

## نمایه سازی ویدئو

زهرا سعدی

کارشناس ارشد کتابداری و اطلاع رسانی، دانشگاه تهران

### چکیده

هدف پژوهش حاضر بررسی شیوه‌های نمایه‌سازی ویدئو جهت ذخیره و بازیابی آن در مخازن اطلاعاتی است. روش پژوهش سندی است. یافته‌ها نشان می‌دهد که شیوه‌های نمایه‌سازی ویدئو هنوز از استاندارد مشخصی پیروی نمی‌کنند؛ اما سه روش عمده در نمایه‌سازی ویدئو عبارتند از: نمایه‌سازی مبتنی بر ویژگی، نمایه‌سازی مبتنی بر گزارمان و نمایه‌سازی براساس حوزه تخصصی هر یک از این شیوه‌ها معایب و مزایای خاص خود را دارد که بسته به ماهیت و کاربرد ویدئو مورد استفاده قرار می‌گیرد. کلیدواژه‌ها: نمایه‌سازی، ویدئو، نمایه‌سازی مبتنی بر ویژگی، نمایه‌سازی مبتنی بر گزارمان، نمایه‌سازی مبتنی بر حوزه تخصصی.

## مقدمه

استفاده از رسانه های دیداری شنیداری، به عنوان منابع اطلاعاتی یا سرگرمی و نیز به عنوان ابزاری تصویری یا ترویجی، به طور فزاینده ای رواج یافته است. در میان تمام رسانه های دیداری شنیداری، ویدئو نه تنها در زمینه های مرسوم پخش برنامه چون تلویزیون، رادیو، و سینما استفاده می شود، بلکه در تجارت، آموزش و پرورش، پزشکی، پیشگیری از جرم، و مانند آن نیز کاربرد دارد (ویکمی، ۱۳۸۰، ص ۱).

در رسانه هایی چون کتاب، معمولاً خوانندگان با مراجعه به قفسه ها و مرور اجمالی فهرست مندرجات و تورق کتاب ها از محتوای اطلاعاتی آنها آگاهی می یابند، ولی "بینندگان ویدئو" تا فیلم را نبینند از محتوای آن باخبر نمی شوند. از آنجایی که در انتخاب ویدئو، مشاهده کامل محتوای آن زمان بر است، نمایه سازی آرشیوهای ویدئویی به عنوان نوعی ابزار سازماندهی، جست و جو، و دسترسی سریع به محتوای موضوعی مدارک ضروری به نظر می رسد. از سوی دیگر، ضرورت آگاهی محققان از اطلاعات روزآمد پایگاه های اطلاعاتی، سرعت و دقت کافی جهت نمایه سازی بهینه مدارک ویدئویی را می طلبد که این امر مستلزم طراحی الگوریتم های کارآمد و دقیق نمایه سازی ویدئو است.

## نمایه سازی ویدئو

بر اساس استاندارد نمایه سازی بریتانیا<sup>۱</sup>، نمایه مجموعه ای منظم از شناسه های نشانه گذاری شده است تا کاربران بتوانند اطلاعاتی که محل آنها در مدرک مشخص شده، پیدا کنند (نیاکان، ۱۳۸۳، ص ۹). فرآیند تجزیه، تفکیک و تحلیل نحوی<sup>۲</sup> ویدئو را نمایه سازی ویدئو گویند (همپاپور، [۲۰۰۶]، ص ۱-۲).

### بررسی سند ویدئو به منظور نمایه سازی

زمانی که یک سند ویدئویی را نمایه سازی می‌کنند چند مسئله باید مد نظر قرار گیرد:

۱. چه چیزی را نمایه می‌کنند (مثلاً کل سند یا برخی از قاب‌های آن)؟
۲. چگونه نمایه سازی می‌کنند (مثلاً کدام رویکرد و شیوه نمایه سازی را به کار می‌گیرند)؟
۳. کدام قسمت را نمایه سازی می‌کنند؟ (مثلاً اسامی بازیکنان، موقعیت‌های زمانی آنها یا هر دو)؟ (اسنوک، ۲۰۰۵، ص ۱-۵).
۴. چرا این مدرک به مجموعه اضافه شده است (انتخاب مدرک)؟
۵. کاربران به چه جنبه‌هایی از مدرک علاقه مندند (نیاز کاربر)؟ (لنکستر، ۱۳۸۲، ص ۱۳)

### تأثیر چند کیفیتی بودن ویدئو در نمایه سازی

از آنجایی که پدیدآور ویدئو از مجراهای متنی، صوتی، و تصویری برای بیان عقایدش بهره می‌برد، محتوای ویدئو ذاتاً چند مدلی (یا کیفیتی) است. بنابراین، در نمایه سازی و بازیابی ویدئو باید سه جنبه معنی، محتوا، و طرح در نظر گرفته شود. از لحاظ معنی:

- در سلسله مراتب نمایه معنایی، پنج سطح مختلف مطرح می‌شود:
۱. هدف: برخی اسناد دارای هدف مشابه‌اند؛ به طور مثال، هدف اکثر فیلم‌های سینمایی سرگرمی است.
  ۲. نوع: برخی اسناد در سبک و نوع مشابه‌اند؛ به طور مثال، ویژگی‌ها و سبک‌های ویدئوهای خبری، مشترک‌اند.
  ۳. زیر نوع‌ها: برخی اسناد دارای محتوای مشابه‌اند؛ به طور مثال، فیلم‌های

ترسناک در جهاتی شبیه یکدیگرند.

