

UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação
ISSN 0103-2569

**Projeto Video Family Archive: Descrições Semânticas de
Objetos de Mídia**

**Rudinei Goularte
Edson dos Santos Moreira**

Nº 179

RELATÓRIOS TÉCNICOS



São Carlos – SP
Dez/2002

SYSNO	1287259
DATA	/ /
ICMC - SBAB	

Projeto *Video Family Archive*: Descrições Semânticas de Objetos de Mídia

Rudinei Goularte, Edson dos Santos Moreira
{rudinei, edson}@icmc.sc.usp.br
Departamento de Computação e Estatística
Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo – Campus de São Carlos
Cx. Postal 668, 13566-570 São Carlos, SP

Resumo

Este relatório descreve o desenvolvimento de uma técnica para associação de informações contextuais a vídeos. Além disso, expõe como tal técnica pode ser aplicada ao domínio de vídeo interativo através do desenvolvimento de uma ferramenta de anotações contextuais e de busca. Tais ferramentas agem sobre um conjunto de objetos de vídeo codificados de acordo com o padrão MPEG-4. A técnica para associação é baseada no padrão MPEG-7 e propõe um método alternativo para integrar descrições MPEG-7 a conteúdo MPEG-4.

1. Introdução

Com a disseminação da tecnologia de vídeo interativo, torna-se possível um grande número de aplicações e formas de interação, permitindo: interação com os objetos sendo apresentados; buscas complexas, na Web ou baseadas em objetos (em complemento às baseadas em texto); e reações automáticas às ações do usuário. As pesquisas na área de vídeo interativo têm produzido padrões, técnicas e ferramentas para criar, distribuir e apresentar conteúdo interativo [ISO 2001a], e para fornecer semântica a esse conteúdo [ISO 2001b]. Tais padrões, juntamente com o crescente interesse das indústrias de comunicação e de entretenimento, têm fornecido as bases para o desenvolvimento da área de TV Interativa como uma das principais aplicações de vídeo interativo.

Conforme a computação vai se tornando mais penetrante e ubíqua os usuários têm a expectativa de interagir com aplicações e serviços em qualquer lugar, a qualquer momento, e independente do dispositivo que estejam usando. Essa expectativa é ainda maior em relação a aplicações de Vídeo Interativo. Para atender tal demanda são necessárias melhorias tanto nas formas de interação quanto nas estruturas de suporte, que são um dos principais objetivos de uma ativa área de pesquisa conhecida como Computação Ubíqua [Abowd 2002]. Uma das divisões da Computação Ubíqua, conhecida como Consciência de Contexto, explora a comunicação usuário-computador durante interações de forma a aproveitar informações contextuais (presentes na comunicação), fornecendo subsídios para a diminuição do esforço do usuário durante interações, e para a adaptação de serviços de acordo com as necessidades (ou desejos) atuais dos usuários [Abowd 1999].

Contudo, a literatura mostra que os trabalhos sobre vídeo interativo não utilizam consciência de contexto (com raras exceções, como [Faria 2001]), e que carecem de flexibilidade nas estruturas que representam e relacionam os objetos de mídia que compõem os programas [TV-Anytime 2001, Hjelsvold 2001].

Este relatório visa descrever uma técnica de associação de informações contextuais a vídeos caseiros, a qual foi desenvolvida durante um estágio no Gerogia Tech Institute, em Atlanta (EUA), no âmbito de um projeto de cooperação internacional (NSF/CNPq). A técnica está baseada em descrever os aspectos técnicos, semânticos e contextuais dos vídeos. Tais descrições seguem o padrão MPEG-7 [ISO 2001b] e servem como indexadores para características dos vídeos. Tais características interessam tanto a autores de apresentações (como formato do arquivo de vídeo, taxas de quadros por segundo, etc.) quanto a usuários comuns (“o aniversário do meu irmão”, “todos os natais da família na década de 80”, “a cena onde aparece o carro preto”, etc.).

Em fase posterior ao estágio, os vídeos caseiros (em formato AVI) foram convertidos para o formato de arquivo do padrão MPEG-4 (MP4), possibilitando a adição de conteúdo interativo aos vídeos originais. A mesma técnica de associação de descrições foi aplicada aos vídeos em formato MP4, especificando-se uma forma de associar descrições MPEG-7 a conteúdo MPEG-4, uma vez que a forma de se fazer tal associação ainda não está definida dentro dos comitês MPEG.

As seções deste relatório estão divididas da seguinte forma: a seção 2 apresenta, brevemente, as principais tecnologias utilizadas no projeto – MPEG-4, MPEG-7 e consciência de contexto; a seção 3 descreve o projeto “Video Family Archive”, o qual utiliza as tecnologias descritas na seção 2 no fornecimento de ferramentas para anotações padronizadas

de características semânticas de vídeos; a seção 4 apresenta algumas considerações finais e as conclusões resultantes deste projeto.

2. Tecnologias Relacionadas

2.1 MPEG-4

O padrão MPEG-4 [ISO 2001a] foi desenvolvido para fornecer soluções para as novas aplicações multimídia através de características como: composição de apresentações formadas por objetos multimídia; *streaming*; recuperação de erros; alta taxa de compressão; e sincronização. A funcionalidade desse padrão é baseada na interação com objetos (vídeo, áudio, gráficos 2D e 3D, animações 2D e 3D, texto, etc.) de uma apresentação. Uma possível lista de interações pode ser: mudar a posição de objetos; mudar a escala; rotacionar objetos; mudar a velocidade; adicionar ou remover objetos que compõem a apresentação [Battista 1999, ISO 2001a].

