی بر حاشیه نویسی[1]: توضیحاتی سمبلیک از یک ویدئو در یک زمینه خاص_ با استفاده از دانش زمینه ای خاص _ ، ارائه می نماید. در این روش، با داشتن دانش زمینه ای مربوط به یک ویدئوی خاص همانند ویدئوهای ورزشی و یا خبری ، توانایی و امکان استخراج خودکار معنی و مفهوم انتزاعی و کلی از ویدئو وجود خواهد داشت. نمونه ای از این کارها را می توان در [8],[7] که اختصاص به ویدئوی خبری دارد و در[9] و [10] که در زمینه ویدئوهای ورزشی است، مشاهده نمود. اخیرا کمیته MPEG یکی از زبانهای متعلق به خانواده زبانهای XML را به عنوان استاندارد تشریح و توضیح در MPEG7 برگزیده است. [3]

مفاهیم مبتنی بر ویژگیهای سطح پایین تصویر[2]: این سطح مفهومی، اشاره به المانهایی دارد که می‌تواند بصورت خودکار و بدون نیاز به در نظر گرفتن دانش خاص مربوط به یک زمینه، از کلیه ویدئوها استخراج شود. این به معنی استفاده از الگوریتمهایی است که بر روی ویژگیهای سیگنالی فریمها کار می کنند و نتیجه را به صورت ویژگیهایی همچون رنگ، اندازه، شکل و ... بیان می کنند. روش های شاخص گذاری بااستفاده از این متد بسیار سطح_پایین (از لحاظ مفهومی)هستند در نتیجه نمی توان از آنها مستقیما در پرس و جو ها استفاده نمود. کارهای زیادی در این زمینه صورت گرفته است که بیشتر بر روی تکنیکهای پردازش تصویر استوار است. نمونه ای از آن شامل [11] است که در آن به شناسایی اشیا متحرک در یک ویدئو به کمک ویژگیهای تصویر می پردازد.

---

[1] Annotation-Based Semantics

[2] Low-Level Content-Based Semantics

مفاهیم مبتنی بر ساختار[1]: در این جا به دنبال پیدا نمودن ساختار یک ویدئو هستیم که این ساختار غالباً بصورت یک سلسله مراتب از برنامه[2] ها، سکانس[3] ها، منظره[4] ها و فریم[5] ها است. به عبارت دیگر مفاهیم ساختاری یک ویدئو تشریح می شود. کار دراین زمینه همچنان ادامه دارد.نمونه ای از کارها که برای تشخیص منظره انجام شده است، [12] است. گروهی از روشها نیز به دنبال پیدا نمودن روش هایی جهت تشخیص سکانس های یک ویدئو هستند. از جمله این کارها می توان به [13] اشاره نمود که در آن به کمک اطلاعات مناظر مختلف تصویر و اطلاعات راهنمای دیگر همچون همبستگی میان صداهای مختلف در مناظر متفاوت تصویر و استفاده از گفتگوها، به دنبال تشخیص سکانسهای مختلف یک ویدئو است. بطور کلی مفاهیمی که در این سطح ارائه می شود، همچون فهرستی بر یک کتاب است که می توان از آن برای ردیابی و تعیین محل یک مورد خاص، استفاده نمود.

بطور اخص، کاربرد این مفاهیم در روش‌های شاخص‌گذاری بر روی انواع داده‌های ویدئویی است[6]. مدل نمودن مناسب و کارآیی هر یک از این مدل ها، تضمین کننده یک عملکرد مناسب از بانک‌های داده‌ه ای ویدئویی خواهد بود. همانطور که پیش تر توضیح داده شد، عصر دیجیتال به همراه شبکه جهانی WEB ،دستیابی به داده ها را در مقیاسی عظیم فراهم آورده اند. بیشتر پایگاه های داده، دستیابی به داده ها را در یک سطح مفهومی پایین فراهم می آورند. کاربر با میزان قابل توجهی از افزونگی داده[6] مواجه است که در نتیجه آن میبایستی مقدار زیادی وقت و انرژی خود را برای بدست آوردن اطلاعات مورد نیاز از میان

---

[1] Structured-Based Semantics

[2] Clip

[3] Scene

[4] Shot

[5] Frame

[6] Data Overload

داده های موجود، صرف نماید. یکی از روشها برای حل این مشکل فراهم نمودن یک جستجوی مفهومی است [6] .فراهم آوردن دسترسی به داده های ویدئویی در سطحی مفهومی، نیازمند سیستمهای مدیریت داده های ویدئویی ، متناسب و درخور این دامنه از داده ها است. شاخص گذاری کارا و بازیابی مناسب برای یک سطح دسترسی بالا نیاز به دانش زمینه ای را ملزوم می دارد.

بازیابی و شاخص گذاری ویدئو، یک زمینه کاری جوان است که بطور اخص در هوش مصنوعی، پردازش سیگنالهای دیجیتال و پردازش زبانهای طبیعی، مفاهیم پایگاه داده، شناسایی الگو و بینایی ماشین ریشه دارد. اما هیچ یک از زمینه های کاری ذکر شده مستقیما قادر به حل این مشکل نیستند. بجای آن، حل مسئله را می توان در اشتراکی میان این زمینه های کاری ذکر شده ، پیدا نمود.

به نظر می آید که در این میان بزرگترین و مهمترین چالش، پر کردن شکاف مفهومی موجود در این زمینه است.به این معنی که بیشتر ویژگیهای سطح پایین بسادگی قابل قابل محاسبه و اندازه گیری هستند اما نقطه شروع در یک عملیات وابسته به یک پایگاه داده ویدئویی، عموما یک پرس و جوی سطح بالا از یک انسان است. ترجمه و تبدیل پرسش های سطح بالای این انسان به ویژگیهای سطح پایین که کامپیوترها قادر به کار کردن با آنها هستند، مشکل پر کردن این شکاف مفهومی را می نمایاند. مشکل اساسی با یک پرس و جو، فهمیدن منظوری است که در پشت آن پرس و جو قرار دارد. مسائلی همچون چگونگی مدل کردن بینایی انسان، تشخیص و شناسایی اشیا و چگونگی قطعه بندی نمودن یک ویدئو، از دیگر چالش های موجود در این زمینه است.[14]

یکی دیگر از بحثهای داغ در زمینه نگهداری از داده های ویدئویی ، خلاصه سازی داده های ویدئویی[1] است .اهمیت این خلاصه ها از آنجاست که می توانند کاربر را به سرعت از اطلاعات مهم یک ویدئوی

[1] Video Summarizations

طولانی با خبر نمایند.روش های خلاصه کردن ویدئو در دو دسته کلی قرار می گیرند، یکی استفاده از روشهای مبتنی بر قانون و دیگری استفاده از روشهایی مبتنی بر ریاضیات. ارائه مفهوم و مدل کردن آن در یک داده ویدئویی به خصوص در روشهای تلخیص ویدئو- مبتنی بر قانون به چالش طلبیده می شود. [15]

همانطور که گفته شد یکی از زمینه های تحقیقاتی مهم در حال گسترش،بحث شاخص بندی و بازیابی اطلاعات تصویری است. مشاهدات نشان می دهد که روشهای ارائه این اطلاعات به کمک ویژگیهای سطح پایین به اندازه کافی مناسب نیستند و گرایش گسترده ای به سمت پر کردن شکاف های مفهومی صورت گرفته است.در این سمینار درباره هر یک از روشهای ارائه مفهوم در داده‌های ویدئویی و نقاط ضعف و کاربردهای آن بحث خواهد شد و نظریات ومشی‌های متفاوت در هر یک از این سطوح تشریح خواهند شد.

۲ مشکلات پیش رو در ساخت بانک های داده ویدئویی

امروز مقادیر بسیار زیادی از داده‌های ویدئویی تولید و در خدمت استفاده کاربران است. علت این امر را می‌توان در وسائل سخت‌افزاری ذخیره و بازیابی ارزانتر و بهتر، شبکه‌های وسیع و سریعتر، عمومی شدن اینترنت و WEB ، گسترش و ایجاد محصولات و سرویسهایی جدید و فرمت‌های ذخیره همچون DVD ها دانست. به هر دلیل فراوانی داده‌های ویدئویی به دلیل محتویات زیادی که شامل تصاویر، صدا متن و حرکات بصورت ترکیب شده میباشد، از اهمیت قابل بحثی برخوردار است. مشکلاتی که در راه ساخت پایگاه داده‌های ویدئو وجود دارد:

داده‌های خام ویدئویی به تنهایی استفاده و کارآیی محدودی دارند که در نتیجه آن نیاز به حاشیه نویسی[۱] احساس می‌شود. حاشیه نویسی به صورت دستی کاری سخت، زمان بر، نادقیق، ناتمام و .. .و از همه مهمتر هزینه داراست. در نتیجه نیاز به پیداکردن الگوریتمها و سیستمهایی خلاق[۲] که اجازه تشریح، سازماندهی و مدیریت داده‌های ویدئویی را با درکی از مفاهیم معنایی آن به صورت اتوماتیک یا نیمه اتوماتیک داشته باشند نیاز می‌شود و در نتیجه آن بحث بازیابی داده های ویدئویی مبتنی بر محتویات[۳] آن مطرح می‌شود.

- حجم بالای داده‌های خام شامل: ویدئو، صدا، متن
- مخازن داده‌ای توزیع شده[۴] که در عین حال همیشه ساختمند نیستند.
- وجود برنامه‌های ویدئویی متنوع که هر یک ساختار و قوانین خاص خود را دارند.
- درک یک ویدئو بسیار به زمینه[۵] وابسته است.
- کاربران متفاوت در پلات فرم های متفاوت، نیازهای متفاوتی دارند.

در این فصل سعی می‌شود تا مفاهیم و نظریات اصلی در پایگاه داده‌های ویدئویی و مشکلات موجود در پس پرده طراحی آن بحث و بررسی شود- به خصوص به موانعی که یک پایگاه داده ویدئویی را از یک

---

[1] Annotation

[2] Creative

[3] Content-Based Video Retrieval

[4] Distributed

[5] Context

پایگاه داده معمولی متفاوت می‌سازد بحث شود درباره بعضی از روشها و محصولات تجاری موجود بحث شود.

۳ مفاهیم پایه در مدیریت سیستمهای بانک دادهٔ ویدئویی

اولین هدف یک سیستم مدیریت بانک داده های ویدئویی دسترسی مستقیم و اتفاقی به داده‌های متوالی[۱] ویدئویی بصورت ساختگی و کاذب است. برای دستیابی به این هدف می‌بایست که یک ویدئو را به قسمت‌هایی تقسیم‌بندی نمود. این قسمتها را ایندکس نمود و این ایندکسها را به گونه‌ای نمایش داد که اجازه Browsing و بازیابی آسانتر را بدهد. بصورت پایه یک سیستم مدیریت بانک داده های ویدئویی یک پایگاه داده از ایندکس و یا نشانه روهاست. تفاوتهای یک سیستم مدیریت بانک داده های ویدئویی با یک سیستم مدیریت بانک داده ای سنتی را می‌توان در موارد زیر دانست:

- داده‌های خام ویدئویی نیاز دارند تا مدل شوند، ایندکس گذاری شوند و به شکلی ساختمند در آیند.
- پرس و جو و بازیابی متعاقب آن عموماً بوسیله الگوریتمهای بازیابی مبتنی بر شباهت[۲] پی‌ریزی می‌شود که ارتباط آن را با بازیابی اطلاعات[۳] نشان می‌دهد.