۴. واحدهای منطقی: بخش‌هایی از محتوای سند ویدئویی در واحدهای منطقی مشابه‌اند؛ به طور مثال، گفت و گوها در فیلم‌های درام یا گزارش‌های مربوط به وضعیت هوا در اخبار تلویزیونی از جهاتی مشابه‌اند.

۵. وقایع مشهور: این وقایع، بخش‌های کوتاهی از ویدئوهای که به یک واقعه مربوط می‌شوند و در طول زمان تغییر نمی‌کنند؛ مانند انفجارها در فیلم‌های پرتحرک.

از لحاظ محتوا

منظور عناصری است که پدیدآور برای ایجاد سندی ویدئویی به کار می‌گیرد و عبارتند از:

الف. تنظیمات: زمان و مکانی که داستان ویدئو در آن اتفاق می‌افتد.

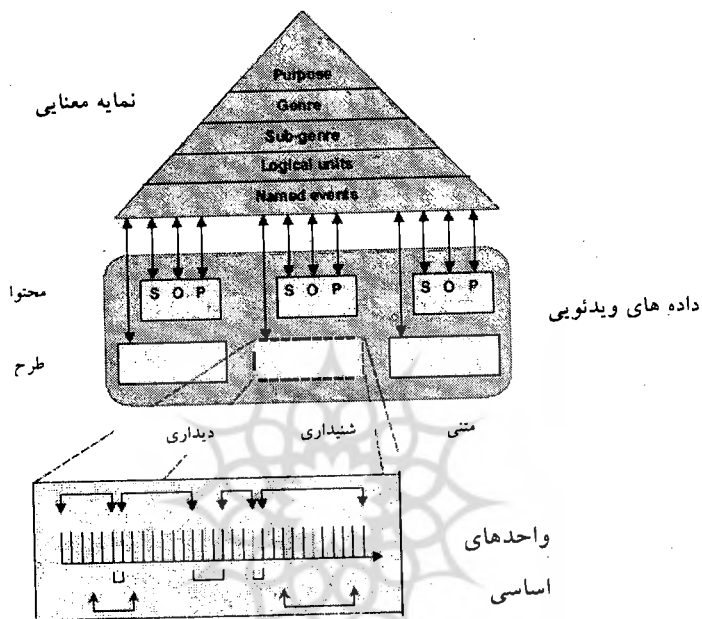
ب. اشیا: شامل اشیای ایستا و پویا در سند ویدئویی.

ج. مردم: شامل وجود انسان در اسناد ویدئویی.

در واقع تنظیمات به واحدهای منطقی مربوط می‌شوند. اشیا و مردم هم عناصر اصلی وقایع‌اند.

از لحاظ طرح

منظور از طرح همان ساختار نحوی است که پدیدآور برای سند ویدئویی به کار می‌گیرد. به طور مثال، در سند ویدئویی، مجرای دیداری مجموعه‌ای از قاب‌های متوالی منظم است اما در مجرای متنی واحدهای اصلی یکایک حروف‌اند. (استوک، ۲۰۰۵، ص ۲-۵). در شکل ۱ لایه‌هایی را که در نمایه سازی ویدئو باید مدنظر داشت، نمایش داده شده است.



شکل ۱. لایه های نمایه سازی ویدئو (اسنوک، ۲۰۰۵، ص ۱۰)

### شیوه های اصلی نمایه سازی

شیوه های اصلی در نمایه سازی ویدئو به سه دسته عمده تقسیم می شوند:

۱. نمایه سازی ویدئو بر اساس ویژگی ها<sup>۱</sup>؛
۲. نمایه سازی ویدئو بر اساس گزارمان (توضیح)<sup>۲</sup>؛
۳. نمایه سازی ویدئو بر اساس حوزه تخصصی (دامنه خاص)<sup>۳</sup>.

### نمایه سازی بر اساس ویژگی ها

این شیوه به سه دسته نمایه سازی مبتنی بر اجزاء<sup>۴</sup>، نمایه سازی مبتنی بر اشیا<sup>۵</sup>، و

1. Feature based video indexing

2. Annotation- based video indexing

3. specific domain-based indexing

4. segment-based indexing

5. Object-based

نمایه سازی مبتنی بر وقایع<sup>۱</sup> تقسیم می شود:

#### ۱. نمایه سازی مبتنی بر اجزاء

در طول فرآیند نمایه سازی متن، یک سند به اجزای کوچکترش مانند بخش، پاراگراف، جمله، عبارت، واژه، حرف، و عدد تقسیم گردیده و سرانجام نمایه ها بر اساس این اجزاء ساخته می شوند. ولی ویدئو به سلسله مراتبی شبیه به فیلمنامه مصور در فیلم سازی تقسیم می شود. این نوع طرح مرور اغلب استوری بورد (فیلمنامه مصور)<sup>\*\*</sup> نامیده می شود، زیرا مجموعه ای از قاب ها را در بر می گیرد که وقایع یا مفاهیم اصلی ویدئو را نشان می دهند. در این شیوه نمایه سازی، یک ویدئو شامل داستان هایی می شود مثل مهمانی تولد، تعطیلات، ازدواج، و مانند آن. هر داستان نیز شامل مجموعه ای از صحنه ها می شود مثل صحنه مهمانی در داستان ازدواج. هر صحنه نیز به نماهایی تقسیم می شود مثل نمایی از مهمانان خاص یا هیجانانگیز عروس و داماد. در نهایت هر نما از یک سکانس از قاب های منفرد تشکیل می شود که هر قاب، یک عکس است. برای نمایش اجزای ویدئو دو راه وجود دارد که عبارتند از:

الف. نمایش مبتنی بر قاب های کلیدی؛

ب. نمایش مبتنی بر صحنه

#### نمایش مبتنی بر قاب های کلیدی

یکی از رایج ترین راه های نمایش بخش های ویدئویی از طریق قاب های کلیدی است. به طور مثال، در یک نما قاب های مهم و معنی داری وجود دارد که معنای اصلی نما را در بر می گیرد. در انتخاب قاب های کلیدی مخصوصاً در فرآیند خودکار سازی، دو نکته را باید در نظر گرفت:

۱. تعداد قاب های کلیدی به کار رفته برای هر نما؛

۲. انتخاب مهم‌ترین قاب از طریق کاهش قاب‌های زائد یا دسته‌بندی قاب‌ها بر اساس میزان ربط. شکل ۲ تجزیه ویدئو به قاب‌های کلیدی را نشان می‌دهد.



شکل ۲. تجزیه ویدئو به قاب‌های کلیدی

ب. نمایش مبتنی بر صحنه

جریان ویدئویی از لحاظ زمانی به صحنه‌هایی تقسیم می‌شود. برش‌های صحنه از طریق تغییرات شدید در محتوای قاب‌ها مانند انتشار سطوح سیاه و سفید یا رنگی، شناسایی می‌شوند. نمایش مبتنی بر صحنه از سه طریق ایجاد می‌شود:

۱. موزائیک پانورامیکس<sup>۲</sup>: پانورامیکس، دارای یک دوربین عکاسی با عدسی گردنده است از نماها به طور مسلسل منظره فضایی می‌گیرد.

۲. تغییرات هندسی<sup>۳</sup>: قاب‌های ویدئویی مختلف به سیستم مناسب با موزائیک مربوط می‌شود. به طوریکه کاربران می‌توانند محل هر نقطه از پشت و جلوی صحنه را بین تصاویر موزائیک دورنما و قاب‌های مفرد، ترسیم کنند.

۳. اطلاعات پویا<sup>۴</sup>: این روش به ترسیم نمایش مسیرهای زمانی ممتد از اشیای متحرک و نقاط ظهور آن‌ها می‌پردازد.

1. cuts

2. Panoramix

3. Geometric transformations

4. Dynamic information

## ۲. نمایه سازی ویدئویی مبتنی بر اشیا<sup>۱</sup>

این شیوه نمایه سازی، به متمایزسازی اشیای خاص در یک سکانس ویدئویی جهت یافتن تغییرات محتوایی می پردازد. استخراج شیء وقتی به سادگی انجام می شود که ویدئو از طریق استاندارد کدگذاری مبتنی بر اشیا، مانند چون ام پگ<sup>۲</sup> فشرده شده باشد. از طریق این استاندارد می توان محتوای ویدئویی را در مقایسه با استانداردهای کدگذاری سنتی، مانند ام پگ<sup>۳</sup> و ام پگ<sup>۴</sup> با سطح بالاتری بیان کرد.

یک بخش ویدئویی<sup>۵</sup> فشرده با استاندارد، از گروهی اشیای ویدئویی<sup>۶</sup> تشکیل شده است. اشیای ویدئویی نیز از لایه های شیئی<sup>۷</sup> تشکیل شده است. لایه های شیئی هم شامل سکانس های منظمی از طرح های شیئی ویدئو<sup>۸</sup> شود. استاندارد ام پگ<sup>۴</sup> مشخص می کند که چگونه اشیا در جریان های بیتی فشرده، باید نمایش داده شود. مسیریابی و ردیابی اشیا از طریق شناسایی لبه های متحرک<sup>۹</sup> انجام می شود.

کیم و هوانگ<sup>۱۰</sup> تکنیکی برای بخش بندی اشیای ویدئویی براساس شناسایی تغییرات لبه، پیشنهاد کردند. تکنیک آن ها از کشف لبه کنی<sup>۱۱</sup> برای شناسایی و پیوند لبه ها در قاب های متوالی استفاده می کند. خروجی، یک نقشه لبه متحرک است که توسط طرح شیئی ویدئو بیان می شود. مرحله پس از پردازش شامل استخراج اشیای متحرک مورد نظر است (کیم، ۲۰۰۲، ص ۱۲۲-۱۲۹).

