Ciência da Informação

REFERÊNCIA

BARRETO, Juliano Serra. Desafios e avanços na recuperação automática da informação audiovisual. **Ciência da Informação**, v. 36, n. 3, p. 17-28, set./dez. 2007. DOI: https://doi.org/10.1590/S0100-19652007000300003. Disponível em: https://www.scielo.br/j/ci/a/3RKGt6c6RY4pCFYWcypKh5B/?lang=pt#. Acesso em: 06 jun. 2022.

Desafios e avanços na recuperação automática da informação audiovisual

Juliano Serra Barreto
Mestre em artes.
Professor da Universidade de Brasília (UnB).
E-mail: serra@unb.br

RESUMO

Exposição sobre processos e métodos utilizados para a indexação e recuperação textual da informação semântica em vídeo, tendo como base a identificação e classificação do seu conteúdo visual e sonoro.

PALAVRAS-CHAVE

Sistemas de recuperação da informação visual. Indexação de vídeos. Recuperação do conteúdo audiovisual.

Challenges and advancements in automatic retrieval of audiovisual information

ABSTRACT

Presentation of methods and processes applied to classification and retrieval of semantic information of video programs, through identification of sound and visual content.

KEYWORDS

Content based image retrieval. Video indexing. Multimedia content retrieval.

INTRODUÇÃO

Se "uma imagem vale por mil palavras", pode-se dizer que para uma descrição total do que vemos em um comercial de televisão de 30 segundos, precisaríamos de cerca de 720 mil palavras. É um número expressivo, mas para um sistema eficiente de recuperação de informação audiovisual não é necessário chegar a valor tão elevado de descritores. Neste artigo serão relatados alguns avanços da pesquisa cujo objeto de estudo é a correspondência entre elementos visuais e significados verbais, e que se desenvolve integrando áreas como a psicologia da cognição, a inteligência artificial, a produção audiovisual e a ciência da informação.

Desde as tábuas sumérias até a atualidade, muitos materiais duráveis serviram para preservar informações mas, o que se pode prever para o que hoje é acondicionado em suportes eletrônicos, e que já constitui em grande escala a nossa herança cultural e intelectual para as futuras gerações? No caso do audiovisual, as vantagens do registro eletromagnético estão condicionadas à enorme fragilidade dos meios, se comparados ao material fotográfico, pois a informação digital, dependente da alta rotatividade da informática, para permanecer exige cuidados especiais, desde a sua criação até a sua conservação. Somente a manutenção de uma política duradoura e de cooperação entre os fabricantes de *hardware* e desenvolvedores de *software*, os distribuidores e produtores de mídia, e com a participação de bibliotecas, arquivos e museus poderemos esperar que nossas mensagens sejam ainda acessadas no futuro.

A invenção do cinema e a rápida multiplicação dos meios e processos, que geraram enorme quantidade de material audiovisual, literalmente transformaram a face do mundo e continuam modificando os padrões da atividade humana. Em alguns países, tal acervo é reconhecidamente um repositório valiosíssimo de informações, mas ainda assim é na prática um tesouro oculto, pois as descrições sobre os conteúdos poucas vezes incluem algo mais que títulos e curtas sinopses. No Brasil porém, pouco mais do que 5 % de todo o material em película produzido até os anos 40 permanece atualmente preservado. A criação de ferramentas que podem permitir a pesquisa por entidades e conceitos registrados em filmes está sendo empreendida não somente por filmotecas e museus, mas também de forma intensa pelos produtores de mídia, que se preparam para oferecer conteúdo audiovisual personalizado via internet e televisão digital.

Na implementação de aplicações que vão de bibliotecas digitais a sistemas de segurança, serão necessárias novas ferramentas que permitam o acesso facilitado ao conteúdo de audiovisuais. A seguir apresentam-se as tendências atuais e propostas de solução para a interpretação semântica automática do que é genericamente denominado produto audiovisual, e que abrange toda a produção de imagens em movimento feita através de câmeras de diversos formatos, utilizadas em ritmo crescente na sociedade contemporânea.

INFORMAÇÃO AUDIOVISUAL

A humanidade vem produzindo ícones há pelo menos 7.000 anos, porém com a revolução digital estamos agora experimentando uma relação muito íntima e quase absoluta com a imagem, alcançando novo patamar que pode nos levar inclusive a situações extremas de vigilância total, como a do *Big Brother* imaginado por George Orwell no romance "1984".

A informação visual tem sido armazenada de forma analógica e indexada manualmente, mas hoje muitos sistemas de base de dados digitais são utilizados para armazenar imagens, juntamente com seus metadados e taxonomias associados. Sistemas híbridos, com indexação automática e análise de conteúdo supervisionada devem ser desenvolvidos, pois existem sérias limitações ao uso de indexadores manuais, uma vez que requerem anotação individual, dificultando seu uso em grandes arquivos, e que sofrem influência tanto do domínio de aplicação quanto do conhecimento da pessoa que realiza a tarefa. O reconhecimento de imagens e sons é parte da área de sistemas de recuperação da informação, em que se colocam grandes desafios relativos ao armazenamento, indexação, formulação de consultas e recuperação de conteúdo semântico.

Ao se considerarem seqüências de imagens, o problema de indexação torna-se mais difícil, pois envolve a identificação e o entendimento de cenas longas e complexas para que seja possível obter uma recuperação precisa e eficiente. Atualmente existem sistemas que permitem aos usuários especificar buscas em repositórios de imagens por meio da seleção de elementos visuais, como cor e textura, pelas comparações de imagens-exemplo e pelo reconhecimento de padrões espaciais, ou temporais no caso do vídeo.

Nas seções seguintes serão consideradas as consequências do aumento da produção audiovisual, assim como as formas de registro e preservação de vídeo. Também serão revistos os processos utilizados na análise fílmica e os padrões propostos para a indexação de materiais audiovisuais.

A aceleração da produção midiática

As estimativas produzidas por Kompatsiaris (2006) revelaram alguns números impressionantes para a produção audiovisual nos anos vindouros. Em todo o mundo, 1-2 exabytes (bilhões de gigabytes) de conteúdo eletrônico serão produzidos e 80 bilhões de imagens digitais serão feitas anualmente. Mais de um bilhão de imagens relacionadas a transações comerciais estão disponíveis e devem aumentar dez vezes nos próximos dois anos. A cada ano, 4 mil novos filmes serão produzidos, além dos 300 mil já disponíveis em todo o globo. E serão ultrapassadas as 100 bilhões de horas de material audiovisual distribuídas por 33 mil estação de televisão e 43 mil de rádio. Como podemos lidar com tal quantidade de documentos e metadados, que já é assustadoramente denominado sobrecarga informacional? Que ferramentas podem viabilizar a organização de tal produção? Nesse contexto, como encontrar a informação necessária, no momento preciso?

A rápida transformação dos procedimentos e materiais de reprodução audiovisual permite grande variedade de formatos e suportes, mas alguns fundamentos básicos ainda prevalecem. A *câmera obscura* ainda é o *design* básico de qualquer aparelho utilizado para registrar imagens da realidade visível, embora o processo eletrônico já não comporte o uso da prata nem as reações químicas. Entretanto, a conservação de documentos baseados em prata, como filmes e fotografias, embora seja delicada, é conhecida e eficiente, obtendo-se documentos que podem se manter inalterados por até mais de um século. Tais produtos presumidamente terão vida útil mais longa do que os documentos guardados, em meio magnético, e mesmo em dispositivos óticos, sobretudo quando dependem de *software* e *hardware* específicos para serem lidos. Com estas e outras preocupações, já vêm sendo pesquisados parâmetros mais permanentes para a preservação da informação em formatos digitais, como é possível encontrar nas definições propostas pela British Library em 1998, que têm sido aprimoradas desde então (BEAGRIE, 1998).