طراحی واسط های کاربری گرافیکی[۱] به شدت اهمیت دارد. [۱۶]

---

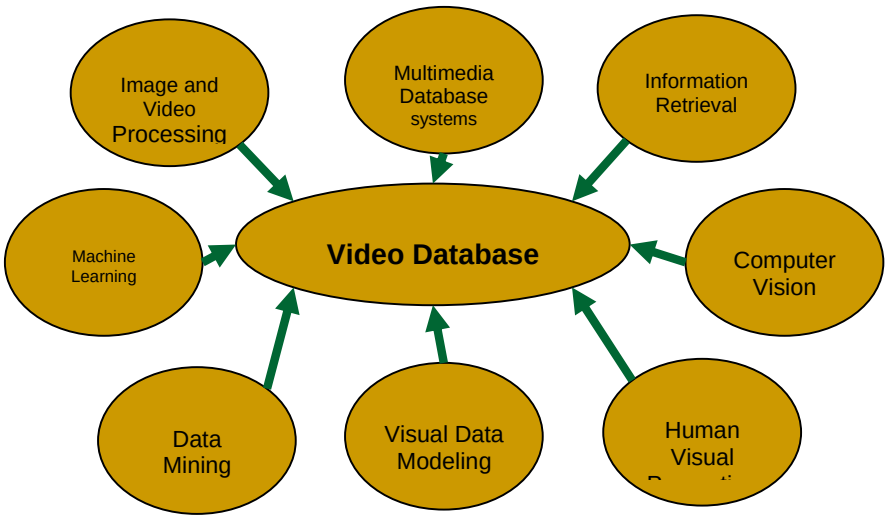
[1] Sequential

[2] Similarity- Based retrieval

[3] Information Retrieval

سیستم های مدیریت بانک داده های ویدئویی[۲] به میزان وسیعی در ارتباط با فیلدهای دیگر همچون پردازش تصویر و ویدئو، بازیابی اطلاعات بنیایی ماشین، مدل کردن داده‌های تصویر و ... (شکل ۱) هستند.

بلاک دیاگرام مربوط به یک سیستم مدیریت بانک داده ویدئویی در شکل ۲ نشان داده شده است. هر یک از بلاکهای تشکیل دهنده یک وظیفه خاص را جهت آماده سازی داده ها بر عهده دارد. شرح وظایف هریک از این بلاکها در ادامه آمده است.

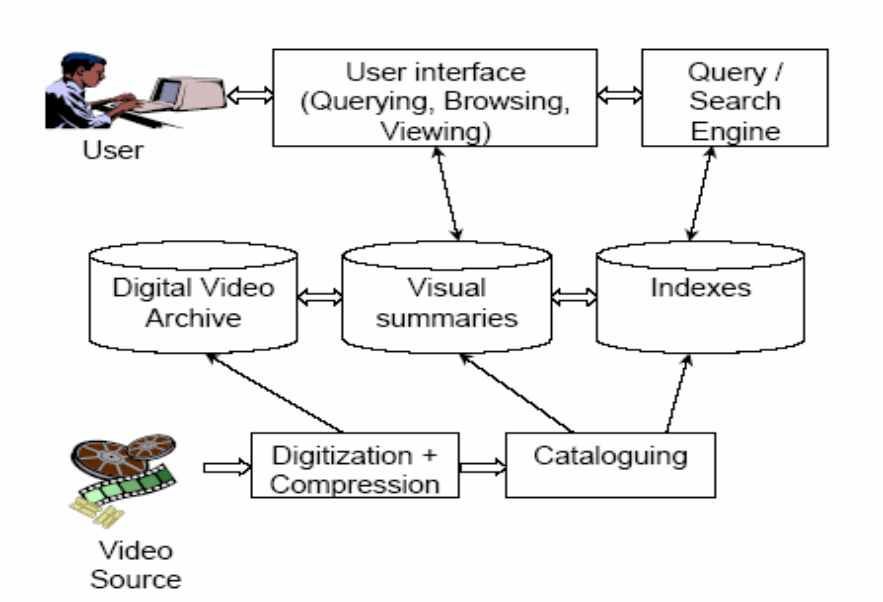


شکل 1. رابطه میان سیستم های پایگاه داده ویدئویی با دیگر زمینه های تحقیقاتی

---

[1] Graphical User Interface

[2] Video Database Management System (VDBMS)



شکل ۲. بلاک دیاگرام مربوط به یک سیستم مدیریت بانک داده ویدئویی

Digitization and Compression. این واحد وظیفه تبدیل داده های آنالوگ ویدئویی خام به داده های دیجیتال را بر عهد دارد. برای این منظور نیاز به نرم افزار و سخت افزار خاص می باشد.

Cataloguing . داده‌های ویدئویی پس از دیجتایز شدن و فشرده‌سازی نیاز به فهرست نگاری[1] دارند. این به معنی پردازشهایی جهت استخراج واحدهای داستانهای با معنی از داده‌های خام ویدئویی و ساختن ایندکس‌های متناسب با آن است.

[1] cataloguing

Visual Summaries. ارائه دهنده محتویات به شکلی موخر و فشرده و غالباً به صورت سلسله مراتبی است.

Digital Video Archive. آرشیو داده های ویدئویی دیجیتال شده

Indexes. آرشیو ایندکس ها که اشاره گرهایی به قطعات ویدئویی هستند .

Query/Search Engine. امکان جستجو بر اساس پارامترهایی که کاربر مشخص نموده است، فراهم می آورد.

User Interface. یک واسط کاربر با رابط های گرافیکی که امکاناتی از قبیل Browsing، پرس وجو[1] و تماشای نتایج بدست آمده را به شکل در تقابل با کاربر[2] به کاربر می دهد

۳,۱ سازماندهی داده های ویدئویی

از آن جا که ویدئو یک رسانه ساختمند است که در آن اتفاقات در زمان و محل، داستان را بیان می‌کند، یک برنامه ویدئویی می بایست به صورت یک مستند در نظرگرفته شود نه به صورت یک رشته متوالی غیر ساختمند از فریمها. پردازش تبدیل داده‌های خام ویدئویی به واحدهای ساخته یافته که می توانند برای

---

[1] Query

[2] Interactive

ساختن یک نمایه از آنچه در یک برنامه ویدئویی دیده می‌شود استفاده گردد به عنوان تجرید ویدئو[1] نیز ارجاع داده می‌شود که در دو دسته کلی تقسیم‌بندی می‌شود:

- مدل نمودن ویدئو و ارائه آن
- قطعه‌بندی ویدئو[2] و خلاصه سازی

به عبارت دیگر تجرید ویدئو فرایند تبدیل داده های خام ویدئویی به واحدهای ساختیافته جهت ساختن یک نمایه از آنچه که در یک برنامه ویدئویی دیده می شود، می باشد[۱۲].

۳,۲ مدل نمودن ویدئو و ارائه آن

پروسه طراحی برای انتخاب شیوه ارائه داده‌های ویدئویی مبتنی بر خصوصیات ویدئو،اطلاعات محتوی ویدئو و کاربردی است که برای آن منظور به کار می‌رود. مدل نمودن ویدئو نقش کلیدی در سیستمهای مدیریت بانک داده ویدئویی بازی می‌کند چرا که بقیه کارها حتماً کم و بیش به آن وابسته است. داده‌های ویدئویی حاوی اطلاعات بیشتری نسبت به مستندات متنی هستند. تعبیرها و تفسیرها عموماً همراه با ابهام هستند و بستگی به بیننده و زمینه کاربردی ان دارد. این مدلها غالباً ساختار روشن و محکمی در زیربنای خود ندارند. روابط میان قطعات ویدئو پیچیده و غالباً تعریف نشده‌اند به عنوان مثال هنوز یک اپراتور نشان دهنده میزان شباهت که عموماً مورد قبول قرار گیرد وجود ندارد. [۱۷]

---

[1] Video Abstraction

[2] Parsing

محتویات یک داده ویدئویی شامل محتویات مفهومی[1] ، محتویات تصویر و صدا و محتویات متنی[2] است. محتوی مفهومی، ایده یا دانشی است که ویدئو به طور ضمنی به کاربر القاء می‌نماید. این دانش غالباً مبهم است، به صورت ذهنی بوده و به زمینه کاربرد آن بستگی دارد. محتویات تصویر صدا شامل رنگ، زمینه، شکل ، حرکت شیءها، روابط میان اشیاء، حرکت‌های دوربین، صداها و ... می‌شود. محتویات متنی شامل عنوان‌ها. زیر عنوان‌ها و ... می‌باشد که نیاز به OCR[3] را مطرح می‌نماید. نکته‌ای که در این بیان توجه به آن ضروری است ان است که محتویات یک داده ویدئویی به یک اندازه از اهمیت برخوردار نیستند.

مدل ویدئویی که ارائه می‌شود باید دارای ویژگی‌های زیر باشد: [۱۷]

- می‌بایستی داده ویدئویی را به عنوان یکی از ساختار داده‌هایش دقیقاً همانند ستون و یا داده‌های عددی پشتیبانی و حمایت نماید.
- مشخصات محتویات داده‌های ویدئو را در ساختار مفهومی خود مجتمع نماید.
- اطلاعات صوتی را به همراه داده‌های مجری همراه نماید.
- توانایی بیان روابط زمانی و ساختاری میان قطعات را داشته باشد.
- بتواند به صورت خودکار ویژگی‌های سطح پایین همچون (رنگ، بافت، شکل ، حرکت) را استخراج نماید و از آنها استفاده نماید.

---

[1] Semantic

[2] Textual

[3] Optical Character Recognition

سطوح سلسله مراتبی از تجرید جریان داده های ویدئویی[۱] در کاهش میزان دانه دانه بودن[۲] به شکل زیر است:

- فریم کلیدی[۳] : فریمی از یک منظره[۴] که بیشترین اطلاعات را داشته باشد.
- منظره : رشته ای از فریمها که به صورت متصل ضبط شده‌اند و ارائه دهنده یک حرکت ممتد (ادامه دار) در زمان و یا فضا است.
- گروه[۵] : موجودیت میانی منظره های فیزیکی و جزء صحنه‌های مفهومی و همانند پلی میان ایندو عمل می‌نماید.
- سکانس[۶] : مجموعه‌ای از منظره های مفهوماً به همه وابسته و از لحاظ زمانی مجاور، که مفاهیم و یا داستانها را در سطحی بالا تعریف و اشاره می‌نمایند
- برنامه ویدئویی: کل یک برنامه ویدئویی[۷] ویدیویی است. [۸]

---

[1] video stream abstraction

[2] granularity

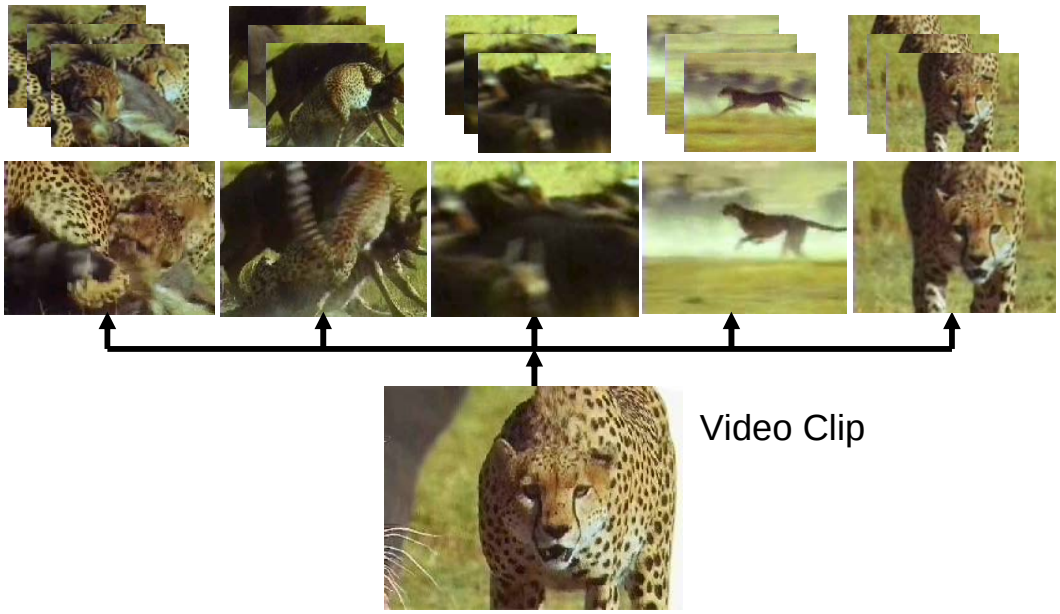
[3] Key frame

[4] Shot

[5] Group

[6] Scene or sequence

[7] Clip



شکل ۳. سطوح تجرید در قطعه بندی یک ویدئو. در بالاترین سطح فریمها و در پایین ترین سطح یک برنامه ویدئویی قرار دارد.