اسکوپ فیلم<sup>۱۲</sup> و دیگران یک الگوریتم مسیریابی مبتنی بر طرح (نقشه برجسته)<sup>۱۳</sup>

- |   |                             |
|---|-----------------------------|
| 1. Object-based video indexing techniques | 2. MPEG-4                   |
| 3. MPEG-1                                 | 4. MPEG-2                   |
| 5. Video session(VS)                      | 6. Video objects(VOS)       |
| 7. Video object layers(VOL)               | 8. Video object plane(VOPS) |
| 9. Moving edges                           | 10. Kim and Hawang          |
| 11. Canny edge detector                   | 12. Schoep film             |
| 13. Contour - based tracking algorithm    |                             |



ایجاد کردند که از یک سکانس تغییر شکل الگو به مسیر شیئی ویدئو استفاده می‌کند. به طور کلی، استخراج اشیا کار مشکلی است. مخصوصاً زمانی که اشیای درون قاب خیلی زیاد یا کوچک‌اند یا خیلی سریع حرکت می‌کنند بگونه‌ای که، تار و مبهم به نظر می‌آید. یاجیما<sup>۱</sup> و دیگران نوعی شیوه نمایه سازی پیشنهاد داده‌اند که به کاربران امکان می‌دهد حرکات اشیا را درخواست کنند. در این حالت، کاربران می‌توانند از طریق رسم مسیر حرکت روی یک صفحه ویدئو، مستقیماً به جست و جو پردازند. این تکنیک به ترسیم رابطه‌های زمانی و فضایی بین اشیای در حال حرکت بستگی دارد.

### ۳. نمایه سازی مبتنی بر وقایع<sup>۲</sup>

این تکنیک به شناسایی خودکار وقایع جذاب و جالب در مواد ویدئویی خام، کمک می‌کند. وقایع، همان روابط بین اشیای موجود در یک فاصله زمانی هستند که ممکنه قبل یا بعد از وقایع دیگر اتفاق بیفتند. یک واقعه شامل موارد زیر می‌شود:

الف. بافت‌های زمانی، مانند جریان آب؛

ب. فعالیت‌ها، مانند پیاده‌روی افراد؛

ج. وقایع حرکتی مجزا، مانند خندیدن.

در شناسایی وقایع ویدئو از تکنولوژی‌هایی مانند تحلیل تکنیک‌های تدوینی دوربین، تخمین حرکت جامع، استخراج پس زمینه و پیش زمینه به منظور تشخیص اشیای قابل مشاهده، کشف جریان‌های شرح کوتاه<sup>۳</sup> استفاده می‌شود. نقطه ضعف اصلی تکنیک‌های شناسایی وقایع این است که بسیار گران و زمان براند و به تحلیل قاب‌های متعددی نیاز دارند: (تیجین درانگورو، ۲۰۰۵، ص ۱۸-۳۲؛ دی جبرا، ۲۰۰۵، ص ۲).

1. Yajima

2. Event-based video indexing techniques

3. Closed caption(CC)

## نمایه سازی ویدئویی مبتنی بر گزارمان

یکی از راه‌های مدیریت داده‌های ویدئویی، توضیح و گزارمان نویسی مفاهیم معنایی بخش‌های ویدئو از طریق کلیدواژه‌ها یا متن‌های آزاد است. در این شیوه، درخواست کاربر از طریق زبان پرسش استاندارد (اس کیو ال)<sup>۱</sup> و مرور سلسله مراتبی موضوعی، کنترل می‌شود. از معایب اصلی این روش این است که تفسیر و یادداشت برداری دستی هر بخش از ویدئو و همچنین فرآیند انطباق<sup>۲</sup> ویژگی‌های ویدئویی سطح پایین با مفاهیم معنایی سطح بالا کار دشواری است.

انتخاب متن آزاد یا کلید واژه‌ها، ذهنی است و اغلب به دانش و تخصص حوزه‌ای بستگی دارد. یک تصویر، معادل هزار واژه است یعنی واژه‌ها اغلب در توصیف کامل یک قاب ناتوان‌اند و نهایتاً برای توصیف یک بخش ویدئو، کافی نیستند. کاربران هم نمی‌دانند چگونه درخواست‌هایشان را از طریق واژه‌ها توضیح دهند و بیشتر تمایل دارند که بر اساس شباهت تصویری یا صوتی در میان مدارک به جست و جو بپردازند.

با وجود این محدودیت‌ها، این روش هنوز بعنوان بهترین روش در نمایش محتوای معنایی ویدئو به کار برده می‌شود. مثلاً در یک ویدئوی ورزشی یا خبری واژه‌های بسیاری وجود دارد که باید یک واژه‌نامه<sup>۳</sup> از آنها تهیه شود تا سبب نمایه سازی یکدست تر و سهولت در امر بازیابی شود (گوردن، ۲۰۰، ص ۱). این شیوه نمایه سازی به دو طریق انجام می‌شود:

۱. نمایه سازی مبتنی بر گزارمان از طریق رده بندی موضوعی؛

۲. نمایه سازی مبتنی بر گزارمان از طریق ام پگ ۷.