Os sistemas de redes distribuídas estão também modificando profundamente a estrutura e a linguagem da experiência cinemática. Novas possibilidades de interação entre autores e públicos permitem a criação de filmes adaptativos, multiplicando os níveis de leitura e explorando eventos em tempo real. Para Paul Virilio, estamos mesmo inaugurando um novo estatuto para a imagem, uma era da lógica paradoxal, em que a imagem se impõe à coisa representada, e que desestabiliza as representações públicas tradicionais, em benefício de uma apresentação, de uma presença paradoxal que supre a própria existência. Em suas palavras: "Esta virtualidade que domina a atualidade, subvertendo a própria noção de realidade" (VIRILIO, 1994).

O custo da produção audiovisual, no que concerne à geração e gravação de imagem e som, tem caído progressivamente, à medida que componentes eletrônicos são fabricados em maior quantidade e com maior capacidade, e consumidos em larga escala. Assim

vemos uma nova e vigorosa popularização do audiovisual no mundo industrializado, que invade todos os pontos da Terra e todos os recantos em que se encontra a presença humana. Desde o ponto de vista das câmeras de vigilância, nas ruas e nas escolas, até o do interior das residências, em webcams e nos celulares, multiplica-se em proporção geométrica a produção de imagens em movimento. A consolidação de sistemas que permitem eficiente catalogação e busca em acervos multimídia permitirá uma relação mais interativa e funcional com o audiovisual, mais personalizada e ao mesmo tempo, mais difusa.

O vídeo digital

O vídeo é uma tecnologia de processamento de sinais eletrônicos, que podem ser analógicos ou digitais, desenvolvida para apresentar imagens em movimento, aproveitando-se do efeito fisiológico da persistência retiniana, assim como é feito no processo cinematográfico. Um filme é uma seqüência de imagens fixas que, exibidas a taxas em torno de 20 quadros por segundo, apresentam uma ilusão visual de movimento no plano bidimensional da tela de projeção. O agrupamento dessas imagens formando um filme ou programa reflete uma organização definida na fase de edição e em geral é prevista por um roteiro. A edição é um processo de colagem linear de trechos de imagens e sons sincronizados que formam conjuntos denominados planos (ou tomadas), cenas e seqüências. As seqüências formam os grandes blocos narrativos, sendo análogas a capítulos de livros, na composição do filme. As seqüências contêm cenas, que são como parágrafos, trechos da narrativa com unidade lógica e visual. Por sua vez, uma cenas é um agrupamento de planos, sendo cada plano um subconjunto dos fotogramas, ou quadros obtidos em operação única da câmera.

Na identificação de conteúdos é importante a discriminação destes níveis hierárquicos, o que é feito com o reconhecimento de padrões nas imagens isoladas e também no fluxo de imagens. No entanto, o vídeo digital é apresentado usualmente de forma comprimida, dificultando este tipo de análise em alguns formatos de arquivo multimídia. A baixa qualidade de exibição afeta negativamente os algoritmos de extração de características visuais e a localização de eventos dinâmicos como transições entre cenas.

No vídeo digital eliminam-se redundâncias entre dois quadros subseqüentes utilizando-se padrões de compressão de imagens, para se obter um arquivo mais leve e fácil de ser manipulado. O algoritmo utilizado para essa compressão é chamado codec, e o arquivo que contém o programa codificado é chamado container. A indústria tem apresentado muitas soluções de formatos que agem como containers, e estes podem incorporar diversos codecs. Os formatos e codecs de arquivos de vídeo mais conhecidos são os seguintes:

- o DivX é um *codec* com elevada taxa de compressão, que pode reduzir o tamanho de um filme em DVD de 6 GB para 700MB sem perder muita qualidade;
- o MJPEG é um *codec* que guarda cada *frame* como uma imagem JPEG separada. A qualidade é ótima e independente do movimento na seqüência de vídeo. Nos vídeos em MPEG, a qualidade decresce quando a seqüência tem muito movimento;
- o MPEG é uma família de formatos de compressão padronizados pelo Moving Picture Experts Group[1], o qual é formado por cerca de 350 organizações. O MPEG-1 é o padrão básico de compressão de áudio e vídeo. O MPEG-2 é um conjunto de padrões voltados para a difusão televisiva de qualidade. O MPEG-4 usa um algoritmo H.264 para altas taxas de compressão. Suporta o *Digital Rights Management* (DRM), para controle de direitos autorais e é hoje o *codec* mais usado para *streaming* multimídia na Internet e na difusão televisiva, com o *container* MP4. De especial interesse para a recuperação da informação audiovisual é o formato MPEG- 7, uma proposta de padronização da descrição de conteúdos multimídia, e que já está sendo usado em repositórios multimídia. É um esquema de metadados que permite descrição espacial e temporal em diferentes níveis de detalhe (Doller, 2007). Existem vários programas para a anotação e recuperação em MPEG-7, como o Caliph & Emir[2] e o VideoAnnex[3];
- o AVI (Audio Video Interleaved) armazena a informação de áudio e vídeo em estruturas intercaladas, geralmente utilizados os *codecs* MPEG, o Divx e o WMV. O WMV (Windows Media Video) é atualmente a versão registrada da Microsoft do MPEG-4, e permite agregar o sistema DRM aos arquivos, ativando assim uma proteção contra cópias. Outros formatos populares são o Quicktime e o RealVideo.

Os suportes físicos para arquivos digitais podem ser magnéticos (HD,disquetes, etc), óticos (CD, DVD etc.) ou *chips*, circuitos integrados de memória (RAM, *pendrive*, cartão). Pesquisas recentes por dispositivos de armazenagem prometem mídias mais duráveis, como um sistema holográfico de registro em cristais fotorrefrativos[4]. Atualmente o arquivo digital exige a presença de um contexto tecnológico para ser acessado e essa dependência tecnológica pode leva-lo à rápida obsolescência. Como forma de evitar a degeneração da informação em meio eletrônico, faz-se necessária a criação de políticas institucionais de longo prazo.

A preservação digital é o conjunto de atividades ou processos responsáveis por garantir o acesso continuado a longo prazo à informação e ao restante patrimônio cultural existente em formatos digitais. As diferentes metodologias que foram propostas se opõem entre as que valorizam estratégias centradas na preservação do objeto físico, e as que preconizam a preservação do objeto conceitual, por meio de conversões e encapsulamento. Para Ferreira (2006),

1 http://www.mpeg.org/
2 http://www.semanticmetadata.net/
3 http://www.research.ibm.com/VideoAnnEx/
4 http://www.inphase-tech.com

a preocupação obstinada pela manutenção do arquivo original vem diminuindo à medida que aumenta a compreensão acerca dos processos informáticos, e difunde-se a idéia de que o foco da preservação não precisa estar na retenção do objeto físico, mas na conservação da experiência sensorial produzida por esse objeto, que abrange um escopo maior do que o próprio documento audiovisual. Assim, uma política de preservação deverá descrever claramente as estratégias adotadas para assegurar a preservação dos materiais em cada um dos níveis de abstração do vídeo, quais sejam, o físico, o lógico e o conceitual, e ao mesmo tempo não pode negligenciar os níveis superiores, como o social, o econômico e o organizacional.