همچنین یک مدل ویدئویی می بایستی که اشیاء فیزیکی روابط زمانی وفضایی میان آنها را تشخیص دهد. روابط زمانی می‌توانند (باید) بصورت عبارتی همچون قبل از، در طی، شروع شد، همپوشانی... غیره بیان شود. روابط فضایی می تواند مبتنی بر یک سیستم مختصاتی دو بعدی یا سه بعدی باشد.

یک مدل ویدئویی می‌بایستی از حاشیه نویسی و یا به عبارت دیگر اضافه نمودن فراداده‌ها به ویدئو پشتیبانی نماید. سه دسته عمومی از متاداده‌ها غالباً بدین منظور به کار می‌روند: ۱- فرا داده‌های وابسته به محتویات به عنوان مثال ویژگی‌های مربوط به صورت یک گوینده خبر ۲- فراداده‌های تشریحی

محتویات به عنوان مثال میزان خوشحالی یا عصبانیت بر پایه حالتهای صورت ۳- فراداده‌های مستقل از محتویات که برای مثال می‌تواند شامل نام کارگردان فیلم باشد.

مدلهای داده‌های ویدئویی شامل گروههای انواع زیر می باشد:

- مدلهای مبتنی بر قطعات ویدئویی[۱]
- مدلهای مبتنی بر لایه‌های حاشیه نویسی [۲]
- مدلهای اشیاء ویدئویی[۳]
- مدلهای داده‌های ویدئویی جبری[۴]
- مدلهای آماری[۵]
- بقیه موارد ...

ساخت مدلهای مبتنی بر قطعات ویدئویی شامل دو قدم اساسی است: ۱- قطعه‌بندی جریان ویدئویی به یک مجموعه از واحدهای پایه منظم شده بر اساس زمان (عموما منظره ها)، ۲- ساخت مدلهای وابسته به دامنه (به صورت سلسله مراتبی و یا ماشینهای حالت محدود) بسته به نوع واحدهای پایه .

---

[1] Based On Video Segmentation

[2] Based On Annotation Layered

[3] Algebraic video data Model

[4] Algebraic video data Model

[5] Statistical Models

در مدلهای مبتنی بر لایه های حاشیه نویسی شده که به نام مدلهای چینه بندی[1] نیز شناخته می‌شوند اطلاعات زمینه‌ای ویدئو را قطعه‌بندی نموده و تقریبا دیدگاه تدوینگر ویدئو را تخمین می‌زند. ایده اصلی آن است که اگر حاشیه نویسی در سطح مناسب و دقیقی از تجرید صورت پذیرد آنگاه اطلاعات کلی را میتوان به سادگی بدست آورد.

مدلهای مبتنی بر اشیا ویدئویی از مقایسه بین مدلهای داده‌ای رابطه‌ای شی گرا[2] و گسترش آن به داده های ویدئویی به دست می آید. مزایای این مدل شامل توانائی ارائه و مدیریت اشیاء پیچیده، Encapsulate نمودن داده‌ها و متدهای مربوط به آن در یک شی و توارث ساختار خصیصه‌ها و متدهای مبتنی بر کلاسهای سلسله مراتبی است. اما این روش محدودیت های خاص خود را نیز به همراه دارد. داده خام ویدئویی مستقل از شکل و محتویات آن و ساختار پایگاه داده‌ای که بعداً در طی مرحله حاشیه نویسی می‌آید، ساخته می‌شود. شمای داده‌ها برای داده‌های ویدئویی، ایستا و ساکن نیست بدین معنی که توضیحات مربوط به یک داده ویدئویی به کاربرو برنامه کاربردی وابسته هستند و اطلاعات غنی دلالت بر آن دارد که منظورهای معنایی به صورت افزایشی[3] اضافه شوند. از آنجا که داده‌های ویدئویی همپوشانی دارند و یا شامل یکدیگر هستند پشتیبانی از توارث فراگیر (شمول) (به اضافه توارث مبتنی بر کلاس) مدنظر است. مثالی از یک سیستم شی‌گرا OVID [۱۸]است. در این سیستم شی ویدئو[4] به صورت یک رشته دلخواه از فریمها تعریف می‌شود. هر شی ویدئو شامل یک کدشناسه یکتا، شماره فریم شروع و پایان، محتویات که به وسیله زوجهای خصیصه و ارزش به صورت دستی آماده شده‌اند می باشد. این مدل بدون

---

[1] Stratification models

[2] Object Oriented

[3] Incremental

[4] Video object

شِما[1] است بدین معنی که اجازه می‌دهد هر خصیصه‌ای به صورت دلخواه به شی ویدئو اضافه شود همچنین سیستم توارث فراگیر بازه‌ای[2] را پشتیبانی می نماید.

مدل های ویدئویی جبری یک جریان ویدئو[3] را به وسیله اعمال یکسری عملگرهای جبری به صورت بازگشتی بر روی قطعات خام ویدئویی تعریف می‌نمایند. موجودیت اصلی ارائه است . ارائه‌ها به وسیله مبین‌های ویدئویی تشریح می‌شوند که از قطعات خام به کمک عملگرهای جبری ساخته شده‌اند. نمونه‌ای از این عملگرهای جبری شامل موارد ذیل است:

- Creation شامل Create و Delay
- Composition شامل Concatenation ، Union ، Intersection
- Output شامل Window و Audio
- Description شامل Description و Hide- Content

مثالهایی از این مدل شامل [۱۹][۲۰][۲۱] می‌باشد.

در مدل های آماری از دانش مربوط به ساختار ویدئو همچون وسیله‌ای برای فراهم آوردن طراحی اولیه مفاهیم ویدئویی استفاده شود. در این روش از تکنیکهای یادگیری ماشین استفاده می‌شود (به عنوان مثال استنتاج بیزی) برای یادگیری معانی از مجموعه‌هایی از نمونه‌های آموزشی بدون آن که تکیه‌ای بر روی خصیصه‌های سطح پایین همچون زمینه، رنگ و یا .. .داشته باشیم.

---

[1] Schema less

[2] interval inclusion inheritance

[3] Video stream

۴ خلاصه سازی و قطعه بندی ویدئو[1]

قطعه بندی کردن ویدئو یا parsing قطعه‌بندی زمانی از محتویات یک ویدئو به واحدهایی کوچکتر است. تکنیک‌های قطعه بندی کردن ویدئو[2] اطلاعات ساختاری یک برنامه ویدئویی را بوسیله شناسایی مرزبندی‌های زمانی استخراج می‌نمایند و قطعات با معنی را تشخیص می‌دهند که عموماً منظره نامیده می‌شوند. خلاصه ویدئویی تلاش می‌کند تا یک خلاصه تصویری از ویدئویی متضمن آن ارائه نماید به شکلی موجزتر،به کمک حذف افزونگی‌ها.

قطعه بندی ویدئو می‌تواند در سطح یک منظره و یا در سطح سکانس اتفاق بیفتد. منظره ها عموما کوچکترین واحد مورد علاقه هستند. یک منظره (یک حرکت متوالی روی صفحه که به نظر می‌آید در نتیجه یک حرکت دوربین باشد) عموماً کوچکترین شی مورد علاقه در این زمینه است. منظره ها می‌توانند به صورت اتوماتیک تشخیص داده شوند و غالباً به کمک فریمهای کلیدی نمایش داده می‌شوند. تشخیص مناظر ویدئویی[3] پروسه تشخیص حرکت میان دو منظره متوالی است که در طی آن رشته فریمهایی که متعلق به هر یک از منظره ها هستند با یکدیگر در یک گروه قرار می‌گیرند. به طور کلی دو دسته از حرکت‌های منظره[4] وجود دارد:

- حرکت‌های ناگهانی که شامل کات[5] یا برش است.

---

[1] Video segmentation

[2] Video Parsing

[3] Shot detection

[4] Shot transitions

[5] Cut

- حرکت‌های تدریجی که شامل fade in و fade out و dissolve است.

تشخیص سکانس های ویدئویی[1] تشخیص اتوماتیک مرزبندی‌های معنایی در یک برنامه ویدئویی است و برخلاف مرزبندی‌های فیزیکی، یکی از کارهای سخت و مشکل زا است و یکی از موضوعات تحقیقاتی روز است چراکه ارائه راه‌حل برای آن نیاز به آنالیز محتویات در سطحی بالاتر، سطح معنایی، دارد. سه استراتژی در این زمنیه ارائه و قبول شده است:

اولین راه بر مبنای قوانین تولید فیلم است (همانند افکتهای حرکتی، تکرار،منظره ها، حضور موسیقی متن فیلم و ...) جهت تشخیص گره‌های زمانی و محلی از تغییرات بزرگ.

راه حل دوم الگوریتم های مبتنی بر خوشه بندی زمانی محدود شده[2] بر پایه این عقیده کار می کند که محتویاتی که از لحاظ معنایی به یکدیگر وابسته‌اند گرایش به محلی شدن یا جمع شدن در یک زمان دارند.

الگوریتم های مبتنی بر مدل های از پیش تعریف شده[3]، بر پایه مدلهای ساختاری خاصی است برای برنامه‌هایی که در آنها ساختارهای زمانی غالباً بسیار محدود و قابل پیش بینی است همانند اخبار و برنامه های ورزشی.