A figura 1 mostra o processo de codificação, multiplexação e streaming de objetos MPEG-4: arquivos de mídia (como vídeos AVI) e uma descrição textual da composição da apresentação [ISO 2001a] são enviados ao codificador. O codificador gera, como saída, objetos audiovisuais MPEG-4 primitivos na forma de *streams* MPEG-4 elementares comprimidos (existem vários tipos de taxas de bits e de algoritmos de codificação de modo a cobrir um amplo domínio de aplicações). Tais *streams* podem ser armazenados ou enviados ao multiplexador que gera como saída um único *stream* MPEG-4, o qual pode ser transmitido via rede.

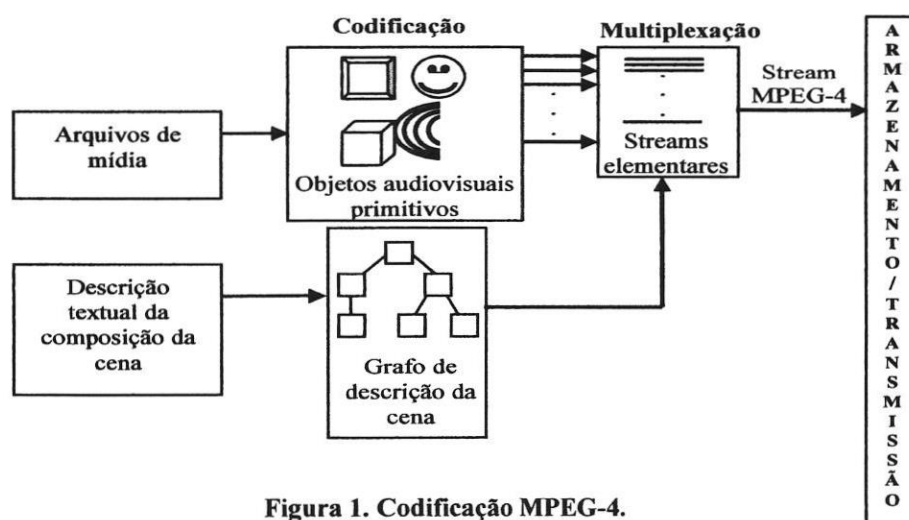


Figura 1. Codificação MPEG-4.

Uma característica do padrão que vale ser ressaltada é que um sinal de vídeo MPEG-4 pode ser codificado em uma ou mais camadas com prioridades diferentes. A camada com prioridade mais alta, a camada base, fornece o nível básico de qualidade. Para refinar incrementalmente essa qualidade, camadas de aprimoramento com prioridades progressivamente mais baixas são decodificadas. As camadas de aprimoramento melhoram a qualidade da camada básica em termos de resolução e de taxa de quadros, fornecendo escalabilidade espacial e temporal [Ebrahimi 2000].

2.2 MPEG-7

O MPEG-7 [ISO 2001b] tem como objetivo padronizar a descrição de dados com conteúdo multimídia. Isso é feito através da especificação de um conjunto padrão de descritores (*Descriptors –D*), os quais podem ser usados para descrever uma variedade de tipos de informação multimídia. O padrão MPEG-7 especifica estruturas pré-definidas de descritores e seus relacionamentos, assim como formas de o usuário especificar suas próprias estruturas. Tais estruturas são chamadas de esquemas de descrição (*Description Schemes –DS*), e a definição de novos esquemas de descrição é feita usando uma linguagem específica para isso, chamada linguagem de definição de descrições (*Description Definition Language –DDL*) [ISO 2001b, Nack 1999]. A DDL é baseada na linguagem XML Schema [W3C 2001a], contendo as partes estruturais e de tipos de dados da XML Schema (*XML Schema structural language* e *XML Schema datatype language*), e adicionando algumas extensões relativas aos tipos de dados *arrays* e matrizes, e a tempo (duração e ponto no tempo) [ISO 2001b].

É importante observar que o padrão MPEG-7 não tem o objetivo de produzir formatos de codificação, mecanismos de busca ou ferramentas de autoria. Também é importante observar que a DDL MPEG-7 não é uma linguagem de modelagem, mas uma linguagem de esquemas para representar o resultado de uma modelagem de dados audiovisuais [ISO 2001b, Nack 1999]. Descrições MPEG-7 não estão limitadas a objetos multimídia. Fotos em papel, documentos textuais, animais, ou mesmo abstrações, como um programa de TV, podem também ser descritos.

Também se deve deixar claro que o padrão MPEG-7 é independente dos outros padrões MPEG (como o MPEG-4), podendo ser utilizado para descrever objetos em qualquer formato, como AVI, MOV, WAV, SMIL, etc.

2.3 Consciência de Contexto

De forma geral, interação vem sendo usada para melhorar interfaces entre usuários e sistemas. Consciência de contexto é um avanço nessa área uma vez que fornece aos computadores informações contextuais sobre a comunicação entre o usuário e a aplicação, permitindo aos sistemas executarem tarefas automaticamente [Abowd 1999].

Na área de computação ubíqua uma das definições de contexto e consciência de contexto mais amplamente utilizada é a de Abowd [Abowd 1999]. Tal definição cobre todos os requisitos, para computação consciente de contexto, apresentados em definições anteriores: apresentação de informações e serviços aos usuários; execução automática de serviços; associação de informação contextual para posterior recuperação. Segundo a definição de Abowd: *“Contexto é qualquer informação usada para definir a situação de uma entidade. Contexto (primário) pode ser categorizado em: localização, identidade, atividade e tempo. Uma entidade pode ser uma pessoa, um lugar ou um objeto relevantes para a interação entre o usuário e a aplicação. Então, um sistema é consciente de contexto se ele usa o contexto para fornecer informação ou serviço relevantes para o usuário (a relevância depende da tarefa que o usuário deseja realizar)”*.