Análise do conteúdo fílmico

No *Dicionário Teórico e Crítico do Cinema*, Jaques Aumont diz que as teorias de análise fílmica produzidas até os anos 70 carregavam principalmente um ideal estruturalista, e pesquisadores da Imagem como Raymond Bellour, Roland Barthes e Jaques Monod: "...procuravam no próprio texto, em sua estruturação e em sua ligação com as condições de sua gênese a explicação de sua forma e de sua relação com o espectador." (AUMONT, 2003). E continua explicando que somente após Cristian Metz e sua sintagmática, da linguística gerativa de Colin e Carrol, e da psicologia da montagem de Jean Mitry, é que Jean Louis Schefer recuperou uma dimensão figurativa na interpretação do filme, em oposição às tentativas de codificação vistas anteriormente, abrindo espaço para conceitos e processos originários da psicologia cognitiva, que são hoje extensamente aplicados na recuperação textual de conteúdos visuais. De fato, a partir daí, e apoiando-se na gestalt e no entendimento de aspectos linguísticos no cinema e na fotografia, chega-se mesmo a construir uma sintaxe da linguagem visual, como na proposição de Dondis (1997) e na gramática fílmica de Arijon (1991).

Em extenso trabalho sobre a análise de filmes, Tárin (2006) considera que a elaboração de uma descrição e de uma interpretação do filme são as etapas básicas, mas que devem ser acompanhadas por outras avaliações externas ao objeto estudado. Assim, inicialmente é necessário:

1) decompor o filme em seus elementos constituintes (desconstruir= descrever);
2) estabelecer relações entre tais elementos para compreender e explicar a constituição de um "todo significante" (reconstruir= interpretar).

Mas este processo se estende na inclusão de parâmetros contextuais que revelam uma situação e uma história para o produto audiovisual:

- o estudo sobre as condições técnicas de produção do filme;
- a reflexão sobre a situação econômico-político-social no momento de sua produção;
- a incorporação de princípios ordenadores, tais como gênero; estilos autorais, *star-system*, movimentos cinematográficos, etc;
- o estudo sobre a recepção do filme, tanto em seu surgimento quanto no correr dos anos;
- a utilização ou não em algum modelo de representação determinado.

Estes pontos são extremamente relevantes, pois a recuperação eficaz do conteúdo visual e sonoro só é possível com uma indexação significativa e discriminante, e que deve estar relacionada com intenções e procedimentos do usuário quando faz a consulta no ambiente real.

O conteúdo visual de imagens pode ser classificado em dois tipos principais:

- conteúdo primitivo de imagens – refere-se aos elementos básicos que compõem a imagem; são características visuais que podem ser reconhecidas e extraídas automaticamente pelo computador com reconhecimento de padrões e visão computacional. Conteúdos primitivos são em geral de natureza quantitativa;
- conteúdo complexo de imagens – refere- se aos padrões de uma imagem que são percebidos por seres humanos como fontes de significados. Dificilmente podem ser identificados por máquinas e são principalmente de natureza qualitativa.

Os índices ou metadados, sejam extraídos automaticamente ou anotados manualmente, podem ser classificados de acordo com a relação que eles têm com a imagem ou vídeo nos seguintes tipos:

- metadados independentes do conteúdo – dados que não concernem diretamente ao conteúdo da imagem ou vídeo, mas estão relacionados com este, como o formato da imagem, autoria, data, local, condições de iluminação, etc.;
- metadados dependentes do conteúdo – dados que se referem a características consideradas de nível baixo e médio, como cor, textura, forma, relações espaciais, movimento e combinações destes. Para alguns tipos de imagens, como as provenientes de satélites, da biomedicina, como tomografias, etc., é possível descrever o conteúdo destas em termos da geometria intrínseca e de configurações topológicas;
- metadados descritivos do conteúdo – dados que se referem ao conteúdo semântico e que concernem às relações das entidades da imagem com entidades do mundo real ou eventos temporais, emoções e significados associados a sinais visuais e cenas.

A maior vantagem associada com a indexação de conteúdo primitivo é que sua extração pode ser automatizada. Entretanto, este conteúdo pode não ser suficientemente rico para grande variedade de aplicações, uma vez que tipos de objetos e características significativas que podem ser reconhecidos pela máquina são ainda limitados. Em contrapartida, o conteúdo complexo da imagem é semanticamente rico, mas sua extração e indexação são custosos, uma vez que um envolvimento manual considerável é geralmente necessário.

Padrões de Indexação

A informação visual tem sido tradicionalmente produzida e conservada em suporte analógico e indexada manualmente, mas com a digitalização dos processos de captura, registro e manipulação da imagem fotográfica, hoje as bases de dados em memórias magnéticas e óticas são utilizadas para armazenar imagens e sons, juntamente com os metadados, taxonomias e tesauros associados.

As diferentes iniciativas para indexação de audiovisuais, quando definem taxonomias específicas, podem valorizar diferentes visões do problema: por um prisma implementacional, mais voltado para aspectos técnicos; ou uma aproximação conceitual, preocupada com a semântica; ou ainda uma visão contextual, que leva em conta a utilização do material. As etapas que geralmente estruturam a indexação de vídeos são as seguintes:

- segmentação do programa em cenas e planos;
- descrição de planos – identificação de elementos de conteúdo;
- descrição de cenas – localização temporal e sumário textual;
- transcrição de voz e classificação de áudio;
- descrição de metadados independentes de conteúdo.

Das inúmeras aplicações desse processo, destacamos a possibilidade da oferta de vídeo sob demanda, de forma que apenas determinado segmento do programa pode ser apresentado como resposta a uma busca, ou oferecido em um cardápio personalizado de preferências em um sistema de televisão interativa e a utilização do conceito de hipervídeo, ou seja, a navegação por meio de segmentos *hiperlinkados*. Modificações profundas na nossa relação com o audiovisual serão provocadas pelo desenvolvimento de sistemas eficientes de indexação, e afetarão boa parte das atividades humana, da cultural à produção industrial, da educação à segurança, da medicina à astronomia.

O que acontecerá mais rapidamente se houver consenso na definição de um arcabouço comum de metodologias para recuperação semântica de imagem e som, o que no entanto ainda não aconteceu. Alguns dos padrões recentemente propostos e aplicados para indexação de audiovisuais são os seguintes:

- Dublin Core –linguagem para descrição de metadados que utiliza duas classes de termos: elementos - organizados em três categorias, conteúdo, propriedade intelectual e instanciação; e qualificadores - divididos em duas classes, elementos de refinamento e esquemas de codificação. Inicialmente criado para descrição de objetos textuais, por meio de diversas extensões e acréscimos tem sido aplicado também em conteúdos audiovisuais.
- RDF – *Resource Description Framework* – tem por objetivo a definição de recursos que podem ser operados independentemente do domínio específico da aplicação, facilitando e automatizando a troca de informações entre máquinas e entre plataformas distintas. O modelo básico compõe-se de recursos, propriedades e declarações, em um sistema de classes extensível, que utiliza a sintaxe XML.
- MPEG-7 – *Multimidia Content Description Interface* – é um padrão para descrição de objetos multimídia e prevê o suporte a certo grau de interpretação semântica. Busca a interoperabilidade em recuperação, indexação, filtragem e acesso a conteúdos audiovisuais entre recursos e aplicações que manipulam esses conteúdos. Para isso utiliza descritores, esquemas de descrição e uma linguagem de definições de descritores, criados com a sintaxe XML, fornece ferramentas que permitem a gestão do conteúdo e sua descrição estrutural e conceitual, além de navegação e acesso randômico e interação com o usuário, inclusive com um histórico da utilização do sistema.
- LOM – *Learnig Objects Metadata* – estrutura metadados para objetos de aprendizagem, que são definidos como uma entidade que pode ser usada para aprendizagem e educação. Objetiva o compartilhamento e a troca desses objetos em diferentes ambientes e contextos, por meio de classificação hierárquica em categorias gerais e específicas.