عموما الگوریتمهای تشخیص منظره ها با توجه به نوع ویدئویی که به عنوان ورودی دریافت می کنند تقسیم بندی می شوند که شامل قطعه‌بندی ویدئو در محدوده غیر فشرده[4] و فشرده[1] می باشد. در

---

[1] Scene-Based Video Segmentation

[2] Constrained Time Clustering

[3] A Priori Model - Based Algorithm

[4] Uncompressed

محدوده ویدئو های غیر فشرده ایده اصلی بر پایه اندازه شباهت تعریف شده میان فریمهای متوالی است. به عبارت دیگر هنگامی که دو تصویر به اندازهٔ کافی با یکدیگر تفاوت داشته باشند ممکن است که یک کات وجود داشته باشد. حرکت‌های تدریجی نیز می توانند با استفاده از اندازه‌گیری تفاوت انباشته[۲] و استفاده از طرح‌های آستانه مناسب تشخیص داده شوند. در این میان سه روش عمده برای اینکار وجود دارد: روشهای پیکسلی[۳]، مبتنی بر بلاک[۴] و روشهای مقایسه میان هیستوگرام[۵] ها.در روش مقایسه‌ای پیکسلی تفاوت میان مقادیر شدت و یا رنگ و یا پیکسلها از دو تصویر متوالی با یکدیگر مقایسه می‌شود. به عنوان مثال می‌توان از قدر مطلق مجموع مقادیر تفاوت‌های پیکسلی استفاده نمود. مثالاً از فرمول زیر می‌توان برای ویدئوهای سیاه و سفید[۶] استفاده نمود.

$$D(i,i+1)=\frac{\sum_{x=1}^{x}\sum_{y=1}^{y}\left|p_i(x,y)-p_{i+1}(x,y)\right|}{x.y}$$

فرمول ۱. یک فرمول بر پایه تفاضل انباشته شده برای تشخیص Shot های ویدئویی برای تصاویر سیاه و سفید

---

[1] Compressed

[2] Cumulative difference

[3] pixel Based

[4] Block Based

[5] Histogram comparison

[6] Gray Level

در این صورت یک کات در صورتی تشخیص داده می‌شود که این مقدار مجموع از یک حد آستانه از پیش تعریف شده T بیشتر باشد. ایراد این کار این است که به اشیا و حرکت‌های دوربین حساس است و نمی‌تواند تفاوتی میان تغییرات بزرگ در یک فضای کوچک و یا تغییرات کوچک در یک فضای بزرگ شود. در روش مقایسه مبتنی بر بلاکها هر فریم i به b بلاک تقسیم بندی می‌شود که با بلاک‌های متناظر خود در فریم i+1 مقایسه می‌شود. بلاک‌های متناظر به کمک یک میزان شباهت[1] λ مقایسه می‌شوند که تابعی از وایانس و مقدار میانگین است. سپس تعداد بلاکهایی که در آنها مقدار λ از یک حد آستانه T بیشتر است محاسبه می‌شود در صورتی که تعداد بلاکهای تغییر یافته به اندازه کافی بزرگ باشد یک کات تشخیص داده می‌شود. در این زمینه مقاله‌های زیادی وجود دارد همچون الگوریتم net که توسط Koprinska & carrato در [۲۲] ارائه شده است.

در روش مبتنی بر هیستوگرام ایده اصلی بر این است که در دو فریم با زمینه ثابت و اشیا بدون تغییر (هرچند دارای حرکت) تغییرات کوچکی در هسیتوگرامهایشان وجود خواهد داشت. مزیت این روش این است که نسبت به چرخش تصاویر و تغییرات آرام ناشی از زاویه دید و مقیاس حساسیت ندارد. اما مشکل آن در این است که دو تصویر با هستوگرامهای یکسان ممکن است دارای محتویاتی کاملاً متفاوت باشند. هرچند که احتمال چنین اتفاقی به اندازه کافی پایین است. علاوه بر این روشهای حل این شکل در مقاله‌های متعددی تشریح شده است.

اما کار در حوزه ویدئوهای فشرده شده دارای مزیت های بیشتری نسبت به حوزه غیر فشرده است. آنچه تا بدینجا گفته شد، مربوط به کار بر روی تصاویر فشرده سازی نشده بود. اما عموماً تصاویر بصورت فشرده شده ارائه می‌شوند. در صورتی که کار بر روی تصاویر فشرده شده مستقیماً به صورت گیرد مزایای زیر

---

[1] Likelihood Ratio

حاصل می‌شود. اول این که هزینه های محاسباتی ناشی از فشرده سازی و غیر فشرده سازی کردن حذف می‌شود که خود از پیچیدگی محاسباتی کاسته و سبب صرفه‌جویی در زمان و فضا می‌شود. از آنجا که در حوزه ویدئوهای فشرده سازی شده با میزان نرخ داده کمتری سروکار داریم عملیات سریعتر صورت خواهد گرفت. همچنین یک ویدئوی از پیش فشرده سازی شده حاوی اطلاعات مطلوبی همچون بردار حرکت اشیا و ... است که می‌تواند در قطعه‌بندی‌های زمانی ویدئو مناسب واقع شوند.

قطعه‌بندی زمانی ویدئوها یکی از بحثهای داغ و زمینه‌های فعال تحقیقاتی در حال حاضر است. کارهای قدیمی‌تر اکثراً بر روی تشخیص برش ها متمرکز شده بودند در حالی که تکنیکهای جدیدتر اکثراً با تشخیص تغییرات تدریجی سروکار دارند. اکثر الگوریتمهایی فعلی تنها قادر به تشخیص حرکتهای تدریجی کوتاه هستند و قادر به تشخیص انواع حرکت‌های تدریجی متفاوت نیستند. کارهای آینده در این زمینه بر روی افزایش توانایی در تشخیص dissolve میان رشته‌هایی با حرکت‌های تند و سریع و افزایش توانایی در تفاوت قائل شدن میان حرکت‌های تدریجی از حرکت‌های اشیا است. بطور خلاصه روشهایی که می‌تواند در این راه کمک نماید شامل استفاده از اطلاعات اضافی همچون صدا و متون، استفاده تواماً از تکنیکهای قطعه‌بندی زمانی و گسترش متدهایی که توانایی یادگیری از تجربه‌های تنظیم پارامترها را دارند. در عین حال وجود یک محک (یک پایگاه داده + ضوابط ارزیابی یکسان) برای مقایسه اسانتر روشها مورد نیاز است.

۴,۱ خلاصه سازی ویدئو

خلاصه های ویدئویی بر روی پیدا کردن یک مجموعه کوچکتر از تصاویر برای ارائه از محتویات تصویری تمرکز نموده است و غالبا فریمهای کلیدی را به کاربر ارائه می‌نماید. خلاصه های ویدئویی غالبا خلاصه‌ای

از تصاویر ثابت[1] به کاربر ارائه می نماید. مجموعه‌ای از تصاویر ثابت با موضوع برجسته یا فریمهای کلیدی هستند که از ویدئوی متضمن خود ساخته شده‌اند. بیشتر تحقیقات در زمینه‌های خلاصه سازی شامل استخراج فریمهای کلیدی و گسترش یک واسط browser مبتنی بر آن که بهترین ارائه دهنده از ویدئو باشد،می‌باشد.

استفاده از تصاویر ثابت برای ارائه مزیت‌هایی را به همراه خود می‌آورد. خلاصه‌های تصاویر ثابت می‌توانند بسیار سریعتر از خلاصه‌هایی با تصاویر متحرک ساخته شوند چرا که هیچ دستکاری در صدا و یا تصویر پ الزامی نیست. ترتیب زمانی فریمهای ارائه شده می‌تواندمشاهده شود در نتیجه کاربر می‌تواند مفاهیم[2] آن را سریعتر بفهمد. علاوه بر این تصاویر ثابت استخراج شده را می‌توان در صورت نیاز چاپ نمود.

انتخاب دیگری که در مقابل تصاویر ثابت[3] قرار می‌گیرد ارائه چکیده ویدئو[4] است. به طور کلی چکیده ویدئو ها، ویدئو کلیپ‌های کوتاه متشکل از مجموعه‌ای از رشته‌های تصاویر و صدای متناظر با آنها است که از یک ویدئوی اصلی طولانی‌تر استخراج شده‌اند. غالبا چکیده ویدئو ها یک خلاصه چند رسانه‌ای ارائه می دهند که بیش‌تر نمایش داده می‌شوند تا این که به صورت ثابت نگاه شوند. هدف آنها بیشتر ارائه رشته ویدئوی اصلی در مقدار زمان بسیار کوتاه‌تر است . دو روش اصلی برای چکیده کردن ویدئو[5] وجود دارد. یکی رشته‌های خلاصه[6] که برای آشنا نمودن کاربر با کلیات یک رشته ویدئویی به کار می‌رود. دومی

---

[1] static storyboard

[2] Concept

[3] Still images

[4] Video skims

[5] Video skimming

[6] Summary sequences

قسمتهای پررنگ[1] است که تنها حاوی جذابترین قسمتهای یک رشته ویدئویی است . انتخابقسمتهای پررنگ از یک ویدئو یک فرایند ذهنی است و در نتیجه بیشتر فعالیتهای استخراج چکیده های ویدئویی بر روی تولید رشته‌های خلاصه تمرکز یافته‌اند.

یکی از مهمترین جنبه‌های تلخیص ویدئو گسترش رابطهای کاربری است که بتوانند به بهترین نحو رشته ویدئوی اصلی را ارائه نمایند. میان سطحی از تجرید که در خلاصه یک ویدئو به کاربر ارائه می‌شود و میزان فهمی که کاربر از آن دارد یک نوع توازن[2] وجود دارد به این معنی که هرچه خلاصه ما متراکم‌تر و با بار معنایی بیشتر باشد ، برای یک کاربر که دارای یک زمینه ذهنی است امکان browsing بهتر وسریعتر وجود دارد اما ممکن است اطلاعاتی که این خلاصه از کل ویدئو نشان می‌دهد به اندازه کافی جامع نباشد و نتواند فهمی کامل از ویدئو را فراهم آورد. یک خلاصه که به صورت دقیقتر ارائه شده است، ممکن است بتواند به کاربر اطلاعات بیشتری را برای فهم ویدئو منتقل نماید اما ممکن است زمان بیشتری برای brows نمودن آن نیاز باشد.