۱. نمایه سازی مبتنی بر گزارمان از طریق رده بندی موضوعی<sup>۴</sup>

رده بندی موضوعی عمدتاً برای سازماندهی مقادیر زیاد اطلاعات مانند کتابخانه و یا

1. Standard query language (SQL)

2. mapping

3. glossary

4. Subject (topic) classification

موتورهای جست و جو به کار می‌رود. برای مرور، کاربران می‌توانند موضوعات موجود را انتخاب کنند و یا بر اساس کلیدواژه‌ها، موضوعات مورد نظر را جست و جو کنند (پتکویک، ۲۰۰۲، ص ۱-۲).

یکی از محدودیت‌های اصلی این روش این است که اکثر کاربران درباره چگونگی رده‌بندی موضوعی، نظرات مختلفی دارند. ولی باید گفت اغلب ویدئوها مانند اخبار و فیلم‌های سینمایی، موضوعات ساختاریافته خوبی دارند. ویدئوهای ورزشی نیز بر اساس نوع ورزش (گروهی یا فردی) و یا نوع وقایع رده‌بندی می‌شوند. لو<sup>۱</sup> رده‌بندی موضوعی ویدئو مطرح کرد. لو ویدئو را به رده‌های حیوانات، هنرها، غذاها، ورزش‌ها، قازّه‌ها، و مانند آن تقسیم کرد. این رده بندی در جدول نمایش داده شده است.

۱. جدول رده بندی موضوعی لو

شماره	رده بندی	مثال ها
۱	حیوانات	سگ‌ها، گربه ها
۲	هنرها	نقاشی، مجسمه سازی
۳	غذاها	نوشیدنی‌ها، میوه‌ها
۴	ورزش ها	فوتبال، بسکتبال
۵	قازّه‌ها	آسیا، اروپا
۶	گربه‌سانان	شیر، ببر
۷	اروپا	فرانسه، آلمان
۸	سرگرمی	موسیقی، ورزش، فیلم
۹	اطلاعات	اخبار، مستندات
۱۰	ارتباطات	آگهی بازرگانی

مأخذ: (تبیجین درانگورو، ۲۰۰۵، ص ۳۴)

اسنوک و دیگران<sup>۱</sup> نیز سند ویدئویی را به سه رده سرگرمی، اطلاعات و ارتباطات تقسیم کردند. طرح رده‌بندی اسنوک از لو پیچیده‌تر است ولی آسان‌تر گسترش می‌یابد (تیجین درانگورو، ۲۰۰۵، ص ۳۴).

در رده‌بندی باید به کاربردهای ویدئو از جمله ورزش، سینما، تلویزیون، امنیت، اخبار و مانند آن، هدف از ایجاد ویدئو، محتوای ویدئو، فرآیند تولید محصول ویدئویی، و روشی که ویدئو به کار گرفته شده است، توجه شود.

به منظور تقسیم رده‌ها به واحدها و وقایع منطقی، از رویکرد نمایه‌سازی مبتنی بر حوزه تخصصی (دامنه خاص) استفاده می‌شود.

## ۲. نمایه‌سازی گزارمانی از طریق اساس ام پگ ۷<sup>۲</sup>

ام پگ ۷ از طرف گروه تخصصی تصاویر متحرک به عنوان یک استاندارد ایزو/آی ای سی (کمیته فنی مشترک)<sup>۳</sup> برای توصیف انواع مختلف اطلاعات دیداری شنیداری آرشیوی بلادرنگ و غیر بلادرنگ معرفی شده است.

طرح‌های توصیفی چند رسانه‌ای ام پگ ۷، محتوای چند رسانه‌ای را از لحاظ معنایی و ساختاری، توصیف می‌کنند. توصیفات معنایی استاندارد ام پگ ۷ توصیف موجودیت‌هایی مانند (اشیا، وقایع، مفاهیم حالت‌ها، حالات و وضعیت‌های، زمان‌ها و مکان‌های معنایی در جهان داستانی و روایی) تأکید می‌کنند. اشیا، وقایع، و موجودیت‌ها در داده‌های چند رسانه‌ای قابل مشاهده‌اند ولی مفاهیم معنایی قابل مشاهده نیستند بلکه به عنوان کلیتی از موجودیت‌های معنایی قابل توصیف‌اند. در واقع، توصیف‌گرهای معنایی برای توصیف مفاهیم انتزاعی<sup>۴</sup> استفاده می‌شوند [که به یک نمونه محتوای چند رسانه‌ای یا مجموعه‌ای از آنها تعمیم داده می‌شود] (تیجین درانگورو، ۲۰۰۵، ص ۲۴-۳۸؛ کوگیا، ۲۰۰۵، ص ۱۷۹؛ آلمدیا، ۲۰۰۴، ص ۲).