RECUPERAÇÃO AUTOMÁTICA DE CONTEÚDO

Os sistemas de recuperação de imagens por conteúdo são denominados CBIR (*Content Based Image Retrieval*) ou CBVIR quando incluem o vídeo, e podem ser construídos para encontrar imagens de duas formas: a busca por exemplo, em que se utilizam como chave de busca as características visuais de uma imagem ou esboço de referência; e a busca textual, realizada a partir da transcrição de significados ou conceitos contidos na imagem que foram previamente relacionados a características visuais específicas. Na pesquisa em texto, as palavras serão procuradas na base criada a partir da análise de significados implícitos no conteúdo visual, processo denominado recuperação semântica, pois fundamentalmente diz respeito à relação entre um signo e aquilo a que ele se refere.

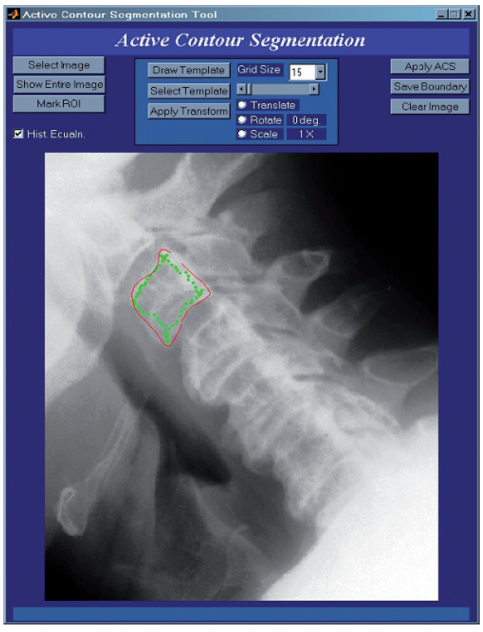
A pesquisa em CBIR é muito abrangente e envolve diversas áreas das ciências sociais e outras mais tecnológicas que contribuem com as ferramentas computacionais necessárias para determinar a informação sintática presente em imagens, especialmente as que estudam a visão artificial e o reconhecimento de padrões. Por exemplo, na *Columbia University* uma pesquisa multidisciplinar desenvolveu o Persival (*Personalized Retrieval and Summarization of Image, Video And Language resources*), sistema automático de identificação de metadados dependentes de conteúdo, especializado em imagens e gráficos úteis na medicina. Na Universidade de Winsconsin, além de ferramentas de análise, pesquisa-se um sistema videográfico autônomo, capaz de produzir vídeos informativos de qualidade simulando os métodos de operação de câmera usados por profissionais(GLEICH, 2002).

No campo da ciência da informação, a Tufts University de Boston, nos Estados Unidos, propôs a iniciativa "*Digital Library for the Humanities*"[5], suportada também por outras universidades americanas, que pretende definir novos padrões para a produção de audiovisuais. Este novo formato de mídia permitiria, aos documentos assim formatados, uma auto-atualização, fruto da interação com outros documentos eletrônicos, e também com seus usuários.

A indexação do produto audiovisual e a extração de dados relevantes apresenta desafios teóricos enfrentados por muitos autores, e é uma discussão extremamente atual diante das transformações midiáticas anteriormente apontadas. A recuperação de informações semânticas contidas em fotografias é ainda uma meta a ser alcançada, mas ao consideramos seqüências de imagens, o problema de indexação torna-se muito mais desafiador, pois envolve a identificação e o entendimento de cenas longas e complexas, compostas por centenas de imagens.

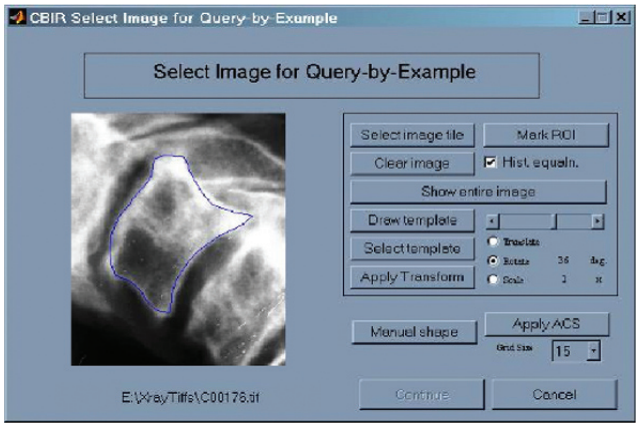
O primeiro passo é o desenvolvimento de sistemas de reconhecimento de padrões capazes de dividir as seqüências de imagens em unidades menores, porém significativas, no processo denominado segmentação de vídeo. É importante considerar a detecção de determinados eventos marcantes, como o instante em que o predador ataca uma presa, a ação eletroquímica em áreas do cérebro, a colisão de veículos e outros registros de curtíssima duração. E finalmente aplicar técnicas de reconhecimento visual, já utilizadas extensivamente em imagens técnicas não-fotográficas, como tomografias e ecografias, como no exemplo mostrado na figura 1, que apresenta imagens radiográficas da coluna vertebral.

Tela 1

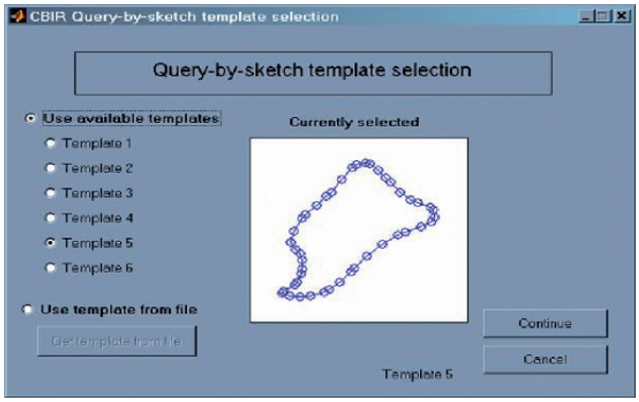


5 http://www.perseus.tufts.edu/

Tela 2



Tela 3



Tela 4

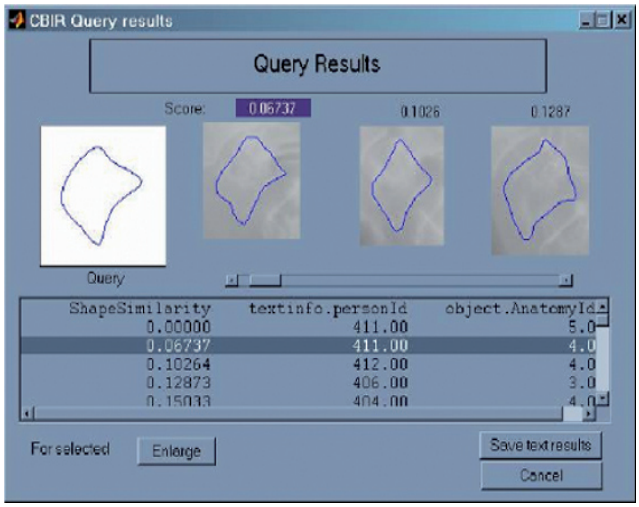


Figura 1 - Reconhecimento em imagens médicas.
Tela 1: Segmentação da imagem; Tela 2: Faz a busca por exemplo;
Telas 3 e 4 permitem a busca por esboço.