بطور کلی کلاسهای متفاوت از کاربران نیازهای متفاوتی دارند. یکی از مهمترین سوال‌هایی که در سیستم های تلخیص ویدئو مطرح می‌شود این است که چگونه shot هایی را انتخاب کنیم که بهترین نماینده هستند؟ که جواب قطعی برای این سوال وجود ندارد علاوه بر این، این سؤال مطرح می‌شود که چه چیزهای دیگر بایستی در خلاصه شامل شود مثلاً متن، صدا و .... تحقیقات مهمی که در این زمینه در حال ظهور هستند شامل قطعه‌بندی و خلاصه بندی تطبیق پذیر و همچنین تهیه خلاصه برای تحویل دادن به کاربران mobileاست. [۲۳]

---

[1] High lights

[2] Trade off

۵ شاخص بندی، پرس و جو و بازیابی داده های ویدئویی

در مقایسه با بانک‌های داده سنتی یا همان بانک داده های متنی، شاخص نمودن ویدئو بسیار مشکل تر و پیچیده‌تر است. چرا که در پایگاه داده‌های سنتی عموماً داده‌ها بر اساس یک کلید اصلی ایندکس می‌شوند و ایندکس نمودن به صورت عملی واضح و روشن صورت می‌گیرد. اما برخلاف داده‌های متنی، تولید ساختن ایندکس به صورت اتوماتیک از محتویات یک ویدئو کاری بسیار سخت و مشکل است . به منظور شاخص نمودن یک ویدئو سه قدم می‌بایستی صورت گیرد:

۱- قطعه بندی[۱]: قطعه‌بندی زمانی از محتویات ویدئو به واحدهای کوچکتر

۲- تجرید[۲]: استخراج و یا ساخت زیر مجموعه نمایش داده‌های ویدئویی از ویدئوی اصلی

۳- آنالیز محتویات[۳]: استخراج ویژگیهای بصری از فریمهای ویدئوی نماینده

کارهای موجود بر روی شاخص‌بندی ویدئوها در حال حاضر می‌تواند به سه گروه تقسیم‌بندی شود:

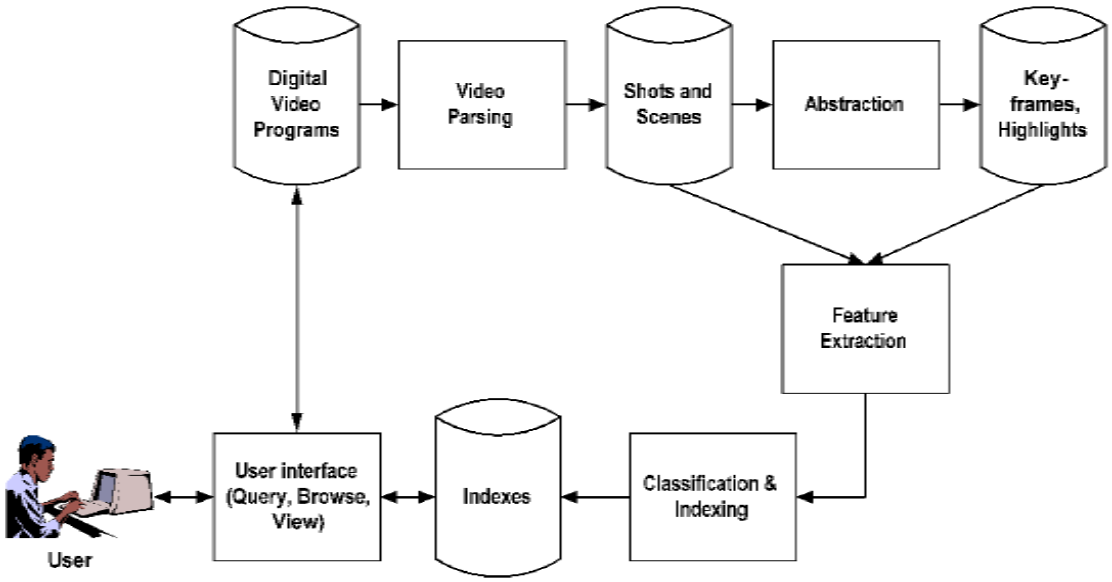
- شاخص بندی مبتنی بر حاشیه نویسی
- شاخص بندی مبتنی بر ویژگیها
- شاخص بندی برای زمینه کاری مشخص

---

[1] Video Parsing

[2] abstraction

[3] Content analysis



شکل ۳. شمای یک سیستم ایندکس گذاری برای بانک داده ویدئویی

بیشتر روشهای شاخص بندی[۱] نیاز دارند تا جریان ویدئو به مناظر قطعه‌بندی شوند پیش از آن که شاخص بندی صورت گیرد. در شاخص گذاری مبتنی بر حاشیه نویسی[۲] ، حاشیه نویسی غالباً یک فرآیند دستی است که به وسیله کاربران ورزیده و با تجربه صورت می‌گیرد، و غالباً با مشکلاتی همچون زمان، هزینه، ابهام و ... است. روش متعارف و تکنیک کنونی در حال حاضر اختصاص دادن یک کلمه کلیدی[۳] به قطعات ویدئویی یا منظره هاست. تکنیکهای حاشیه نویسی مبتنی بر حاشیه نویسی برای شاخص بندی مرتبط با انتخاب کلمات کلیدی، ساختمان داده و واسطها جهت تسهیل کار کاربران هستند. حتی به

---

indexing [1]

Annotation-based indexing [2]

keyword [3]

وسیله کمکهای اضافی، روش حاشیه نویسی مبتنی بر کلمات کلیدی ضعیف است. علت امر در این است که کلمات کلیدی نمی‌توانند روابط زمانی و فضایی موجود را نمایش دهند. علاوه بر این کلمات کلیدی نمی‌توانند به طور کامل ارائه دهنده اطلاعات معنایی باشند و علاوه بر این توارث ، همانندی و شباهت و یا استنتاج میان توصیف‌گرها[1] را پشتیبانی نمی‌نمایند. کلمات کلیدی نمی توانند روابط میان توضیحات را شرح دهند و بدتر از همه اینکه کلمات کلیدی مقیاس نمی‌گذارند.

پیشنهادهای دیگری که در مقابل روش حاشیه نویسی مبتنی بر کلمات کلیدی قرار می‌گیرد شامل:

- حاشیه نویسی چند لایه و با استفاده از شکلک ها
- حاشیه نویسی قطعه بندی شده[2] (ساختار حاشیه نویسی سلسله مراتبی بر بالای یک جریان ویدئویی فیزیکی)

استفاده از منطق زمانی _ مکانی[3] [۲۴]

شق دیگر در مقابل روشهای مبتنی بر حاشیه نویسی روشهای مبتنی بر ویژگیها است. در روشهای شاخص‌بند مبتنی بر ویژگی ها [4] هدف نهایی توانایی در شاخص‌گذاری تماماً خودکار از یک برنامه ویدئویی بر اساس محتوای آن است. بر پایه تکنیکهای پردازش تصویر ویژگیهای کلیدی تصویر همانند رنگ،زمینه، حرکت اشیاء و ... را از اویدئو استخراج و از این ویژگیها برای ساختن ایندکس‌ها استفاده شود. بزرگترین

---

[1] descriptor

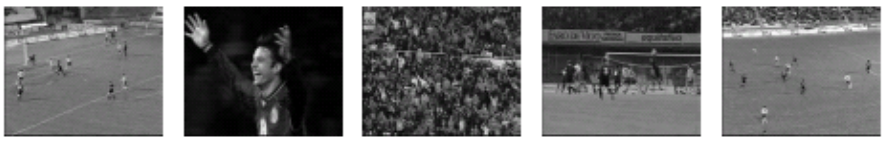
[2] Segmented annotation

[3] Spatial Temporal Logic

[4] Feature-Based

مشکل در این روش شکاف معنایی[1] است. به صورت گسترده‌ای در طی ۱۰ سال گذشته بر روی این روش تحقیقات گرفته است.

علاوه بر آنچه که گفته شد، در مطالعه سیستم های ایندکس گذاری، با توجه به مستقل و یا وابسته بودن این سیستم ها به حوزه کاری خاص، یک دسته بندی صورت میگیرد. هدف اصلی رسیدن به یک سیستم شاخص بندی مستقل از دامنه کاربردی خاص است. روشهای وابسته به دامنه کاری خاص برای انجام پردازش‌های بیشتر بر روی استخراج ویژگی‌های سطح پایین ویدئو و آنالیز نمودن نتایج بدست آمده غالبا از مدلهای ساختاری ویدئویی سطح بالا (منطقی) به عنوان دانش ابتدایی استفاده می کنند. . مثالهایی از این روش پردازش و خلاصه‌سازی برنامه‌های ورزشی است.



شکل ۵. چگونگی استفاده از دانش حوزه ای خاص برای پردازش یک ویدئو. در مثال بالا با توجه به ساختار خاص یک ویدئو فوتبال و توالی معمول تصاویر می توان ویدئو های مختلف مربوط به بازی فوتبال را پردازش و آنها را حاشیه نویسی نمود

[1] Semantic gap

۵,۱ پروسه بازیابی داده‌های ویدئویی

با توجه به آنچه که گفته شد عموما پروسه بازیابی داده های ویدئویی شامل ۴ مرحله است.

- کاربر یک پرس و جو را به کمک واسطهای کاربری گرافیکی مشخص می‌کند.
- پرس و جو پردازش و ارزیابی می‌شود.
- مقدار یا ویژگی بدست آمده برای تطابق و بازیابی داده ویدئویی ذخیره شده در بانک داده های ویدئویی استفاده می‌شود.
- داده ویدئویی بدست آمد (منتج شده) برای Browsing ، تماشا و در صورت نیاز برای پرس و جوهای دقیق‌تر در صفحه نمایش برای کاربر پخش می‌شود.

انواع مختلفی از پرس و جوها برای بازیابی داده های ویدئویی وجود دارد. پرس و جو ها را می توان از دیدگاه های متفاوتی دسته بندی نمود. دسته بندی پرس و جو ها برحسب محتوای پرس و جو، به پرس و جوهای اطلاعات معنایی (سخت‌ترین)، پرس و جو بر روی فرا اطلاعات (شبیه آنچه که در بانک های داده رایج صورت می‌گیرد) و پرس و جوهای صوتی تصویری مبتنی بر ویژگیهای سطح پایین ویدئو که شامل ویژگیهای فضایی، زمانی و یا فضایی- زمانی هستند تقسیم بندی می شوند.

همچنین پرس و جو ها را می توان برحسب نوع تطابق‌ها[1] دسته بندی نمود. پرس و جو های مبتنی بر تطابق دقیق و پرس و جوهای مبتنی بر تطابق میزان شباهت دو دسته اصلی هستند. انواع پرس و جو برحسب میزان دانه دانگی بودن یا اندازه مورد انتظار از نتیجه پرس و جو عبارت است: پرس و جوهای مبتنی بر فریم ، مبتنی بر برنامه ویدئویی[2] و مبتنی بر جریان ویدئو[3].

از طرف دیگر پرس و جو ها برحسب رفتار و وضعیت نیز به دسته‌های زیر تقسیم‌بندی می‌شوند. پرس و جو های معین[4] که کاربر دارای ایده و نظری روشن و واضح است از آنچه که به عنوان کاربر انتظار دارد. پرس و جو های Browsing که کاربر ممکن است در مورد نیازهای بازیابی خود شک داشته باشد و یا با ساختارها و انواع اطلاعات موجود در سیستم های مدیریت بانک داده ویدئویی ناآشنا باشد.

علاوه بر این انواع پرس و جو ها را می توان بر اساس خصوصیات آنها به دسته های زیر گروهبندی نمود:

- پرس و جو های مستقیم[5]
- پرس و جو به کمک مثال [6]
- پرس و جوهای تکرار شونده[7]

---

[1] matching type

[2] Clip

[3] Video Stream

[4] Deterministic

[5] Direct Query

[6] Query By Example

[7] Iterative Query

پرس و جوهای بانک های داده ویدئویی می توانند به وسیله گسترش SQL[1] برای داده‌های ویدئویی مشخص و تعریف شوند همانندِ TSQL2یا STL و یا Video SQL و غیره . اما عموماً پرس و جو ها به وسیله مثال[2] یا پرس و جوهایی با ظرح کلی از پیش تعریف شده[3] و روشهای در تقابل با کاربر[4] همانند Browsing، مشاهده به همراه بازخوردهایی[5] از طرف کاربر بیشتر از پرس و جوهای شبیه به SQL استفاده می‌شود.

پردازش پرس و جو ها عموما شامل رویه زیر است:

- پیمایش و قطعه بندی پرس و جو[6]: شرایط موجود در پرس و جو یا درخواستهای موجود در آن غالباً به واحدهای پایه شکسته می‌شوند و سپس ارزیابی می‌شوند.
- ارزیابی پرس و جو[7]: از ویژگیهای تصویری سطح پایین از پیش استخراج شده از ویدئو استفاده می‌نماید.
- جستجو در شاخص های بانک داده ویدئویی.