1. snoek

2. MPEG-7 based annotation indexing

3. ISO/IEC<sup>40</sup> Real time

4. abstractions

نمایه سازی مبتنی بر دانش حوزه تخصصی(دامنه خاص)

برای از بین بردن شکاف و فاصله معنایی بین مفاهیم سطح بالا و ویژگی های سطح پایین از شیوه نمایه سازی مبتنی بر دانش حوزه ای استفاده می شود. روش های پل زدن بین فواصل معنایی<sup>۱</sup> به دو دسته تقسیم می شوند:  
۱. درخواست از طریق مثال با اندازه گیری میزان ربط<sup>۲</sup>:

از آنجایی که استخراج ویژگی دیداری شنیداری آسانتر از فهم معنایی است و ممکنه کاملاً خودکار انجام شود، بازیابی از راه درخواست با مثال، بسیار مفید است. به طور مثال، اگر یک نمونه نمای ویدئو درخواست شود، سیستم باید بخش های نمایه شده ای که نزدیک ترین شباهت را با آن دارند، پیدا کند. مانند، شباهت اوج سخنرانی با حالات صورت فرد.

کیم و چوآکیم<sup>۳</sup> تکنیک تطبیق (جورسازی)<sup>۴</sup> مشترک ترین زیرسکانس ها را توسعه دادند. آن ها به اندازه گیری شباهت زمانی بین بخش های مشابه در ویدئوی خبری پرداختند (یانگ تا، ۲۰۰۴، ص ۶۸-۷۵).

مسئولیت تعیین اهمیت ها و میزان ربط نتایج بدست آمده با کاربر است و لازمه که آنها را علامت گذاری کند تا تصاویری که مربوط به پرسش اند، بازیابی شوند. اهمیت ها و وزن ها در درخواست برای نمایش مفاهیم سطح بالا و درک ذهنی بیشتر مورد توجه واقع می شود (تیجین درانگورو، ۲۰۰۵، ص ۳۸-۴۰؛ پولیت، ۱۳۸۰، ص ۱۱۵).

۲. سطحی نگری هوشمند و سریع (یا سریع نگری هوشمند)<sup>۵</sup>:

سطحی نگری ویدئو شبیه سطحی نگری کتاب است. به طور مثال، اگر کاربری بخواهد از طریق یک کتاب بفهمد که چگونه یک وسیله ای را می سازند، ممکنه به جای خواندن کل کتاب فقط به مثال ها و نمودارها در بخش های مشخصی از کتاب مراجعه

1. Indexing by bridging semantic Gap

2. Query by example(QBE) with relevance

3. Kim & chua kim

4. matching

5. Intelligence fast-forward(smart skimming)

کند. در یک ویدئو هم وقتی کاربر می‌خواهد بخش مشخصی از ویدئو را بدون در نظر گرفتن مابقی به کار گیرد، سطحی‌نگری و مرور کلی می‌کند. سطحی‌نگری کاربر را قادر می‌سازد تا هدف محتوای ویدئو را از نسخه کوتاه‌تری نسبت به اصل ویدئو بشناسد. سطحی‌نگری زمان لازم برای تحلیل را کاهش می‌دهد (تیجین درانگورو، ۲۰۰۵، ص ۴۰-۵۳). از تکنیک‌های سطحی‌نگری می‌توان به موارد زیر اشاره کرد:

#### الف. مرور از طریق نمایش فشرده<sup>۱</sup>

اسناد ویدئویی به دلیل ابعاد فضایی و زمانی، حجیم‌اند. بنابراین، هنگام تحلیل و کاوش، کاربران باید از نمایش فشرده‌تر اسناد ویدئویی بهره‌برند. برای فشرده‌سازی اطلاعات ویدئویی سطح بالا از تکنیک‌های تلخیص یا سطحی‌نگری استفاده می‌شود؛ مانند کادرهای پیش‌نمایش کاربرپسند در واسطه کاربری موتور جست‌وجوی گوگل.

#### ب. تلخیص<sup>۲</sup>

تلخیص فرآیند بیان زیرمجموعه‌های قابل‌نمایش از کل یک بریده فیلم<sup>۳</sup> است که جنبه‌های معنایی مهمی داشته باشد.

فایزولین<sup>۴</sup> در تلخیص، قواعد سی‌پی‌آر<sup>۵</sup> را از طریق سه مفهوم اصلی یعنی اولویت قابی، ادامه‌تلخیص، و غیر تکراری‌ها پیشنهاد کرد. از نظر او، اصل مفهوم تلخیص ویدئو شامل قاب‌هایی می‌شود که مهم‌تراند و اولویت آن‌ها بالاتر از وقایع غیرمهم است؛ ادامه و اتصال تلخیص باید برای کاربران قابل‌درک باشد ولی تلخیص نباید تکراری باشد. موریا<sup>۶</sup> و اسکاچی<sup>۷</sup> در تلخیص از محتوای روان‌شناسی استفاده کردند. ما<sup>۸</sup> مدل

- 
- |   |                  |
|---|------------------|
| 1. Browsing with compact representation | 2. summarization |
| 3. Video Clip                           | 4. fayzullin     |
| 5. CPR                                  | 6. moriyama      |
| 7. skauchi                              |                  |