Grande número de centros de pesquisa e de empresas estão envolvidos com esta questão, percebida como urgente pelos grandes distribuidores mundiais de audiovisual. Em pesquisa que envolve a Sharp, a Phillips, a Microsoft e a AT&T,

desenvolvida nas universidades de Berkley e de Illinois, são explorados quatro processos coordenados: extração de elementos, análise de estruturas, abstração e indexação, para a obtenção de um sistema automático de segmentação e identificação de conteúdos em qualquer tipo de vídeo (DIMITROVA, 2002).

Na Holanda, o projeto DMW[6](*Digital Media Warehouses*) trouxe muitos avanços na modelagem multimodal e inteligente para o reconhecimento de padrões em vídeos e a identificação de conceitos sobre eventos e objetos. Os pesquisadores do DMW definiram soluções lógicas e físicas, além de padrões para aquisição e indexação do produto multimidiático, que formam uma arquitetura integrada para armazenagem de metadados acoplada a uma linguagem de consulta de alto nível.

No Brasil, a Universidade Federal de São Carlos desenvolve o sistema SisRMi-CN (Serrano, 2003), um ambiente para a criação e gestão de aplicações multimídia, oferecendo diferentes formas de recuperação de informações, usando lógica nebulosa. E na UFMG, o Núcleo de Processamento Digital da Imagem vem apresentando pesquisas consistentes na área, oferecendo *workshops* e implementando programas de preservação junto ao Patrimônio Histórico do Estado de Minas Gerais (ARAÚJO, 2003).

Antologias, catálogos, resenhas e inúmeras outras fontes de informação sobre filmes, vídeos e programas televisivos são publicados regularmente para suprir as necessidades estratégicas de uma indústria cultural cada vez mais profícua. À medida que redes de televisão vão ocupando os canais na Internet e integrando-os em um sistema mundial multimídia, os mecanismos de classificação, busca e indexação de programas e eventos tornam-se serviços essenciais. Algumas das grandes difusoras mundiais de TV e rádio já permitem o acesso livre à parte de seus arquivos de programas. No Brasil, os distribuidores de mídia principais são a All TV, a RedeTV, a TV Vírgula, o Globo Media Center, o canal do portal Terra, a TV Cultura, e também a InterneTV. E à semelhança do YouTube, filmes curtos podem ser vistos no PortaCurtas e no CurtaoCurta. No entanto, as descrições de conteúdo são obtidas em bancos de dados manualmente indexados e resumem-se a curta sinopse e comentários ou críticas de usuários. O Internet Movie Database[7] é provavelmente o mais completo registro da produção cinematográfica disponível atualmente na rede.

Identificação de elementos visuais

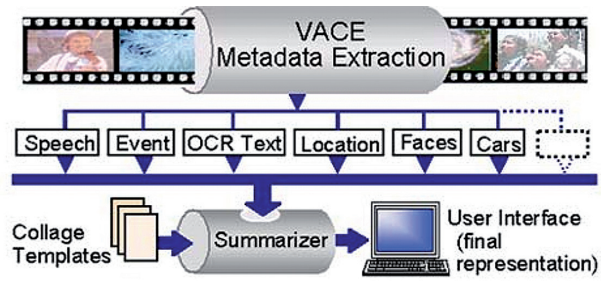
Muitos sistemas de CBIR utilizam a similaridade de características como formas, bordas, cores ou textura para criar índices, o que tem produzido bons resultados, quando utilizado para imagens em movimento. O projeto Informedia[8], da Universidade Carnagie Mellon, foi pioneiro na área ao utilizar estas técnicas e a segmentação do vídeo para indexar programas de notícias em tempo real. A figura 2 mostra uma tela da interface de busca textual em noticiários, e a figura 3 apresenta as etapas de processamento do sistema Informedia, evidenciando os diferentes domínios em que são extraídos os metadados.

Figura 2

Projeto Informedia – Tela do aplicativo de busca e anotação supervisionada

Figura 3

Projeto Informedia – Esquema conceitual do sistema



Outras linhas de pesquisa experimentam o levantamento de gráficos estatísticos de características dinâmicas no vídeo. No trabalho de Guimarães (2003), o fluxo de vídeo é transformado em fatias espaço-temporais por amostragem dos pixels que formam as imagens. Cada quadro 2-D é transformado em uma linha vertical. O gráfico resultante representa o ritmo visual de um vídeo, e estas fatias podem indicar os pontos de transição, onde há grandes mudanças no conteúdo da imagem.

Os sistemas mais reportados para extração de metadados dependentes de conteúdo são os que reconhecem semelhanças entre características visuais, onde usualmente temos como

6 http://monetdb.cwi.nl/acoi/DMW/index.html
7 http://imdb.org
8 http://www.informedia.cs.cmu.edu/

base de pesquisa a similaridade de cores, de formas ou de texturas, ou uma combinação destes parâmetros:

A *cor* é uma das características mais utilizadas pelos seres humanos para reconhecimento e discriminação visual. A aparência de uma cor em objetos do mundo real geralmente é alterada pela textura da superfície, pela iluminação e sombra de outros objetos, e pelas condições de observação e captura. As operações de reconhecimento de similaridades permitem encontrar as seguintes imagens: que contêm uma cor especificada por meio de proporções aditivas; cujas cores são próximas daquelas de uma imagem exemplo; que contêm regiões coloridas como especificado em esboço; que contêm um objeto conhecido com base nas propriedades de composição espectral. A extração de cores automatizada ainda não é capaz de fazer referências ao contexto, o que dificulta a distinção entre uma informação de cor do objeto e de uma alteração cromática introduzida pelo ambiente.

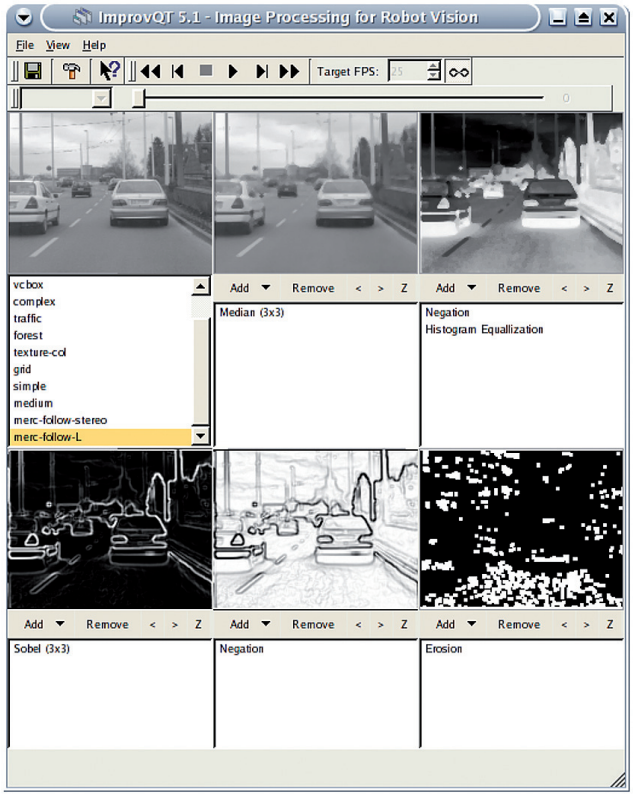
A percepção da *textura* é um fator importante da visão humana, pois ajuda a identificar em uma cena a profundidade e orientação das superfícies, além de revelar suas características tácteis. A textura refere-se a um padrão visual que tem algumas propriedades de homogeneidade que não resultam simplesmente da cor ou da incidência da luz, como a repetição de linhas e as características físicas superficiais dos objetos. Pela extração de características de textura obtém-se um descritor importante para indexar imagens da natureza, e muito útil nas pesquisas em grandes repositórios de imagens .