A tarefa de definir qual informação é importante ou relevante é facilitada utilizando-se algumas diretrizes: identidade (quem é o usuário – **who**), localização (onde o usuário está – **where**), atividade (o que o usuário está fazendo ou deseja fazer – **what**, e como isso é feito – **how**), e tempo (**when**). As dimensões who, where, when, what e how [Abowd 1999, Truong 2001] definem o conjunto mínimo de questões para o projeto de sistemas baseados em contexto. Tal conjunto é chamado de contexto primário já que outros tipos de informação contextual podem ser derivados do contexto primário.

3 – O Projeto “Video Family Archive”

Tal projeto tem como objetivo desenvolver técnicas para associação de informações contextuais a respeito de um vídeo sendo apresentado, em especial as informações envolvendo as dimensões apresentadas na seção 2.3. O processo de associação é feito de forma semi-automática, obedecendo ao padrão MPEG-7. O projeto está dividido em três fases:

1. Desenvolvimento de uma ferramenta que receba informações contextuais sobre um vídeo e, automaticamente, gere uma descrição MPEG-7 sobre o mesmo.
2. Desenvolver um *player* instrumentado para assistir vídeos e inserir informações contextuais (durante a apresentação do vídeo).
3. Desenvolver uma ferramenta de busca que se beneficie da infra-estrutura dos itens 1 e 2, tornando as pesquisas (por vídeo, ou peças de vídeos) mais fácil, interativa e amigável.

O projeto destina-se ao uso familiar facilitando a organização, a procura e a exibição dos vídeos gravados pelos familiares nas mais diversas situações (daí o nome *Video Family Archive*). O projeto utiliza como caso de teste um arquivo de 30 anos (quase 33 horas) de vídeos caseiros sobre uma família em ocasiões como nascimentos, batizados, aniversários, natais, e outras onde a família se reunia (ou se reúne).

Optou-se por implementar esse projeto utilizando-se a linguagem Java, devido a esta possuir uma ampla biblioteca de APIs, e de Java deixar portas abertas à integração desse sistema com outros desenvolvidos em linguagens com C ou C++.

3.1 – Esquemas de Descrição para Contexto

Um dos desafios da área de consciência de contexto é a questão de como representar contexto. Abowd [Abowd 2002] argumenta que, sem boas representações, os desenvolvedores de aplicações são levados a implementar esquemas proprietários e limitados para armazenamento e manipulação desse importante tipo de informação.

A categorização do contexto em cinco dimensões (seção 2.3) ajuda a decidir quais informações são relevantes. Porém, os projetistas ainda necessitam analisar os requisitos da aplicação e decidir como modelar e representar as informações dentro de cada dimensão. Uma solução seria a construção de um esquema de descrição (usando padrões para representação semântica como MPEG-7, XML Schema ou RDF) para estruturar e organizar a modelagem de informações contextuais. Contudo, uma vez que os requisitos de contexto variam muito de aplicação para aplicação, acaba sendo difícil (ou talvez impossível) cobrir todas as possibilidades de uso de contexto de uma forma flexível e eficiente.

Uma proposta alternativa é o fornecimento de uma biblioteca de elementos contextuais semi-estruturados que: facilite a tarefa do projetista em criar representações para as informações contextuais (usando as representações fornecidas na biblioteca); e que permita criar estruturas que atendam a necessidades específicas.

A tabela 1 mostra alguns dos elementos de contexto mais comuns da biblioteca. Uma lista completa dos elementos, bem como explicações detalhadas sobre cada um, fogem aos objetivos deste relatório. A biblioteca proposta importa alguns elementos do Multimedia Description Schemes (MDS) MPEG-7 [ISSO 2002b], como e-mail e coordenadas GPS; outros são tipos primitivos da DDL MPEG-7, como ID; e outros são criados a partir da DDL, como *address*.

Tabela 1. Exemplos de elementos presentes na biblioteca.

Dimensões	Categorias	Elementos
Who		ID, Role (usuário, autor, ou ator).
	Name	given name, middle name, family name, title.
	Personal Information	age, height, weight, eyes (cor), hair (cor), birth date.
	Preferences	title, location, creator, keyword, language, subject, period, media format.
	History	user name, user ID, user action, time, date, period.
Where		Location ID, Location Name, Location Description
	OutDoor Location	address, street, number, state, country, zip code, GPS coordinates, geographical coordinates.
	InDoor Location	floor, room, corridor, gate, exit.
	Electronic Location	URL, e-mail address, path, screen coordinates, segment, frame.
When		Occasion, date, time, period, duration
What		user ID, action.

O fato de se ter um amplo conjunto de elementos na forma de uma biblioteca permite a criação de hierarquias variadas para um mesmo tipo de informação contextual. Por exemplo, um “nome” pode ser representado como: um nome somente (uma string); nome, nome do meio e nome de família; nome e nome de família; somente nome de família. A tarefa do desenvolvedor fica facilitada no sentido de que ele não mais necessita projetar e implementar a representação para “nome”. Ele apenas deve decidir o quanto de informação é necessário para representar “nome”. Dessa forma nota-se o surgimento de uma nova dimensão – o quanto (How Much). Deve-se deixar claro que o uso dessa dimensão na biblioteca é implícito.