---

[1] Standard Query Language

[2] Query By Example

[3] Query By Sketch

[4] Interactive

[5] Feedback

[6] Query Parsing

[7] Query Evaluation

- بازگرداندن نتایج بدست آمده : در صورتی که درخواست های[1] موجود در پرس و جو برآورده شوند[2] و یا میزان اندازه شباهت ماکزیمم شود داده‌های ویدئویی بازیابی می‌شوند.

نظریات مربوط به طراحی واسط های کاربری گرافیکی، Browsing و Viewing

واسط کاربر نقش حیاتی و حساس در کارایی کلی یک سیستم مدیریت بانک داده های ویدئویی بازی می‌کند. کلیه واسطها می بایستی گرافیکی باشند و به صورت ایده آل بایستی پرس و جو کردن، browsing of summaries of result ، viewing هر یک از نتیجه‌های بدست آمده را به همراه فراهم آوردن بازخورد های مرتبط و یا تصفیه اطلاعاتی پرس و جوها در سیستم را فراهم آورد.

ابزارهای browsing می‌توانند در دو کلاس اصلی قرار گیرند.:

- مشاهده فریم های ویدئو و یا شکلک[3] ها به شکل خط-زمان[4]: واحدهای ویدئو در یک ترتیب زمانی مبتنی بر وقوع آنها مرتب و سازماندهی می‌شوند. (شکل ۶)

- به شکل سلسله مراتبی[5] و تابلو داستان[6] مبتنی بر گراف : ارائه‌ای مبتنی بر گراف که تلاش می‌کند ساختار ویدئو را به صورت خلاصه و روشی موجز بیان کند. (شکل ۷)

---

[1] Assertion

[2] Satisfy

[3] icon

[4] Time-Line

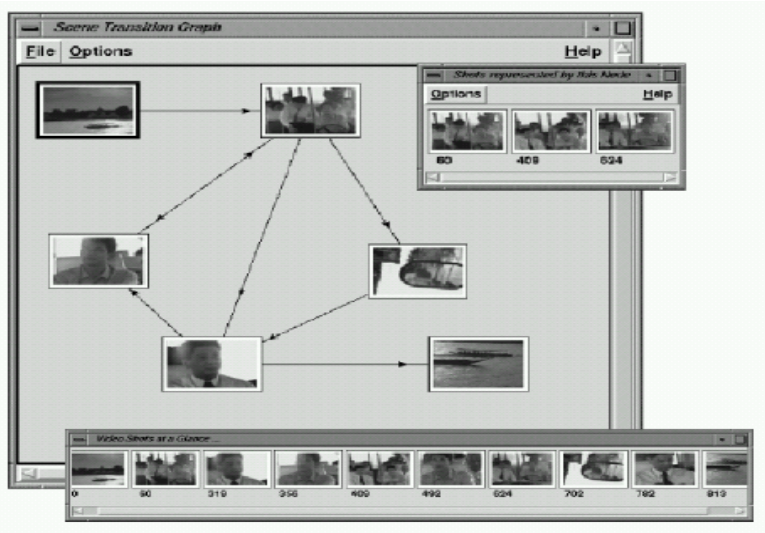
[5] Hierarchical

[6] story board

به طور کلی چگونگی طراحی واسط های کاربری یکی از مسائل و زمینه‌های بسیار فعال تحقیقاتی در زمینه پایگاه های داده ویدئویی است. مثالی از روش خط-زمان پروژه [۲۵] است. مثال از روش سلسله مراتبی شامل STG [1]است که توسط (1997-yeung , yeo) در [۲۶] ارائه شده است.

شکل ۶. نمونه ای از ارائه خط-زمان

[1] Scene Transition graph



شکل ۷. نمونه ای از ارائه مبتنی بر گراف

۶ استانداردهای ویدئویی و نقش آنها در پایگاه داده‌های ویدئویی

اتخاذ و قبول استانداردهای فشرده‌سازی و رمزنگاری[1] ویدئو سبب پیشرفت در این زمینه شده است. الگوریتمهای فشرده‌سازی ویدئو برای دامنه وسیعی از برنامه‌های کاربردی برای فشرده‌سازی ویدئو بر کار می‌روند. کارایی روشهای مدرن فشرده‌سازی همچون MPEG1 ,2 و یا 4، H.261 و H-263 بسیار گیرا و مطلوب بوده است. داده خام ویدئویی از ۱۵ تا ۸۰ برابر بدون از دست رفتن میزان قابل توجهی داده درهنگام ساخت دوباره آن فشرده می شوند. جدول ۲، یک خلاصه از استانداردهای فشرده سازی با توجه به نرخ داده در آنها نمایش می دهد.

---

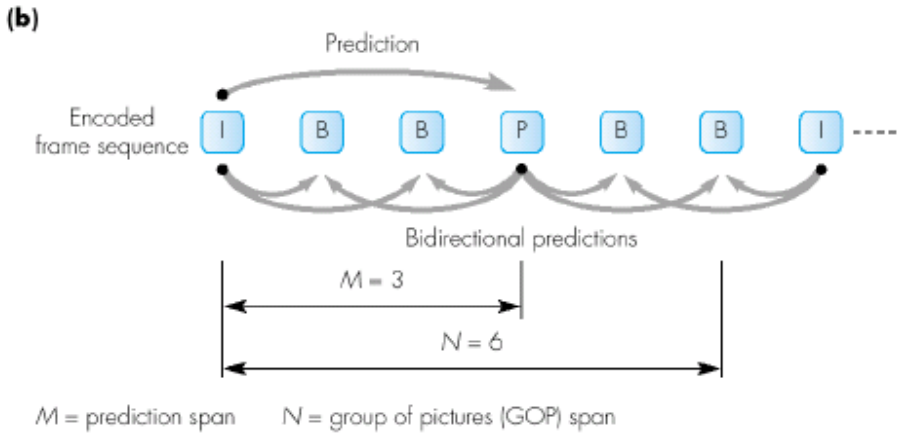
[1] Encoding

- Very-low-bit-rate
  - Frame size: 144 x 176 (QCIF)
  - Frame rate: 5 – 15 fps
  - Target bit rates: 4.8 – 64 Kbps
- Medium-bit-rate
  - Frame size: 288 x 352 (CIF) to 576 x 720
  - Frame rate: 25 – 30 fps
  - Target bit rates: 200 Kbps – 1.5 Mbps
- Professional and high-end applications
  - Frame size: 576 x 720 and larger
  - Frame rate: 25 – 30 fps and larger
  - Target bit rates: 1.5 – 35 Mbps

جدول ۱. انواع استانداردهای ارائه شده برای داده های ویدئویی

استاندارد گروه خبرگان تصاویر متحرک یا MPEG [1]معتبرترین و مورد قبول‌ترین استاندارد جهانی برای فشرده‌سازی ویدئوهای دیجیتال است. این استاندارد از دو تکنیک اصلی برای فشرده سازی زمانی و فضایی استفاده می‌کند. جریان MPEG از سه نوع تصویر I، P و B تشکیل شده است که اینها در یک الگوری تکرار شونده که به نام گروه تصاویر یا (GOP) خوانده می‌شود ترکیب می‌شوند. فریمهای Intra(I) نقاط دسترسی تصادفی به داده‌های فشرده شده را فراهم می‌آورند و تنها به وسیله اطلاعات ارایه شده در تصویر کد شده‌اند. فریمهای P(Predicted) با استفاده از نزدیکترین تصویر مبداء (reference) قبلی (IOSP) با forward notion compensated کد شده‌اند. تصاویر B یا Bi-directional همچنین motion compensated هستند، این بار با توجه به هر دو فریم مبداء قبلی و بعدی.

[1] Moving picture Expert Group



شکل ۸. روابط موجود هنگام فشرده سازی تصاویر در استاندارد MPEG

در هنگام پروسه رمزنگاری[1] آزمایشی بر روی هر بلاک متحرک از فریمهای B, P انجام می‌شود تا ببینم که آیا هزینه Motion Compensation بیشتر است یا Intra Coding. به عنوان نتیجه هر بلاک متحرک از یک Pفریم می‌تواند هر دو Intra و Forward کد شود در حالی که برای هر بلاک متحرک از یک فریم B چهار امکان وجود دارد: intra، Forward ، Back Ward و یا Interpolated . تا به حال چندین نسخه از استاندارد MPEG منتشر شده است.MPEG1 هنوز به طور وسیعی در ویدئو برای PC ها استفاده می‌شود. استاندارد MPEG2 کیفیتی شبیه DVD و به طور وسیع در مصرف کننده‌های الکتریکی استفاده می‌شود. استانداردMPEG4 با عنوان استاندارد کد نمودن محتویات ویدئو[2] پا به عرصه ظهور نهاده است. این استاندارد مشکلاتی در پیداکردن استفاده وسیع دارد، بخصوص به دلیل حمایت از خصوصیات ذهنی و نیاز به گسترش شماهایی جهت قطعه‌بندی کارا و اتوماتیک. MPEG7 یک واسط تشریح محتویات چند

---

[1] encoding

[2] Content Based Video Coding

رسانه‌ای را مشخص می‌کنند علاوه بر این MPEG7 یک استاندارد از فراداده‌های چند رسانه‌ای در غالب XMLارائه می‌نماید[۳]. استاندارد جدیدتر MPEG7 محتویات چند رسانه‌ای را در تعدادی سطوح مختلف شرح می‌دهد شامل ویژگیها،ساختار ، مفاهیم مدلها. هدف MPEG 7 فراهم آوردن یک سیستم فراداده ای کار کننده در خود را ارائه می دهد تا اجازه ایندکس نمودن سریع و کارا ، جستجو کردن و فیلتر نمودن داده های چند رسانه ای بر اساس محتویاتشان فراهم شود اما متاسفانه روشی برای چگونگی استخراج این مفاهیم ارائه نمی دهد.

۷ معنای ویدئو[1]

منظور از معنا[۲] مفاهیم سطح بالا همچون اشیا و اتفاقات در یک داده ویدئویی است. محتویات معنایی یکسان می توانند توسط ارائه های متفاوت تصویری نمایش داده شوند که از مجموعه داده های خام ویدئویی مختلف تشکیل شده اند. معنا شرح اطلاعاتی است که ویدئو در بر دارد و از این طریق به کاربر اجازه می دهد تا داده های ویدئویی را بر حسب محتوای مفهومی آن بازیابی نماید.[۲۷] دسته بندی متدوال از معنا ویدئو ، بر پایه سه گروه شکل می گیرد:

معنای مبتنی بر ویژگیهای سطح پایین محتویات[۳]: المانهایی که می توانند بصورت خودکار از جریان ویدئو استخراج شوند بدون در نظر گرفتن دانش خاص در رابطه با یک زمینه خاص.

معنای مبتنی بر ساختار[۱]: ویدئو به عنوان مسندی با ساختار سلسله مراتبی در نظر گرفته می شود.

---

[1] Video Semantic

[2] Semantic

[3] Low Level Content Based Semantics

معنای مبتنی بر حاشیه نویسی ها[۲]: مفاهیم از طریق حاشیه نویسی داده های ویدئویی نمایش داده می شود.