Forma é um critério que permite identificar na projeção bidimensional parte da estrutura física dos objetos. Para aplicações de recuperação, as características da forma podem ser consideradas como sendo globais ou locais. Características globais são propriedades derivadas da forma inteira, como simetria, circularidade, localização de eixos, etc. Características locais são aquelas derivadas do processamento parcial da forma, incluindo tamanho e orientação de segmentos consecutivos de bordas, pontos de curvaturas e ângulos de curvas. As características de forma podem também ser classificadas em parâmetros internos, que descrevem a região envolvida pelo contorno do objeto, e parâmetros externos, que descrevem as bordas do objeto. Na figura 4 pode-se ver a filtragem de bordas nas imagens de uma aplicação CBIR que usa a forma para a identificação de veículos em trânsito.

Para indexar imagens fixas extraindo-se os metadados dependentes do conteúdo (cor, textura, forma), pode ser necessário pré-calcular para cada imagem um conjunto de características distintivas, e então as consultas são expressas como comparações com exemplos visuais. Para começar a consulta, o usuário seleciona as características que são relevantes e define uma medida de similaridade. Os exemplos tanto podem ser esboços preparados pelo usuário (com ajuda de um programa de desenho) quanto imagens selecionadas em um banco de dados, dentre amostras preparadas. O sistema verifica a similaridade entre o conteúdo da imagem usada na consulta e das imagens da base de dados. Como nem sempre os resultados obtidos em resposta a uma consulta são plenamente satisfatórios, em geral procura-se melhorar este resultado com uma metodologia em que se mantém o número de objetos não encontrados o mais baixo possível, às custas de um número mais alto de falsas respostas.

Figura 4

Filtragem de imagens para o reconhecimento de formas no programa ImprovQT



As técnicas de reconhecimento aplicadas em imagens fixas são empregadas nos quadros obtidos na fase de segmentação do vídeo, apresentada a seguir.

Segmentação do vídeo

Métodos estruturados de representação compacta do conteúdo dos produtos audiovisuais têm sido desenvolvidos com objetivo de facilitar o acesso ao vídeo não só para navegação por imagens relevantes, ou quadros-chave, como para a recuperação e pesquisa via texto. Tais métodos buscam modelar os dados de forma que todas as informações estejam disponíveis de maneira clara e rápida para os usuários que a estão requisitando, além de tornar transparentes as informações pertinentes sobre os dados (metadados).

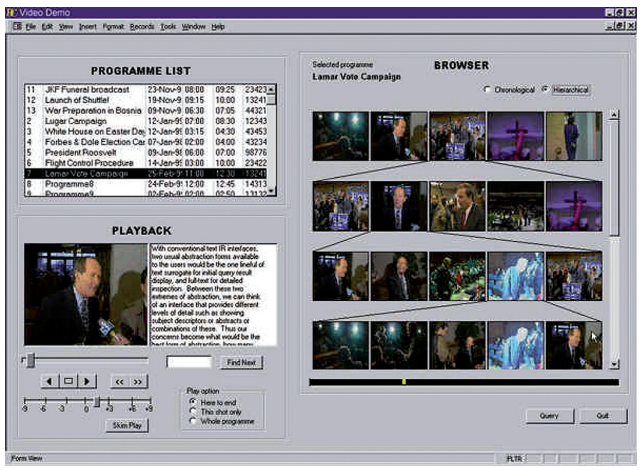
O desafio para a indexação e recuperação de imagens orientadas pelo conteúdo está no desenvolvimento de mecanismos automáticos e precisos, porém genéricos e abrangentes. Uma possibilidade é começar com a extração de conteúdo primitivo e subseqüentemente fazer uso de regras de conhecimento e aprendizado sobre a informação contextual relevante, permitindo uma identificação, ou inferência automática, do conteúdo significativo que uma pessoa observaria em uma fotografia ou vídeo. Essa solução é adequada em algumas situações específicas.

No noticiário televisivo típico, por exemplo, podem-se considerar algumas características particulares importantes que facilitam o trabalho de indexação automática: a trilha sonora é composta na maior parte por falas de entrevistados ou narradores; usam-se extensivamente as legendas e letreiros; e a imagem é freqüentemente uma ilustração do tema tratado.

O vídeo digital é uma apresentação de eventos dinâmicos que possuem imagens, sons, textos e gráficos, uma estrutura complexa que pode ser dividida em partes mais simples. O problema da segmentarão em vídeo começa na identificação dos momentos de mudança radical do conteúdo visual, nos cortes de montagem e na seqüência de quadros estáticos do filme. A abordagem clássica para resolver este problema é baseada no cálculo de medidas de dissimilaridade ou diferenças entre os quadros. Em novas abordagens, a segmentação em vídeo é transformada em um problema de detecção de padrões, no qual cada evento de vídeo é visto como padrões em um imagem espaço-temporal 2D, que constituem um ritmo visual. Nesse caso, são utilizadas basicamente ferramentas morfológicas e topológicas com o objetivo de identificar os padrões específicos que são relacionados a eventos do vídeo, como cortes, *fades*, *dissolves*, *flashings* e outros.

Figura 5

Identificação de planos

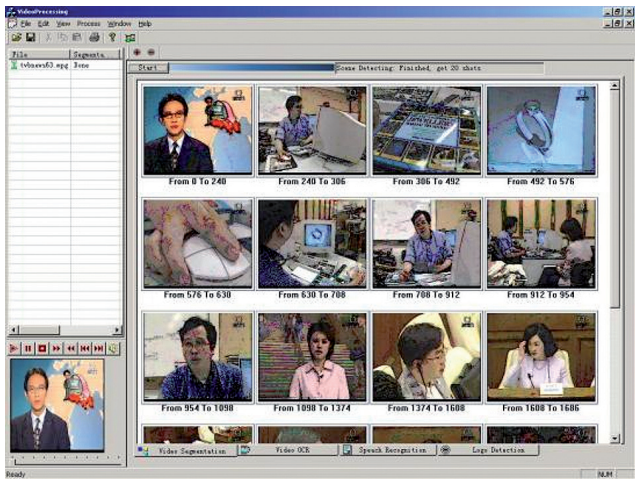


Na segmentação do vídeo, a unidade fundamental é o PLANO, que é capturado a partir de uma operação contínua da câmera. O plano é constituído por uma seqüência ininterrupta de QUADROS ou fotogramas gerados pela câmera, e pode ser uma imagem estática no tempo ou mostrar tanto o movimento produzido pela própria câmera, como por exemplo, *zoom* ou panorâmica, quanto o realizado por objetos da cena.

Uma CENA é usualmente composta de número pequeno de planos seqüenciais unificados pela posição temporal ou características similares. Enquanto o plano é uma unidade física do vídeo, a cena representa uma unidade semântica do mesmo, possuindo algum significado intrínseco. O processo de identificação destas unidades, a segmentação do vídeo, começa por determinar os limites (início e fim) dos planos e cenas. (figura 5).