Como o conjunto de elementos presentes na biblioteca é grande e variado, alguns foram agrupados em categorias. Tais categorias estão expressas na forma de esquemas de descrição (DSs) para serem diretamente usados pelo desenvolvedor; o mesmo acontece com as dimensões. Como a biblioteca utiliza a DDL MPEG-7, ela se torna extensível, permitindo

ao desenvolvedor construir elementos a partir dos fornecidos ou criar novos. Tais elementos podem, então, ser adicionados à biblioteca. A figura 2 ilustra um exemplo da dimensão *Who* personalizada usando-se a categoria *Name*, o elemento *role*, e criando-se um novo elemento – *PersonalInfo* – a partir dos elementos *Age*, *Height*, *Weight*, e *BirthDate* (customização de *Personal Information*).

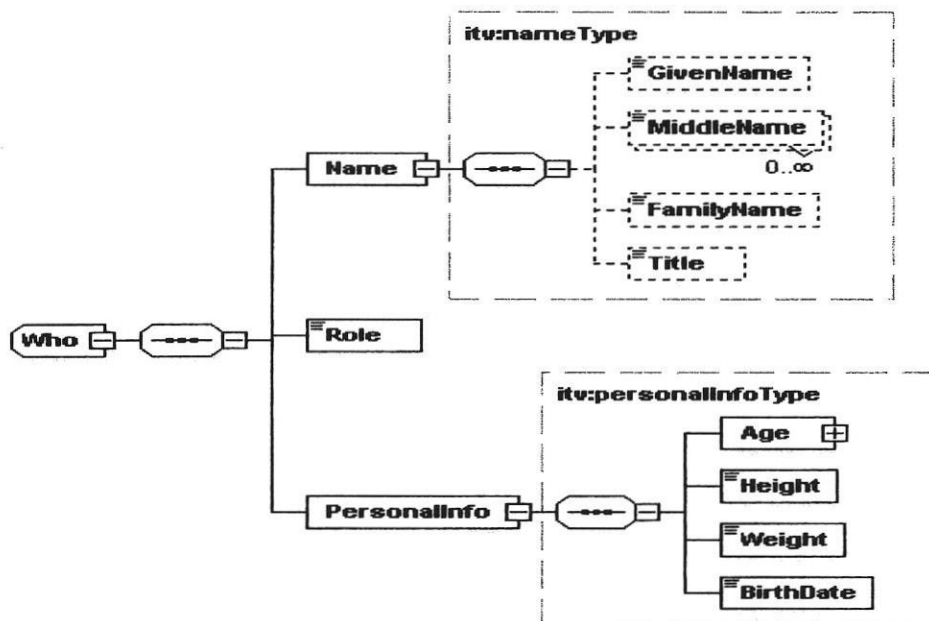


Figura 2. Dimensão Who definida com elementos da biblioteca.

3.2 – O Parser MPEG-7

As cinco dimensões (who, where, when, waht e how) foram modeladas através da construção de esquemas MPEG-7, em uma forma estruturada. Esses esquemas são usados para descrever a informação semântica desejada sobre os dados (multimídia ou não) sendo capturados ou acessados.

Tais descrições podem ser associadas ao vídeo no momento da captura ou posteriormente. Então, as informações contidas nas descrições são usadas para facilitar buscas e acesso a dados previamente armazenados, facilitando a recuperação de informações e o desenvolvimento de aplicações.