توجه کاربری را برای تلخیص ویدئو پیشنهاد کرد. ارائه وقایع مهم داخل بسته‌بندی برجسته، در صرف زمان و انرژی کاربر هنگام جست و جوی کل ویدئو بمنظور بازیابی یک واقعه مهم، صرفه جویی می‌کند. به طور مثال در ویدئوهای امنیتی دوربین در مکان خاصی معمولاً ثابت واقع می‌شود و ممکنه شامل ساعت ها ویدئوی غیرفعال باشد در این ویدئوها، به جای اینکه کاربران کل ویدئو و سکانس‌ها را تماشا کنند، می‌توانند از تلخیص نماهای مهمی که فعالیت مورد نظر آن‌ها را شامل می‌شود، استفاده کنند (ما، ۲۰۰۲، ص ۱-۱۰).

به منظور نمایه‌سازی بر اساس دانش حوزه تخصصی، باید ابتدا مدل دانش ویدئو مشخص شود.

#### مدل دانش

معمولاً الگوریتم‌های نمایه‌سازی خودکار ویدئو سه نوع دانش فیزیکی، سینمایی، و معنایی را به کار می‌گیرند.

۱. دانش فیزیکی<sup>۹</sup>: این دانش شامل مسائل ناشی از محیط فیزیکی می‌شود؛ به طور مثال، در ورزشی مانند بسکتبال شامل رنگ توپ، ساختار تخته پشت حلقه، و سبد می‌شود.

۲. دانش سینمایی<sup>۱۰</sup>: این دانش شامل جزئیاتی از چگونگی وقایع خاص فیلم و تولید آن‌ها می‌شود. به طور مثال، حرکت دوربین، تدوین و مکان‌هایی که در تولید ویدئو استفاده شده است.

۳. دانش معنایی<sup>۱۱</sup>: شامل دانشی درباره ویدئو می‌شود. در ویدئوی ورزشی شامل ساختار زمانی، قوانین بازی، و سایر اطلاعات سطح بالای ورزشی می‌شود؛ به طور مثال، بازی بسکتبال چهارتا یک ربع با ۱۳۰ ثانیه استراحت بین آنها دارد.

8. ma

9. physical knowledge

10. cinematic knowledge

11. semantic knowledge

با مطالعاتی که محققان روی تکنیک‌های مختلف نمایه سازی انجام داده‌اند، مشخص شده است که اکثر تکنیک‌ها برای بیان برجسب‌های<sup>۱</sup> وقایع و ویژگی‌ها، بر روی دانش سینمایی و فیزیکی تأکید دارند تا دانش معنایی لذا برای حل این مشکل، محققان یک مدل دانش ارائه کرده‌اند. همان طور که مدل داده‌ای در یک سیستم پایگاه اطلاعاتی مهم‌ترین عنصر است، در سیستم نمایه سازی ویدئو نیز مدل دانش<sup>۲</sup>، هسته اصلی است. هر مدل دانش شامل چند مدل فرعی می‌شود که انتخاب مدل‌های فرعی و طرح نمایشی دانش کاربردی، به دامنه و ماهیت درخواست‌هایی که از سوی سیستم حمایت می‌شود، بستگی دارد.

مقایسه شیوه‌های نمایه سازی

در جداول ۲ و ۳ به بررسی برخی مشخصات اصلی سه شیوه نمایه سازی، مزایا و معایب آنها پرداخته شده است.

پژوهشگاه علوم انسانی و مطالعات فرهنگی  
پرتال جامع علوم انسانی



جدول ۲. مشخصات و ویژگی‌های شیوه‌های نمایه سازی ویدئو

مشخصات	مبیتی بر ویژگی	مبیتی بر گزارمان	حوزه تخصصی
پایه‌های نمایه سازی	بخش‌ها، اشیا، و وقایع	مفاهیم معنایی	ساختار نوعی
تکنیک‌های مورد نیاز	تحلیل اطلاعات صوتی،	تشخیص کاراکتر نوری،	ترکیب هرنوع تکنیکی،
تصویری، و ویدئویی	تشخیص گفتار، تحلیل	اما تخصصی برای آن	کاربرد
توضیح کوتاه			
نحوه انجام	کاملاً خودکار	نیمه خودکار	دستی
ساختار نمایه ها	براساس بخش‌ها و	رده بندی موضوعی	حوزه و نیازمندی‌های
قالب‌ها	استفاده از ام پگ ۷	خاص کاربر	مرور و جست و جواز
مرور جست و جو	-مرور روی بخش‌ها	-مرور روی موضوعات	طریق ویژگی‌های استخراج
اشیا، و وقایع	- جست و جو از طریق	کلید واژه‌ها	شده یا گزارمانها بر
- جست و جو از طریق	مشخصات صوتی، تصویری	اساس کاربرد خاص	
یا پرسش از طریق مثال			

مأخذ: تیجین درانگورو، ۲۰۰۵، ص ۵۸.