مشکل اصلی در این حوزه، شکاف مفهومی[۳] است. شکاف مفهومی، فاصله یا شکاف میان اطلاعاتی است که می تواند بصورت اتوماتیک از داده های تصویری استخراج شود و ، تفسیر هایی که همان داده ها می توانند برای یک کاربر در یک موقعیت مشخص داشته باشند. توجه به این نکته لازم است که منظور و مفهوم[۴]، یک داده (Datum) نیست که در تصاویر یا ویدئو ارائه شده باشد و بشود از همان ابتدا آن را محاسبه و Decode نمود. در سالهای اخیر تلاشهای زیادی جهت برقراری ارتباط میان ویژگی های سطح پایین و مفاهیم سطح بالا و یا به عبارتی پر کردن شکاف مفهومی صورت گرفته است. به تازگی سعی شده است جهت غلبه بر مشکل شکاف مفهومی تعریف جدیدی از مفهوم و نقش معنی در یک سیستم اطلاعات تصویری ارائه شود.نظریه Emergent Semantic می گوید که اجازه دهید تا معنی ها بوسیله تقابل میان کاربر و ماشین گسترش یابد. [۲۸]

۷,۱ معنای مبتنی بر حاشیه نویسی و مدل نمودن معانی

حاشیه نویسی ها ارائه دهنده هرگونه توصیف سمبلیک از یک ویدئو یا قسمتی از آن است. این حاشیه نویسی ممکن است بصورت دستی فراهم شود. در حال حاضرحاشیه نویسی تنها در صورت فراهم آوردن

---

[1] Structured Based Semantics

[2] Annotation Based Semantics

[3] Semantic Gap

[4] Semantic

دانش زمینه ای کافی در یک محدوده خاص می تواند بصورت خودکار فراهم شود. نحوه ارائه معنای ویدئو نقش کلیدی در آنالیز و بازیابی داده های ویدئویی دارد. از آنجا که فهم ماشین از داده های ویدئویی هنوز یکی از مسائل تحقیقاتی حل نشده است، عموما از حاشیه نویسی برای تشریح محتویات ویدئو استفاده می شود.

مدل نمودن معنی به مراتب مشکل تر از مدل نمودن ساختار و یا ویژگیهای تصویری سطح پایین ویدئو است چرا که برای ارائه آن نیاز به دانش زمینه ای و یا تقابل با کاربر و یا هر دو آنهاست. در سطح فیزیکی ویدئو یک جریان زمانی از پیکسل های متوالی است که هیچ رابطه مستقیمی با محتویات معنایی خود ندارد. این مشکل هنگامی پیچیده تر می شود که مفاهیم معنائی دیگر همچون استعاره ها، معانی مخفی، تشبیهات و ... را در نظر بگیریم.

ساده ترین روش جهت مدل نمودن مفاهیم موجود در محتویات ویدئو استفاده از متن های آزاد جهت حاشیه نویسی دستی است. این همان چیزی است که از آن با نام حاشیه نویسی دستی متون آزاد یاد[1] می شود. در این روش ایده اصلی فراهم آوردن یک لایه بندی مفهومی در بالای جریان داده ویدئویی است. این کار به دو روش کلی صورت می گیرد. چینه بندی[2] و حاشیه نویسی کلمات کلیدی/خصیصه ها از قطعات ثابت[3].

روش کار در حاشیه نویسی کلمات کلیدی/خصیصه ها از قطعات ثابت به این شکل است که ویدئو به قسمت هایی ثابت(معمولا منظره ها) قطعه بندی می شود .مفاهیم این قطعات بصورت مستقل از هم،

---

[1] Free Text Manual Annotation

[2] the stratification

[3] The keyword/attribute annotation of fixed segments

توسط متون آزاد و حاشیه نویسی خصیصه/ واژه کلیدی تشریح می شود. ایرادات این روش عدم وجود انعطاف پذیری و قطعه بندی ثابت است که یک و تنها یک قطعه بندی از داده اصلی ارائه می نماید. همچنین قسمت قطعه بندی شده از Context خود جدا می شود و در نتیجه اطلاعات زمینه ای مهمی در رابطه با آن از دست می رود.

در مقابل روش گفته شده روش انعطاف پذیر تری قرار دارد که ویدئو را برحسب ضوابط مفهومی آن به قطعاتی منطقی تقسیم بندی می نمایند. این همان روش چینه بندی است. در این روش تشریحگرهای متنی چینه[1] نامیده می شوند که ممکن است در هر بخش از ویدئو دارای همپوشانی باشند. در سیستم های جدید تر چینه می تواند یک لیست از تصاویر ثابت باشد. می توان یکسری از عملگرهای جبری برای کار برروی چینه فراهم نمود. به عنوان مثال می توان از جبر زمانی آلن [۲۹] بر روی آنها استفاده نمود.

روش هایی برای بهبود هریک از این روش ها پیشنهاد میشود. به عنوان مثال برای حل مشکل تعداد کلمات کلیدی محدود و یا خصیصه های از پیش تعریف شده استفاده از روشهای پیشنهاد میشود. در این روش کاربر خصیصه های مورد نظر خود را بر حسب نیاز تعریف می کند. مثال آن را می توان در سیستم OVID دید. علاوه بر این استفاده از مکانیزمهای ارث بری سبب تسهیل کار می شود. به عنوان مثال OVID مکانیسم ارث بری بازه ای متداخل[2] را معرفی نموده است که هر قطعه ای از ویدئو می تواند خصوصیات قطعه ای دیگر را به ارث برد.

همانطور که گفته شد یکی دیگر از تکنیک ها برای افزایش سرعت و افزایش کارایی، استفاده از Strataهای تصویری است. به جای استفاده از حاشیه نویسی های متنی کاربر می تواند از یکسری شکلک

---

[1] strata

[2] Interval Inclusion Inheritance

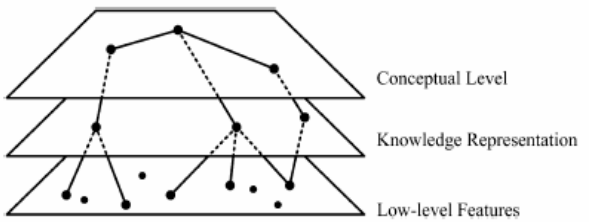
استفاده نماید. این شکلک ها در یک ساختار سلسله مراتبی مرتب می شوند. علاوه بر این امکان تعریف روابط زمانی فضایی وجود دارد. ایده اصلی در اینجا آن است که ادعا می شود ارائه مفاهیم به کمک تصاویر، بهتر و مناسب تر از داده های متنی است چراکه تصاویر توانایی ارائه مفاهیم پیچیده تری هستند.

۷,۲ نمونه های عملی از مدل های معنایی

تا به حال کارهای زیادی برای ارائه مفاهیم در حیطه داده های ویدئویی صورت گرفته است. از جمله آنها می توان از [۳۰] [۳۱] [۳۲] [۳۳] را نام برد. در هریک از سیستم های ذکر شده از روشی متفاوت جهت ارائه مفاهیم استفاده شده است. ارائه مفاهیم برای داده های چند رسانه ای هنوز یکی از زمینه های فعال هوش مصنوعی و مدیریت پایگاه داده های چند رسانه ای است.

در [۳۰] سیستم ارائه شده، رشته ویدئویی را بصورت اتوماتیک با استفاده از دانش مجموعه داده های از قبل حاشیه نویسی شده ، حاشیه نویسی می نماید. این کار توسط روش های مبتنی بر قانون که از تئوری مجموعه های فازی و تکنیکهای داده کاوی بهره می گیرد، صورت می گیرد. رویه کاردر سیستم پیشنهادی به شکل زیر است. ابتدا دانش از مجموعه داده های از قبل حاشیه نویسی شده استخراج می می شود. قوانینی که مفاهیم سطح بالای تعریف شده توسط کاربر در واژگان را به ویژگی های سطح پایین مرتبط می سازد، استنتاج می نماید. به کمک قوانین بدست آمده و استنتاج فازی، ویدئو های جدید را بصورت خودکار حاشیه نویسی می شود. به عبارت دیگر سیستم پیشنهادی به کمک یک مجموعه آزمایشی برای استخراج مفاهیم و حاشیه نویسی ویدئوهای جدید آموزش داده می شود. مشکل سیستم فوق در آن است که پیش از استفاده از آن می بایتسی سیستم را توسط ویدئوهای از پیش حاشیه نویسی شده آموزش داد. این سبب می شود که سیستم حتی پس از آماده سازی یک مجموعه آموزشی تنها برای

یک زمینه خاص قابل کاربرد باشد. به عبارت دیگر سیستم پیشنهادی به شکل غیر مسفقیم از وجود ناظر و همچنین دانش زمینه ای از پیش تعریف شده استفاده می نماید. شکل زیر لایه های متفاوت مفهومی را در این سیستم نمایش می دهد. همانطور که دیده می شود در پایین ترین سطح تصاویر و ویژگیهای سطح پایین تصویر قرار دارند و در بالاترین سطح مفاهیم سطح بالا. فاصله میان این دو سطح که همان شکاف مفهمومی موجود است، سعی شده است توسط ارائه مناسب از دانش و استفاده از تکنیکهای داده کاوی پوشش داده شود.

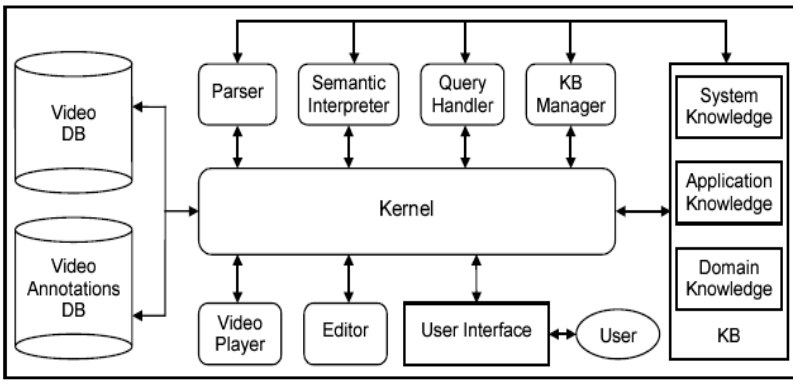


شکل ۹. سطوح مفهومی در نظر گرفته شده در سیستم Dorado

نمونه ای دیگر که در آن به ارائه مفاهیم برای داده های ویدئویی پرداخته شده است، سیستم Smart Videotext است [۳۱]. این سیستم بر پایه مفاهیم قطعات ویدئویی منطقی[1] و حاشیه نویسی متون آزاد

[1] Logical Video segment

عمل می نماید و یک نگاشت دلخواه میان این دو فراهم می آورد. در این سیستم از گرافهای مفهومی[۱] جهت نمایش دانش استفاده می شود. شمای کلی سیستم در شکل زیر مشاهد می شود.



شکل ۹. شمای کلی سیستم Smart Videotext

ذخیره داده های ویدئویی بر عهده بانک داده ویدئویی است. علاوه بر این از یک بانک داده مستقل برای ذخیره حاشیه نویسی های داده های ویدئویی استفده می شود. هر حاشیه نویسی به همراه قطعه ویدئوی منطقی مرجع مربوط به خود ذخیره می شود. پایگاه دانش [۲]بانک دانش سیستم است و در آن دانش مربوط به زمینه های متفاوت ذخیره می شود. پایگاه دانش سیستم[۳] که به شکل بانک قوانین نمود پیدا می کند قوانین استاندارد سیستم را شکل می دهد که مربوط به چگونگی شکل گیری قوانین و دانشی

---

[1] Conceptual Graph

[2] KB

[3] System Knowledge

است که مستقل از زمینه کاربرد هستند. دانش مربوط به کاربرد[1] متشکل از گرافهای مفهومی، انواع مفاهیم، مفاهیم و روابط میان این مفاهیم است که از ویدئو مشتق شده اند. در هر زمان تنها دانش مربوط به یک کاربرد استفاده می شود در حالیکه ممکن است این بانک دانش محتوی دانش کاربردی متفاوت باشد. مراتب مفاهیم به همراه مفاهیم و روابط مفهومی میان آنها در بانک دانش زمینه ای[2] نگهداری می شود. به عبارت دیگر دانش مربوط به یک کاربرد خاص که به صراحت در حاشیه نویسی های ویدئو ذکر نشده است در این بانک دانش ذخیره میشود.