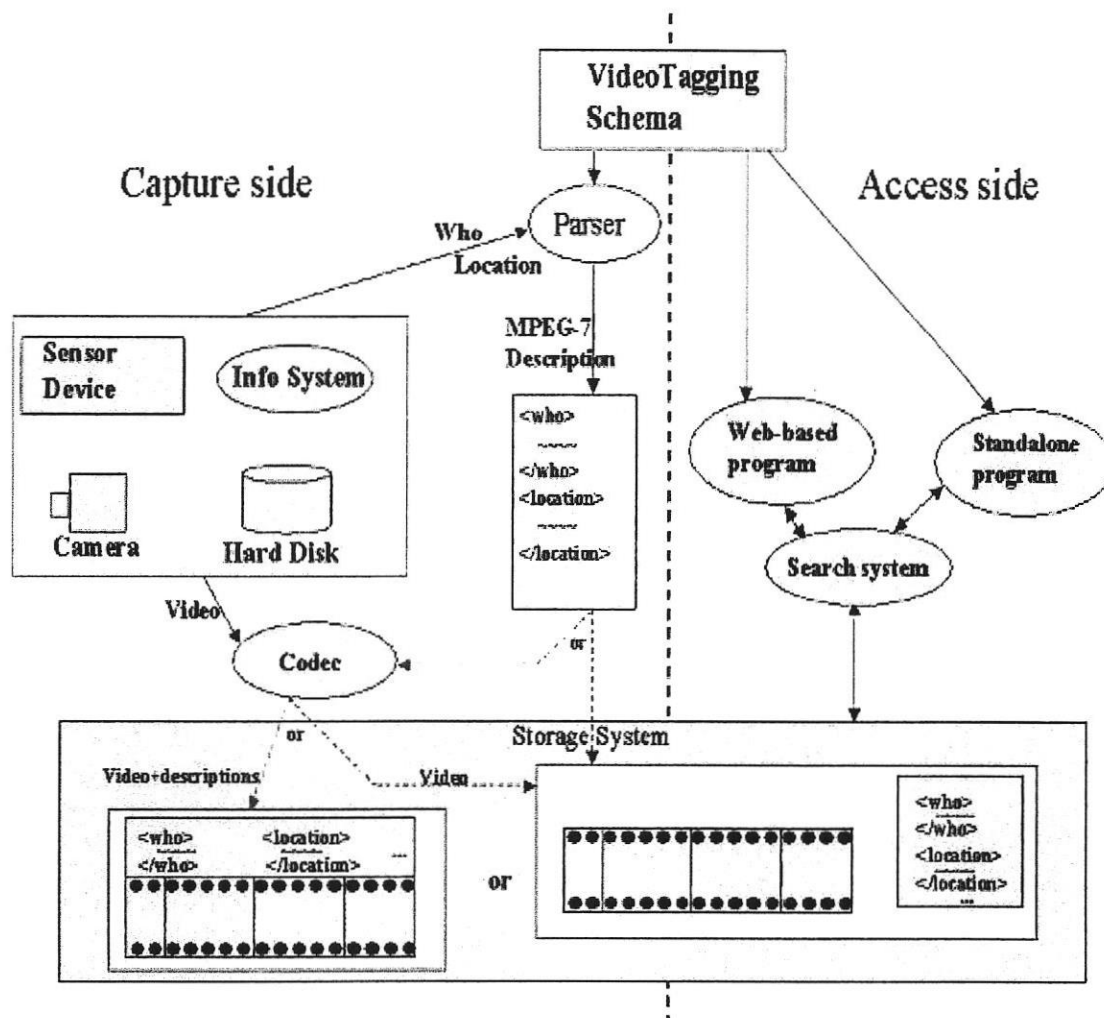


Figura 3. Arquitetura do sistema.

A arquitetura do sistema está ilustrada na figura 3. Nessa figura estão representadas as partes de captura de informação (vídeo, informações contextuais), no lado esquerdo da figura, e de acesso à informação, no lado direito da figura. Ambos os lados utilizam um modelo de informações constituído por esquemas MPEG-7 (Video Tagging Schemas, na figura). Tais esquemas ditam as regras semânticas e sintáticas que governam a descrição dos itens de interesse.

Na parte de acesso, as aplicações, baseadas em web ou programas *standalone*, utilizam os esquemas para validar e associar relacionamentos entre diferentes itens de descrição durante buscas complexas (note-se que tais programas podem ser o próprio *search engine*).

Na parte de captura um *parser* recebe como entrada dados contextuais (vindos de sensores, câmeras, sistemas de informação, etc.) e os esquemas. O *parser* utiliza as regras sintáticas e semânticas contidas nos esquemas para gerar uma árvore em memória contendo as *tags* (definidas pelo projetista através dos esquemas) que irão representar os dados dando sentido semântico aos mesmos. Então, o *parser* percorre a árvore redirecionando a saída para

um arquivo, intercalando as *tags* com os dados. Como saída tem-se um documento estruturado chamado de descrição MPEG-7 (figura 4), que é na verdade uma instância XML de um XML Schema (deve-se lembrar que a DDL do MPEG-7 é o XML Schema estendido). A descrição pode então ser armazenada como um arquivo separado do objeto que representa (opção utilizada no projeto), ou pode ser enviada a um codificador, de vídeo no caso do exemplo da figura 3, que gerará um único arquivo (de vídeo) contendo as imagens e as descrições.

```
<?xml version="1.0" encoding="UTF-8" ?>
<vt:VideoTagging xmlns:vt="http://cc.gatech.edu/VT/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://cc.gatech.edu/VT/ ..\parser\VTSchema.xsd">
<Profile1>
<DateOfRecord>
  <BeginningDate>2000-07-19</BeginningDate>
  <BeginningTime>11:20:11</BeginningTime>
</DateOfRecord>
<SceneBegin>2.86168024</SceneBegin>
<SceneEnd>3.97368024</SceneEnd>
<VideoMediaDescription>
  <MediaInformation>
    <MediaIdentifier>tape1_13.avi</MediaIdentifier>
    <MediaDuration>0:0:18</MediaDuration>
    <VideoEncoding>IV50</VideoEncoding>
    <VideoSize>640x480</VideoSize>
    <VideoFrameRate>24.000383</VideoFrameRate>
    <AudioEncoding>No Audio</AudioEncoding>
    <AudioSampleRate>Unknown</AudioSampleRate>
    <AudioChannels>Unknown</AudioChannels>
  </MediaInformation>
  <MediaLocator>file:C:\users\rudinei\VT\videos\tape1_13.avi</MediaLocator>
</VideoMediaDescription>

  ...
</Profile1>
</vt:VideoTagging>
```

Figura 4. Exemplo de uma descrição gerada pelo *parser*.

3.3 – O *Player*

O *player* tem como funções permitir a visualização de vídeos e a associação de informações contextuais, fornecidas pelo usuário, ao vídeo em si ou à peças do vídeo (cenas ou quadros). Outra função do *player* é servir como interface para as ferramentas de busca que utilizam as informações contextuais.

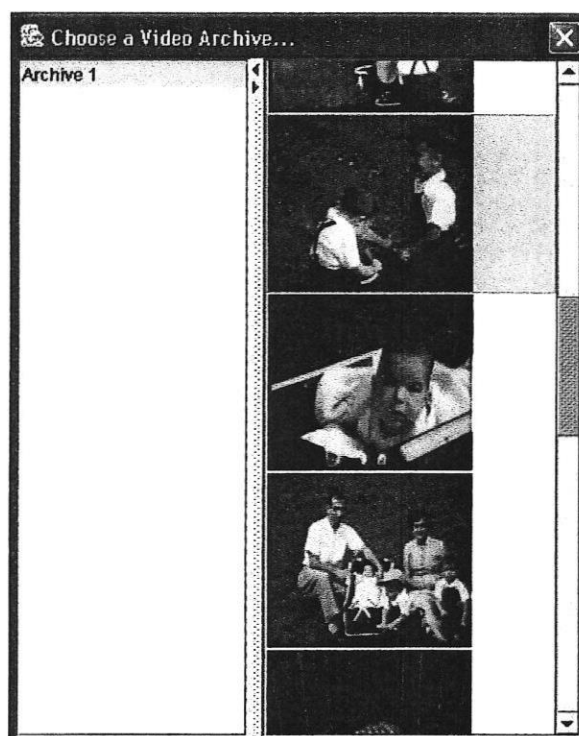


Figura 5. Janela de seleção de arquivos e cenas.