Figura 6

Um navegador visual de vídeos, exibindo os quadros-chave, e a duração das cenas identificadas.



Uma cena é um agrupamento de planos, que por sua vez são constituídos por seqüências de quadros. Por ser grande a quantidade de cenas contidas em um filme, e para facilitar a representação, os planos podem ser sintetizados e apresentados de forma resumida, por meio de quadros selecionados que representam o conteúdo da cena, e são chamados de QUADROS-CHAVE.

Quadros-chave são um ou mais quadros que representam todo o conteúdo visual de um plano da maneira mais aproximada possível. Técnicas para a extração de quadros-chave lidam com os limites do plano (isolamento dos quadros inicial e final), e com padrões de conteúdo visual (a ocorrência de determinado elemento, ou em agrupamentos de elementos distintos). A taxa de amostragem pode variar em função do grau de precisão desejado.

Para a identificação das cenas pode-se recorrer à similaridade visual e a medidas de proximidade temporal. A similaridade visual pode ser medida com a análise de histogramas (gráfico da distribuição de pixels), de estatísticas de fluxo ótico, e da localização de elementos visuais recorrentes. A figura 6 apresenta uma tela de aplicativo para seleção de quadros-chave e identificação de cenas, desenvolvido na Universidade Chinesa de Hong-Kong.

A IBM, em sua unidade de Watson, Califórnia, tem se dedicado especialmente à pesquisa em recuperação de informação audiovisual e desenvolvido inúmeros projetos relacionados, a começar por um ambiente de programação sofisticado para pesquisa semântica, o UIMA (Unstructured Information Management Architecture), que permite a integração com programas em JAVA de ferramentas de análise capazes de descobrir significados, relações e fatos mediante análise de documentos de texto, imagens, *e-mail*, áudio e vídeo. A indexação de imagens e vídeo é feita com o sistema denominado MARS[9] (Multimedia Analysis and Retrieval System), que usa técnicas de inteligência artificial para inferir conceitos semânticos a partir de uma biblioteca de modelos. Assim é possível a busca com base no conteúdo primário, a partir de similaridade em cores, texturas e formas, por conceitos semânticos que descrevem cenas, objetos e eventos. O sistema também faz a pesquisa por metadados e textos contidos na imagem, além da transcrição de diálogos e identificação de gênero musical. Ainda limitado em sua biblioteca de conceitos, com o MARS é possível extrair elementos como prédios ou o tipo de cenário (p. ex. praia, neve, mar, céu), e transcrever locuções treinadas.

Recuperação da informação em áudio

Para o reconhecimento e representação do conteúdo sonoro deve-se inicialmente discriminar os três níveis que costumam compor a trilha sonora de filmes; o nível da fala, sejam diálogos ou narrações; o de ruídos; e a trilha musical, quando houver. O segundo e terceiro níveis também contêm informações semânticas, porém as pesquisas se concentram na identificação de palavras faladas, portanto na extração imediata de significantes lingüísticos. O reconhecimento da fala ou ASR (*Automatic Speech Recognition*) consiste em discriminar fonemas, sílabas e palavras para recuperar uma mensagem de voz, e geralmente acontece em três etapas:

1) aquisição do sinal de voz – a simples transformação do sinal mecânico em sinal elétrico, feita normalmente por microfones, conectados a uma placa de captura de som, e acionada por um *software* de gravação;

2) extração paramétrica – filtragem, quantização e preparação do sinal digital, através de *software* de edição e tratamento de sons;

3) reconhecimento de padrões – a identificação de palavras e frases na representação matemática discreta de sinais contínuos. Algumas das técnicas de processamento digital de sinais usadas são Codificação Preditiva Linear, baseada na diferença entre os tipos de sons; Modelo de Mistura Gaussiano, baseada em classes vocais individualizadas; Transformada Rápida de Fourier (FFT), modelagem do sinal de palavras isoladas.

O reconhecimento de voz é feito por um algoritmo capaz de segmentar o áudio em pequenos trechos que isolam os fonemas. A transcrição é específica para cada língua e cada som individual pode ser identificado e comparado a uma lista previamente construída de palavras ou frases. Existem basicamente dois tipos de transcrição da voz humana: o primeiro permite ativar comandos predefinidos, como "Negrito" ou "Abrir programa", com a fala de um usuário específico. Os do segundo tipo são os chamados programas de ditado, que permitem transcrever textos. Estes podem ser dependentes de locutor, do qual se exige treino prévio e que são comuns hoje em dia, ou independentes de locutor, sistemas ainda em desenvolvimento e que apresentam grandes desafios na sua implementação (NETO, 2006).

Para a classificação da informação musical foram reportados resultados consistentes na busca, por exemplo, ou QBE (Query by Example), método por comparação, capaz de identificar diversos gêneros musicais. Uma segunda linha de pesquisa trabalha no reconhecimento de ritmos e melodias para permitir a busca por solfejo, ou QBH (Query by Humming). Ambos os processos são de interesse especial para a telefonia móvel(AHMAD, 2006).

Métodos de inteligência artificial

Com a consolidação da Internet, o tratamento da informação modifica-se e busca alternativas para uma nova ordem de catalogação e pesquisa, conseqüentemente revolucionando os métodos tradicionais de difusão do conhecimento. Novas práticas impõem a redefinição dos *Gêneros de Informação*, observando-se a nova demanda marcada pela produção multimídia, que absorve múltiplos formatos e subverte as categorias tradicionais que distinguem os tipos de informação; e da noção de *Campos de Informação*, pois o processo de dividir grupos de informações por temas já não corresponde ao potencial da rede, que nos permite navegar de uma informação à outra, correlata, e consolida uma grande infoteca sem divisões rígidas e que facilita a pesquisa interdisciplinar e sem fronteiras; além da flexibilização do conceito de *Agentes da Informação*, a distinção entre emissor e receptor se torna ambígua com a enorme interatividade permitida pela rede, pois surgem os co-autores e os co-leitores. E também se transforma, é claro, o processo de criação, pois a obra agora pode ser modificada por aquele

9 http://www.alphaworks.ibm.com/tech/imars

que a usufrui e que com ela interage em diversos níveis. A par destas mudanças paradigmáticas, o acesso aleatório à informação sugere que o acaso pode se tornar importante ferramenta de pesquisa para obtenção de soluções, ajudando na correção de erros e mesmo na otimização dos resultados da pesquisa.

O equilíbrio entre a revocação e a relevância em procedimentos de busca e a criação de sistemas de reconhecimento de padrões capazes de interpretar aquilo que identificam são questões críticas das tecnologias da recuperação da informação, atualmente. As soluções podem estar no desenvolvimento de máquinas inteligentes, aptas à compreensão de conteúdos e observação dos contextos, e também na criação de interfaces instrutivas entre usuários e sistemas. Algumas ferramentas de inteligência artificial utilizadas atualmente para isto são as seguintes:

- **Redes Neurais –** uma classe de modelagem de prognóstico realizado por ajuste repetido de parâmetro. A rede neural consiste em um número de elementos interconectados e organizados em camadas, que aprendem pela modificação da conexão, criando vínculos entre as diversas camadas.