هسته[3] نقس کنترلر دارد و همچنین شامل یک موتور استنتاج بر پایه Prolog است. تجزیه کننده[4] ساخت درخت نحوی جملات مربوط به حاشیه نویسی ها را بر عهده دارد. ترجمه درختهای بدست آمده به گرافهای مفهومی بر عهده تشریحگر معنایی[5] است. علاوه بر آن فراهم نمودن امکان گرفتن پرس و جو از پایگاه داده ویدئویی بر عهده رسیدگی کننده به پرس و جو ها[6] است. پرس و جوها به شکل گرافهای مفهومی بیان می شوند.

کار در زمینه مدل کردن معنا در این حوزه یکی از مسائل جذاب و فعال تحقیقاتی در سالهای اخیر بوده است. کار در این زمینه همچنان ادامه دارد. به تازگی رویکردهایی که در آن به دنبال یک ارائه یک شکل

Application Knowledge[1]

Domain Knowledge[2]

Kernel[3]

Parser [4]

Semantic Interpreter [5]

Query Handler [6]

از دانش است رونق بیشتری داشته است. همچنین استفاده ار هستان شناسی ها به عنوان ابزاری جهت ارائه دانش یکی از زمینه های تحقیقاتی مهم را در پیش روی محققین گشوده است.

۸ نتیجه گیری

در مستند ارائه شده درباره داده های چند رسانه ای و به خصوص داده های ویدئویی به عنوان یک داده چند رسانه ای مهم بحث شد. مشکلاتی که بر سر راه گسترش سیستم های مدیریت بانک داده های ویدئویی وجود دارد به بحث کشیده شد و درباره اهیمت معانی در کارائی این سیستم ها نکاتی به اختصار گفته شد.

ارائه معنی برای داده ها یکی از زمینه های تحقیقاتی با قدمت طولانی در هوش مصنوعی است. با کاربرد و گسترش روزافزون استفاده از داده های چند رسانه ای به خصوص گسترش تکنولوژی در زمینه بانک های داده ویدئویی و ویدئوهای دیجیتال و پدید آمدن حجم عظیمی از این داده ها جهت پردازش و ارائه به کاربر، وجود ابزارهایی جهت پردازش مفهومی و جستجو و بازیابی مبتنی بر معانی به جای دیگر ویژگیهای سطح پایین، از مباحث تحقیقاتی مهم در سالهای اخیر می باشد.

بزرگترین مشکل در این زمینه شکاف معنایی است و کلیه تلاشها برای حل این مشکل متمرکز شده است. هدف نهایی دستیابی به مدلی است که با استخراج خودکار ویژگیهای سطح پایین تصویر به مفاهیم معنایی سطح بالا برسد. این به معنی پوشش شکاف معنایی است. تا به حال هیچ یک از سیستمها چنین ویژگی را ارئه نداده اند.

جهت رسیدن به این هدف، نیاز است تا با توجه به خصیصه های این نوع داده، هماهنگی دقیق میان روشهای ذخیره و بازیابی و مدل کردن معنایی آن در نظر گرفته شود. در حال حاضر پردازش و بازیابی این

داده ها به کمک مفاهیم معنایی محتوی داده ویدئویی با نقاط ضعف فراوانی همراه است. بهبود روشهای ارئه معنی و مدل کردن آن، می تواند راه گشا و کلید حل مسئله در این زمینه باشد. استفاده از ابزارهای ارائه دانش یکپارچه همانند هستان شناسی ها می تواند یک راه حل مناسب برای بازنمایی دانش معنایی باشد. علاوه بر اینکه در استفاده از هر ابزار ارائه دانش نیاز است تا به ویژگیهای مختص داده های ویدئویی، یعنی وابستگی زمانی و فضایی توجهی خاص مبذول گردد.

همانطورکه گفته شد، از آنجا که مسئله بینایی ماشین و درک تصویر از مسائل حل نشده در زمینه هوش ماشین است، استفاده از حاشیه نویسی های متنی به عنوان آخرین دستاورد جهت بازنمایی معنایی مورد توجه قرار گرفته است. اما استفاده از این روش نیز همراه با مشکلاتی از جمله محدود بودن معنی ارئه شده توسط متن به جای تصاویر، هزینه بر بودن حاشیه نویسی و ... است. اخیرا سیستم هایی جهت حاشیه نویسی خودکار از داده های ویدئویی معرفی شده اند اما در همگی آنها نیز وجود ناظر چه بصورت مستقیم و یا غیر مستقیم دیده می شود. مشکل دیگری که در این زمینه وجود دارد، وابستگی این سیستم ها به زمینه کاربری آنهاست. ارائه یک چهارچوب کلی برای ارائه دانش زمینه ای که سیستم را برای کاربرد در زمینه های مختلف راهنمایی نماید به عنوان یک راه حل می تواند به کمک طلبیده شود.

مشکل دیگری که در این زمینه به چالش طلبیده می شود نبود و یا کمبود یک محک[1] استاندارد و مناسب برای مقایسه بین این سیستم ها است. اخیرا تلاش هایی برای حل این مشکل از طرف TREC صورت گرفته است.

[1] Benchmark

۹ فهرست منابع و ماخذ

[1] تهذیبی، ب. روشهای شاخص‌گذاری در بانک‌های تصویری، سیمنار کارشناسی ارشد ، دانشگاه علم و صنعت ، دانشکده کامپیوتر، ۱۳۷۹.

[2] Sundaram, H. Chang S. F. Video analysis and summarization at structural and semantic levels, Columbia University publishing, USA.

[3] MPEG-7 Committee, Overview of the MPEG-7 Standard, Report ISO/IEC JTC1/SC29/WG11 N4509, J. Martinez Editor, 2001.

[4] Mulhem , P. Gensel , J . Martin, H. Adaptive video summarization, IPAL-CNRS.

[5] Kokkoras, F. Jiang, h. Vlahavas, I. Aref, W.G. , Smart VideoText: a video model based on conceptual graphs , Multimedia systems , Springer-Verlag, 2002.

[6] Hampapur, A. Semantic Video Indexing: Approach and Issues , Internal report , IBM TJ Watson Research Center.

[7] Merialdo, K. Lee, T. Automatic construction of personalized TV news programming, Proceedings of the seventh ACM international conference on Multimedia, 1999.

[8] Zhang, J. Tan, Y. Automatic parsing and Indexing of news Video, Multimedia Systems, Vol2.

[9] Babaguchi, N. Kawai , Y. Event based video Indexing by internal collaboration, Proceedings of the first international workshop on multimedia intelligence Storage and retrieval management (MISRM'99) ,1999

[10] Haung, S. Hang, L. , A semantic network modeling for understanding baseball video , institute of electrical engineering national Tsinghua university.

[11] Li, J. Szafron, D. Modeling of moving object in a video database, IEEE international conference on multimedia computing and systems, (ICMCS) ,1997.

[12] Mulhem, P. Two system for temporal video segmentation, CBMI'99, France, 1999.

[13] Li, Y. Ming, W. Semantic video content abstraction based on multiple cues, IEEE international conference on multimedia and Expo , (ICME) 2001, Tokyo, Japan, 2001.

[14] Lew, M. S. Sebe , N. Challenges of Image and video retrieval, International conference on image and video retrieval, Springer, 2002.

[15] Yahiaoui, I. Meraldo, B. Automatic video summarization, internal report, Multimedia Communications Department, Institute EURECOM.

[16] Milan Petkovic, Willem Jonker, Content-Based Video Retrieval, Springer, ISBN: 1402076177, 2003.

[17] S Hassas, M S Hacid, Video Data, Kogan Page, ISBN: 1903996228, 2003.

[18] E. Oomoto , K. Tanaka, OVID: Design and Implementation of a Video-Object Database System, IEEE Transactions on Knowledge and Data Engineering, v.5 n.4, p.629-643, August 1993.

[19] S.Adali, M.L. Sapino, and V.S. Subrahmanian. An Algebra for Creating and Querying Multimedia Presentations. ACM/Springer Multimedia Systems Journal, 8(3):212-230, 2000.

[20] A. Picariello, M. L. Sapino and V. S. Subrahmanian. Algebraic Video Environment, in Handbook of Video Data Bases (e.g. B. Furht and O. Marques), CRC Press ISBN 084937006X, 2003.

[21] R. Weiss, A. Duda, D. K. Gifford, Content-based Access to Algebraic Video, Int. Conf. on Multimedia Computing and Systems, IEEE Press, 140-151.

[22] Koprinska I, Carrato S, Hybrid rule-based/neural approach for segmentation of MPEG compressed video. Multi-media Tools Application 18(3):187–212, 2002.

[23] Mohamed Ahmed, Roger Impey, Ahmed Karmouch: Developing Video Services for Mobile Users, HICSS 2003.

[24] A. Del Bimbo, E. Vicario, D. Zingoni. Symbolic Description and Visual Querying of Image Sequences Using Spatio-Temporal Logic. IEEE Transactions on Knowledge and Data Engineering, 7(4): 609–621, August 1995.

[25] H. D. Wactlar, T. Kanade, M. A. Smith, and S. M. Stevens. Intelligent access to digital video: Informedia project. IEEE Computer, 29(5):46--53, May 1996.

[26] M. M. Yeung, B. L. Yeo, W. Wolf, and B. Liu, "Video Browsing Using Clustering and Scene Transitions on Compressed Sequences," Multimedia Computing and Networking 1995, Proc. SPIE 2417, 399-413,1995.

[27] A SEMANTIC REPRESENTATION FOR IMAGE RETRIEVAL, Lei Wang and B.S. Manjunath,IEEE ICIP 2003,

[28] Emergent Semantics, Steffen Staab, IEEE INTELLIGENT SYSTEMS, IEEE, 2002.

[29] James F. Allen: Towards a general theory of action and time. Artificial Intelligence, 23:123-154, 1984.

[30] A Rule-Based Video Annotation System, Andres Dorado, Janko Calic, and Ebroul Izquierdo, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, VOL. 14, NO. 5, MAY 2004

[31] Smart VideoText: a video data model based on conceptual graphs, F. Kokkoras1, H. Jiang, I. Vlahavas1,, A.K. Elmagarmid, E.N. Houstis, W.G. Aref, Multimedia Systems, Springer-Verlag 2002.

[32] SEMANTIC CONTENT ANALYSIS OF BROADCASTED SPORTS VIDEOS WITH INTERMODAL COLLABORATION, Naoko Nitta, PhD Thesis, Osaka University,2003.

[33] An Ontology Framework For Knowledge-Assisted Semantic Video Analysis and Annotation, S. Dasiopoulou, V. K. Papastathis, V. Mezaris, I. Kompatsiaris and M. G. Strintzis, Egov Open Source Conference, March 2003 .