Ao iniciar o a aplicação é apresentada ao usuário uma janela para seleção de um dos arquivos¹ registrados. No exemplo representado pela figura 5, no lado esquerdo, temos apenas um arquivo registrado (Archive 1). Cada arquivo contém uma coleção de vídeos que o usuário pode inserir ou remover. Do lado direito da janela representada na figura 5 estão pequenas imagens (*thumbnails*) representando cenas que o usuário registrou (primeiro quadro de cada cena). Ao selecionar o nome do arquivo o *player* exibirá, continuamente, todos os vídeos do arquivo. Ao selecionar uma das imagens, o *player* exibe, continuamente, o arquivo de vídeos a partir da cena selecionada.

Conforme ilustra a figura 6, o *player* possui três componentes: a janela de visualização de vídeo (ao alto e à esquerda), uma janela para entrada e visualização de dados contextuais (ao alto e à direita) e um janela com controles (em baixo).

¹ No contexto do projeto *Video Family Archive* “arquivo” é uma coleção de vídeos armazenados.



Figura 6. Player e seus componentes.

A janela principal restringe-se a exibir vídeos e a controlar a exibição de quadros (avanço e retrocesso quadro a quadro, *play* e *pause*, e um *slider* para posicionamento rápido).

3.4 – A Interface para Exibição e Entrada de Dados

A figura 7 ilustra uma janela com dupla função no sistema. Essa janela é utilizada para visualizar os dados de uma cena específica, previamente registrada pelo usuário, e para a entrada de dados. Os dados a respeito de cenas, previamente registrados, são exibidos nessa janela assim que o *player* exibe as respectivas cenas. Por exemplo, em um vídeo de 300 segundos foram registradas as seguintes cenas: cena 1, iniciando no segundo 23; cena 2, iniciando no segundo 34, cena 3, iniciando no segundo 125 e cena 4, iniciando no segundo 256. Do segundo 0 ao segundo 23 não são exibidas informações, do segundo 23 ao segundo 34 são exibidas as informações sobre a cena 2, do segundo 34 ao 125 as informações da cena 3, e assim por diante.

Toda vez que o usuário deseja registrar uma nova cena no arquivo, ele deve utilizar o botão *New Scene* na janela de controles do *player* (figura 8). Isso fará com que a janela de

visualização de informações apresente suas tabelas, com informações sobre Quem, Onde, Quando e Atividade, em branco, esperando o usuário preencher os campos.

The screenshot shows a window titled "Scenes' Information" with standard window controls (minimize, maximize, close). It contains several tables and a set of radio buttons on the right.

Name	Primary Role	Relation
John	true	brother
Kevin	false	brother

Location Description	City	State
Central Park	New York	NY

Month	Day	Year	Occasion
12	25	1956	Christmas

Activity
playing

On the right side of the window, there are four radio buttons:

- ☒ **Who**
- ☐ **Where**
- ☐ **When**
- ☐ **Activity**

Figura 7. Janela de entrada e exibição de dados.

Nesse momento o sistema automaticamente realiza a captura implícita de informações sobre o vídeo sendo exibido (nome do arquivo de vídeo, tipo de compressão, tamanho do arquivo, taxa de quadros por segundo, duração em segundos, etc.) e sobre a cena que se está registrando (arquivo de vídeo ao qual a cena pertence, tempo inicial da cena, data, etc.).

Assim que o usuário termina de preencher os campos que escolheu, ele pressiona o botão *Save Scene* da janela de controles. Então, o sistema captura, explicitamente, as informações das tabelas enviando-as em um formato padrão ao *parser*. O *parser* recebe as informações, consulta os esquemas pertinentes e gera uma descrição MPEG-7 da cena. A descrição é então gravada em disco para ser usada no futuro como forma de facilitar buscas.

3.5 – A Janela de Controles do *Player*

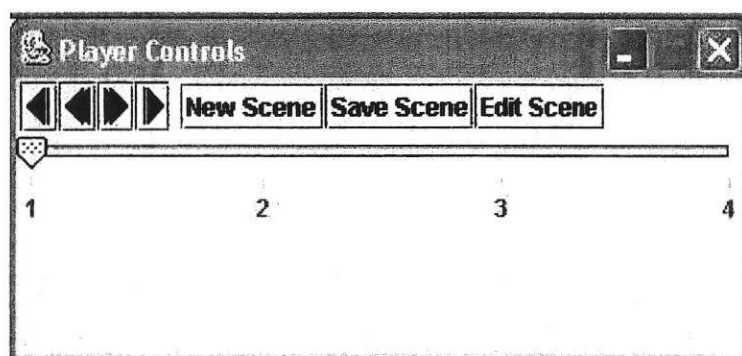


Figura 8. Controles do *Player*.