- **Modelos de Markov –**representações matemáticas utilizadas para prever comportamentos de um sinal através de uma seqüência de observações. Em uma cadeia de Markov supõe-se uma fonte gerando tais saídas observáveis, denominada Fonte de Markov. Os símbolos gerados a partir dessa fonte são dependentes apenas de observações anteriores, as quais foram geradas da mesma forma e assim sucessivamente. O número de seqüências anteriores consideradas para gerar uma saída é conhecido como ordem da Cadeia de Markov. Cada estado de uma cadeia de Markov representa uma observação/símbolo de um evento físico correspondente, o que permite computar, a partir de uma dada seqüência de símbolos, quais foram os estados que geraram tal seqüência. No Modelo Escondido de Markov (MEM) cada estado representa uma probabilidade, de certa forma "escondida" no conjunto dos símbolos que está representado. Um MEM, portanto, possibilita computar a seqüência de estados com maior probabilidade de ter gerado o conjunto observado de símbolos do estado corrente.

- **Lógica Nebulosa –** também chamada *fuzzy*, é um algoritmo que permite simular um aspecto do raciocínio humano, a habilidade de tomar decisões racionais em condições de incerteza e imprecisão. Ao manipular inteligentemente informações imprecisas e conceitos indefinidos, pode inferir uma resposta precisa para um problema cujo enunciado é inexato e incorporar tanto o conhecimento objetivo quanto o conhecimento subjetivo.

- **Algoritmos Genéticos –** desenvolvidos a partir dos princípios da evolução das espécies de Darwin, e em leis e procedimentos da genética. A partir de uma população de indivíduos, representados por cromossomas (palavras binárias), cada um associado a uma aptidão (funções), que são submetidos a um processo de evolução (seleção, reprodução, cruzamento e mutação), repetido em vários ciclos em direção à sobrevivência dos mais bem adaptados.

Para Sims (1991), podemos desenvolver modelos procedurais a partir da seleção interativa com humanos, levando o sistema ao aprendizado das estratégias de preferências do usuário e da lógica de sobrevivência dos resultados mais complexos e interessantes. Significa que poderemos ter o auxílio da inteligência artificial não somente para aplicações de reconhecimento de padrões, indexação e busca, mas também para a própria estruturação e modelagem do sistema e obtenção de modelos de indexação semântica adaptáveis a diferentes domínios.

CONCLUSÕES

A preservação de documentos garante a análise histórica e é fundamental como política de consolidação de uma identidade nacional e planetária, mas diante de gigantesca massa documental, muitas são as dificuldades que se apresentam. A tecnologia digital pode nos ajudar a resolvê-las, porém as soluções não são triviais e exigirão muitos anos de pesquisa e desenvolvimento. A fragilidade dos meios e a inovação contínua de processos e padrões são grandes desafios que devem ser encarados por iniciativas integradoras de longo prazo, que sustentem a conservação e o acesso futuro ao que estamos produzindo hoje em suporte eletrônico.

Verificamos que a recuperação de conteúdos em audiovisuais vem obtendo sucesso especialmente na área de reconhecimento de padrões e na identificação de imagens de cunho técnico, porém a pesquisa pela decodificação semântica de imagens, a extração automática de metadados descritivos está apenas começando, e faz parte da criação da máquina ideal, semelhante a nós mesmos. No caso da visão, provavelmente melhor em certos aspectos, não somente pela capacidade ampliada de perceber outras freqüências luminosas, mas também na possibilidade de analisar maior quantidade de informação visual, uma vez que estejam maduros os sistemas de recuperação da informação audiovisual que foram aqui brevemente examinados.

Artigo recebido em 21/08/2007
e aceito para publicação em 16/05/2008

REFERÊNCIAS

AHMAD, I. et al. *Audio-based queries for video retrieval over java enabled mobile devices.* [S.l.]: Nokia Corporation, 2006. Disponível em: <http://muvis.cs.tut.fi/Documents/SPIE_05_06.pdf>. Acesso em: 2008.

ARAÚJO, A. de A. *RIBC recuperação de informação com base no conteúdo visual.* Belo Horizonte: Universidade Federal de Minas Gerais, 2003.

ARIJON, Daniel. *Grammar of the film language*. Los Angeles, CA: Silman-James Press, 1991.

AUMONT, J.; MARIE, M. Dicionário teórico e crítico do cinema. Campinas, SP: Papirus Editora, 2003.

BEAGRIE l.; GREENSTEIN, D. *A strategic framework for creating and preserving digital collections*. London, UK: UK's Arts and Humanities Data Service, 1998. Disponível em: <http://ahds.ac.uk/strategic.pdf>. Acesso em: 2008.

DIMITROVA, N. et al. Applications of video-content analysis and retrieval. IEEE *MultiMedia,* v. 9/3, p. 42–56, 20-?. Disponível em: <http://2002.csdl.computer.org/comp/mags/mu/2002/03/u3toc.htm>. Acesso em: 2008.

DOLLER, M.; LEFIN, N.; KOSCH, H. *Evaluation of available MPEG-7 annotation tools*. Germany: University of Passau, 2007. Disponível em: <https://www.i-know.at/content/download/870/3615/file/D%C3%B6ller.pdf>. Acesso em: 2008.

DONDIS, D. A. *Sintaxe da linguagem visual.* São Paulo: Martins Fontes, 1997.

FERREIRA, M. *Introdução à preservação digital:* conceitos, estratégias e actuais consensos. Minho: Editora Escola de Engenharia da Universidade do Minho, 2006. Disponível em: <https://repositorium.sdum.uminho.pt/bitstream/1822/5820/1/livro.pdf>. Acesso em: 2008.

GLEICHER, M.; HECK, R.; WALLICK, M. *A framework for virtual videography.* 2002. Disponível em: <http://www.cs.wisc.edu/graphics/Papers/Gleicher/Video/smartgraphics-2002.pdf>. Acesso em: 2008.

GUIMARÃES, S. J. F. *Video transition identification based on 2D image analysis*. 2003. Tese (Doutorado)- Departamento de Ciência e de Computação, UFMG, 2003.

KOMPATSIARIS, Y. *Multimedia semantic analysis technologies and their potential uses*. 2006. Disponível em: <http://www.samt2006.org/presentations/ITI_MM%20Analysis.pdf>. Acesso em: 2008.

NETO, N.; SILVA, E.; SOUSA, E. Software usando reconhecimento e síntese de voz: o estado da arte para o português brasileiro. In: 2005 LATIN AMERICAN CONFERENCE ON HUMAN-COMPUTER INTERACTION, 2005, México. *Electronic proceedings…* Disponível em: <http://doi.acm.org/10.1145/1111360.1111396>. Acesso em: 2008.

SERRANO, M. *Um sistema de recomendação para mídias baseado em conteúdo nebuloso.* 2003. Dissertação (Mestrado)- UFSCar, São Paulo, 2003.

SIMS, Karl. Artificial evolution for computer graphics. In: SIGGRAPH '91, 1991. *Proceedings...* [S.l.: s.n.], 1991.

TARÍN, F. J. G. *El análisis del texto fílmico.* [S.l.]: Biblioteca Central da Universidade da Beira Interior, 2006. Disponível em: <http://www.recensio.ubi.pt/modelos/documentos/documento.php3?coddoc=1597>.Acesso em: 2008.

VIRILIO, Paul. *A máquina de visão.* Rio de Janeiro: José Olympio, 1994.

URL DAS FIGURAS

Figura 1 - http://archive.nlm.nih.gov/pubs/long/spie-sd2003/spie-sd2003.php

Figuras 2 e 3 - http://www.informedia.cs.cmu.edu/dli2/

Figura 4 - http://www.ee.uwa.edu.au/~braunl/improv/

Figura 5 - http://www.irishscientist.ie/DCUAS125.htm

Figura 6 - http://www.2002.org/CDROM/alternat/XS3/image006.jpg