جدول ۳. جدول مقایسه مزایا و محدودیت‌های شیوه‌های نمایه سازی

شیوه ها	هدف اصلی	مزایا	محدودیت ها
مبتنی بر ویژگی	نمایه سازی، اجزای ویدئو براساس ویژگی‌ها	- سادگی خودکارسازی - حمایت پرسش از طریق مثال	- کمبود مفاهیم معنایی - عدم ربط مستقیم ویژگی‌های سطح پایین با مفاهیم سطح بالا
مبتنی بر گزارمان	نمایه سازی سکانس‌های ویدئو از طریق گزارمان‌های استخراج شده از طریق تشخیص گفتار، تشخیص کارکتر نوری ویدئو، یا شرح کوتاه	حمایت از بازیابی معنایی	- مشکل بودن خودکارسازی کامل - محدودیت کاربرد هنگام درخواست از طریق ویژگی‌های سطح پایین
حوزه تخصص (دامنه خاص)	توسعه مناسب‌ترین استراتژی‌ها برای نیازمندی‌های خاص - حمایت از درخواست‌های تخصصی تر - هماهنگی مفاهیم معنایی با دانش کاربر	- یکپارچه سازی و انطباق گزارمان‌ها و ویژگی‌ها	تفاوت در نیازها و تخصص‌ها

## مآخذ

پ پولیت، ا. استون (۱۳۸۰). نظام های ذخیره و بازیابی اطلاعات: خاستگاه، توسعه و کاربردها. ترجمه محمد حسین دیانی، جعفر مهاد. شیراز: کتابخانه منطقه ای علوم و تکنولوژی.

لنکستر، فردریک (۱۳۸۲). نمایه سازی و چکیده نویسی: مبانی نظری و عملی. ترجمه عباس گیلوری، تهران: چاپار.

نیاکان، شهرزاد (۱۳۸۳). نمایه سازی ماشینی. تهران: مرکز اطلاعات و مدارک علمی ایران.

ویکمی، کریس (۱۳۸۰). مدیریت مجموعه های فیلم و ویدئو. ترجمه لیلا مرتضایی. تهران: مرکز اطلاعات و مدارک علمی ایران

"Audiovisual Archive with MPEG-7 video de scription and XML DATABASE"(2004).Pedro Almedia and etal.[on-line]. available: <http://www.ieeta.pt/ujsp/papers/>

"content-based video indexing for the support of digital library search"(2002).M Petkovic and etal[on-line].available : [http://csdi.computer.org/comp/preceeding/icde\(IEEE computer society\)](http://csdi.computer.org/comp/preceeding/icde(IEEE computer society))

Cuggia,Marc(2005)."indexing method of digital audiovisual medical resources with semantic web integration".Fleur Mougine,pierrele Beux [online].available: <http://www.ncbi.nlm.nih.gov> (university of rennes.1,france)

Djerba,Chabane(2005)." Systems and architectures for multimedia information retrieval".Nicu sebe.lew.[on- line].available:[www.lifl.fr/djeraba/acm,mm.2005.pdf](http://www.lifl.fr/djeraba/acm,mm.2005.pdf)

Gordon,Andrew S.(2006)"conceptual indexing for video retrieval".Eric A.Domeshek.[on-line].available:[www.ict.usc.edu/Gordon/ijca195](http://www.ict.usc.edu/Gordon/ijca195)(North western university)

Hampapur,Arun(2006)."semantic video indexing: approach and issues".IBM tjwatson research centre.[on-line].available : <http://www.research.ibm.com/evcvy/pubs/arun-sem.pdf>

"Handbook for digital projects: a management tool for presentation (2006).(on-line).available :<http://www.nedcc.org/digital/iiii.htm>O#project"

Kim,C.(2002)."fast and automatic video object segmentation and tracking for content-based applications".J.N.Hawng.[on-line].available:<http://ieeexplore.ieee.org/Xplore/login.jsp?url</tel5/76/21285/00988659.pdf>

tp< &arnumber<988659&isnumber=21285L.LU.

Ma, Y.F.(2002)."A user attention model for video summarization.L.LU,

Z.Hong-jiang,andMLi.[on-line].available:[http://research.microsoft.com/users/llu/Publications/ACMMM02\\_Attention.pdf](http://research.microsoft.com/users/llu/Publications/ACMMM02_Attention.pdf)

Snoke,Cees G.M(2005)."Multimodal video indexing: A Review of the state-of-the art".marcel worring  
.[on-line].available:<http://www.ecrim.org/pub/bscw.cgi/d28911/mmta.pdf>

Tjondronegoro,Dian(2005)phd thesis:"content-based video indexing for sports applications using  
multi-modal approach".[on-line].available:[http://eprints.qut.edu.au/archive\(Queensland university of  
technology\)](http://eprints.qut.edu.au/archive(Queensland%20university%20of%20technology))

Yaung-tae,K.(2004)."retrieval of news video using sequence  
matching".Tat-seng.[on-line].available:<http://csdl2.computer.org/persagen/DLAbsToc.jsp?resourcePath</dl/proceedings/&toc<comp/proceedings/mmm/2005/2164/00/2164toc.xml&DOI<10.1109/MMMC.2005>