Além dos botões *New Scene* e *Save Scene*, já explicados, a janela de controles (ilustrada na figura 8) possui o botão *Edit Scene*. Esse botão suspende a transição de informações de cenas da janela da figura 7 para que o usuário tenha tempo para fazer alterações nas informações anteriormente registradas. Ao término da edição basta pressionar o botão novamente para que o sistema grave as novas informações. A janela de controles possui ainda os botões triangulares (ao alto e à esquerda). Esses botões servem como controles auxiliares para avanço e retrocesso rápidos e quadro a quadro. O último componente da janela de controles é um *slider* com escala. A escala, nesse caso, serve como indicador dos vídeos que constituem o arquivo (quatro no exemplo), e como meio de acesso aos mesmos.

3.6 – Objetos MPEG-4 e A Interface de Busca

A interface de busca, originalmente projetada para trabalhar com os vídeos em formato AVI, foi aprimorada para suportar vídeos em formato MPEG-4. A idéia é segmentar e descrever um vídeo em vários objetos de mídia (cenas, quadros de vídeo, regiões dentro de um quadro de vídeo, etc.) através das ferramentas já descritas. Tal segmentação ocorre sem replicação (nenhum arquivo de vídeo ou imagem novos são gerados), mantendo suas relações hierárquicas dos objetos e fornecendo meios de se estabelecer relações implícitas (semânticas). As funcionalidades originais da ferramenta de busca foram preservadas, adicionando-se apenas os requisitos necessários para se ter acesso a objetos MPEG-4. Isto incluiu a utilização de um *plug-in* MPEG-4 [Envivio 2002] para o QuickTime *player*, uma vez que o *player* desenvolvido no projeto não suporta tal formato. Então, aqui é apresentada apenas a versão com suporte à MPEG-4.

Para estabelecer as relações estruturais, espaciais, temporais, semânticas e contextuais, foi utilizado o esquema de descrição MediaObject [Goularte 2002], o qual, por sua vez, utiliza

a biblioteca de contexto descrita na seção 3.1. Tal esquema é também responsável por associar as instâncias (descrições MPEG-7) às mídias.

Gerou-se uma descrição de cada um dos objetos de mídia (segmentos), com atenção especial às características contextuais. Os objetos de mídia foram disponibilizados no servidor, assim como suas descrições.

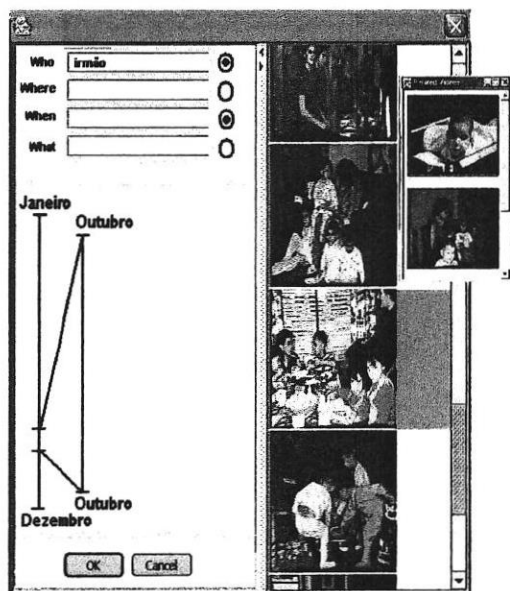


Figura 9. Protótipo do mecanismo de busca.

Para acessar os objetos de interesse foi implementado o protótipo de ferramenta de busca ilustrado na figura 9. A ferramenta utiliza as descrições contextuais dos objetos de mídia, permitindo que o usuário realize buscas complexas apenas usando as dimensões de contexto. A entrada de dados pode ser feita via os campos de texto ou via ferramentas gráficas. O exemplo ilustra uma busca pela palavra chave “irmão” na dimensão “Who”, e pela especificação de um intervalo de tempo na dimensão “When” (Outubro). O processo de busca verifica as descrições dos objetos procurando por tags <Who> onde o conteúdo seja “irmão”, e por relações entre objetos onde o conteúdo do atributo “arcrole” do XLink [Goularte 2002] representado a ligação seja “irmão”. A resposta da busca é uma lista de imagens representando objetos que satisfazem à busca. A lista pode ser indexada de acordo com o tipo de mídia do objeto.

O protótipo ainda fornece um serviço de sugestões de quadros de vídeos ou de cenas relacionados ao conteúdo pesquisado (janela menor, ao alto e à direita). No exemplo, o sistema sugeriu objetos que, apesar de não conterem a palavra “irmão” na dimensão “Who” nem em relações, são objetos que foram assinalados com “Outubro” na dimensão “When”. A estratégia de sugestão pode também ser baseada no histórico de uso do usuário ou nas preferências do usuário (a biblioteca de contexto possui as categorias “History” e



Figura 10. Objeto de mídia MPEG-4 sendo apresentado.

“Preferences” que podem ajudar nesse propósito). Assim que um objeto é selecionado da lista, o mesmo passa a ser apresentado em um *player* que possua o *plug-in* EnvivioTV instalado, como ilustrado na figura 10.

As descrições MPEG-7 podem ser codificadas e armazenadas junto com os objetos de vídeo, em um único arquivo MPEG-4, ou pode ser armazenadas separadamente, em formato texto ou binário. Esta versão da ferramenta utiliza as descrições em formato texto, realizando as buscas em memória com o auxílio do *parser* descrito na seção 3.2. Deve-se lembrar que o *parser*, regra geral, “lê” um documento XML (.xml ou .xsd) gerando uma representação estruturada desse documento em memória: uma árvore DOM [W3C 2002]. Tal árvore pode então ser utilizada para fins de busca.

O padrão MPEG-7 especifica um formato binário (BiM) [ISO 2002b] para suas descrições. Porém, os codificadores ainda não estão disponíveis e a implementação de tais codificadores está além dos objetivos deste trabalho. A estratégia adotada por este trabalho envia os vídeos (ou objetos de vídeo) requisitados pelo usuário juntamente com suas descrições MPEG-7, permitindo que buscas locais (dentro ou a respeito do mesmo vídeo) sejam realizadas sem a necessidade de se recorrer ao servidor. O fato de se ter as descrições dos objetos junto aos objetos no lado do cliente, o fato de estas descrições estarem organizadas hierarquicamente, e o fato de que essas descrições fazem referências às mídias que descrevem preenchem uma lacuna: como associar descrições MPEG-7 a conteúdo MPEG-4.

4 - Conclusões

Este relatório apresentou o desenvolvimento de uma ferramenta para associação de informações contextuais a objetos de vídeo, contribuindo, principalmente, em duas áreas: consciência de contexto e integração MPEG-4/MPEG-7.

Na área de consciência de contexto o trabalho contribuiu fornecendo alguns subsídios para o desenvolvimento de uma forma de representação de contexto. A biblioteca proposta, apesar de não ser uma solução definitiva, auxilia no processo de desenvolvimento de sistemas conscientes de contexto uma vez diminui o esforço do projetista em definir elementos de contexto que representem informações relevantes.

O grupo MPEG, apesar de já estar trabalhando na integração MPEG-4/MPEG-7, ainda não definiu a forma como descrições MPEG-7 serão associadas a conteúdo MPEG-4. Baseado

em algumas propostas apresentadas aos comitês e no histórico do grupo MPEG, tudo indica que essa associação se dará em formato codificado (binário), adicionando o BiM MPEG-7 como um *stream* MPEG-4.

A alternativa apresentada é uma solução imediata e factível para esta associação, que pode trabalhar em conjunto com representações binárias. Outro ponto a ser levantado é que as instâncias do esquema para descrições, podem no futuro, ser convertidas para o formato binário MPEG-7 e inseridas no *stream* MPEG-4 final. A vantagem de ser usar esta abordagem é a possibilidade de utilização dos elementos de contexto presentes no esquema, abrindo espaço para o desenvolvimento de sistemas conscientes de contexto baseados em MPEG-4.

Referências

- [Abowd 1999] Abowd, G. D.; Dey, A. K.; Brown, P. J.; Davies, N.; Smith, M.; Steggles, P.: Towards a better understanding of context and context-awareness, In Gellersen, H. (ed.): *Handheld and Ubiquitous Computing, Lecture Notes in Computer Science*, Springer-Verlag, v. 1707, pp. 304-307, 1999.
- [Abowd 2002] Abowd, G. D.; Mynatt, E. D.; Rodden, T.: *The Human Experience. Pervasive Computing*. Janeiro-Fevereiro, p. 48-57, 2002.
- [Battista 1999] Battista, S.; Casalino, F.; Lande, C.: MPEG-4: A Multimedia Standard for the Third Millennium, part 1, *IEEE MultiMedia*, v. 6, n.4, pp. 74-83, 1999.
- [Ebrahimi 2000] Ebrahimi, T.; Horne, C.: MPEG-4 Natural Video Coding - an Overview, *Signal Processing, Image Communication*, v. 15, n 4-5, pp. 365-385, 2000.
- [Envivio 2002] Envivio. Overview, (2002), on-line:
<http://www.envivio.com/MIRRORSITE/solutions/>
- [Faria2001] Faria, G. B.; Santos Jr, J. B.; Goularte, R.; Moreira, E. S.: Uso de perfis em aplicações de televisão interativa conscientes de contexto. *Anais do SBMídia 2001 VII Simpósio Brasileiro de Sistemas Multimídia e Hiperemídia*, pp. 139-154, outubro de 2001.
- [Goularte 2002] Goularte, R.; Moreira, E. D. S.: Representação de Objetos de Mídia para Aplicações Conscientes de Contexto em TV Interativa. *Anais do SBMídia 2002 VIII Simpósio Brasileiro de Sistemas Multimídia e Hiperemídia*, pp. 150-165, outubro de 2002.

- [Hjelsvold 2001] Hjelsvold, R. et. al.: Web-based Personalization and Management of Interactive Video. Anais do décimo International World Wide Web Conferece, p. 129-139, 2001.
- [ISO 2002a] ISO/IEC JTC1/SC29/WG11 N4668: Overview of the MPEG-4 Standard, on-line (2002): <http://mpeg.telecomitalialab.com/standards/mpeg-4/mpeg-4.htm>.
- [ISO 2002b] ISO/IEC JTC1/SC29/WG11 N4509: Overview of the MPEG-7 Standard, on-line (2002): <http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm>.
- [Truong 2001] Truong, K. N.; Abowd, G. D.; Brotherton, J. A. Who, What, When, Where, How: Design Issues of Capture & Access Applications. Lecture Notes on Comptuer Science 2201 - Ubicomp 2001: Ubiquitous Computing, pp. 209-224. Atlanta, GA, USA. Setembro de 2001.
- [TV-Anytime 2001] TV-Anytime Forum: Specification Series: S3: Metadata (Normative), on-line (2001): <ftp://tva:tva@ftp.bbc.co.uk/pub/Specifications/SP003v11.zip>
- [W3C 2001a] World Wide Web Consortium (W3C): XML Schema Part 0: Primer, on-line (2001a): <http://www.w3.org/TR/xmlschema-0/>, [11/12/01].
- [W3C 2002] World Wide Web Consortium (W3C): Document Object Model (DOM) Level 3 Core Specification, on-line (2002). <http://www.w3.org/TR/2002/WD-DOM-Level-3-Core-20020409/>