10

Content Organization

John R. Smith[1] and Ana B. Benitez[2]

[1]*IBM T. J. Watson Research Center, Hawthorne, USA,* [2]*Columbia University, New York, USA*

10.1 INTRODUCTION

With the tremendous growth of digital multimedia data, multimedia systems and technologies are evolving rapidly to meet this challenge. The evolution can be described as one in which the technology focus is moving from treating multimedia content as signals to more advanced extraction and processing of multimedia features, semantics and knowledge, as illustrated in Figure 10.1. By initially handling multimedia content as signals, the MPEG-1 and MPEG-2 audiovisual coding standards have allowed efficient storage, compression and communication. More recently, by extracting and analyzing features of the multimedia signals, MPEG-4 has further improved coding efficiency by using, for example, object-based coding techniques. In addition, MPEG-7 allows content-based retrieval of multimedia by extracting and searching features using similarity measures. However, the challenge remains for automatically extracting semantic labels of multimedia content, including labeling of objects, events, places, people and so forth. By labeling multimedia content at the semantic level, the content will be easier to search, filter, index, summarize, personalize and repurpose. Ultimately, semantic labels allow the extraction of knowledge through the mining of multimedia archives. This can enable business use of multimedia and decision support and allow seamless marketplaces for multimedia content.

In bridging these gaps from multimedia signals to knowledge, models have an important role. Models provide parameterized or statistical representations of the multimedia information. Given a collection of digital images of sunsets, the color features of the images can be used for forming a model of sunset images. This can be accomplished by extracting color histogram descriptors from the images and by computing a centroid color histogram feature vector. By modeling features and semantic information in this way, unknown images and, in general, multimedia content can be labeled by fitting its features to the models, or equivalently, by classifying the content according to the models. The color histogram model can be used to classify unknown images by comparing a color histogram feature vector of an unknown image with the centroid color histogram feature vector of the sunset image semantic class.

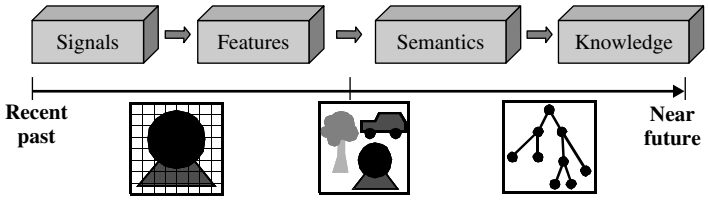

Figure 10.1 Evolution of multimedia applications and technologies involves transition from dealing with signals to features to semantics to knowledge. Models help in bridging the gap from signals to knowledge

Given the importance of semantic labels and the need for modeling content at the feature level in order to automatically annotate and classify multimedia content, MPEG-7 has standardized tools for describing collections and models as follows:

- *Collection tools*: The collection tools describe collections of multimedia content, segments, descriptor instances, concepts or mixed content. Collections can be used for the tasks of describing an album of songs, a group of objects or a cluster of color feature descriptors. The model tools describe parameterized models of multimedia content or collections.
- *Model tools*: The model tools include probability models, analytic models and cluster models. The probability models describe statistics or probabilities associated with the attributes of multimedia content, descriptors or collections. The analytic models describe the association of labels or semantics with multimedia content or collections. The cluster models further describe the association of probability models with the analytic models. Cluster models can describe the centroid of a collection of color feature descriptions of a semantic class of images. Classification models are composed of sets of cluster models or analytic models, and describe information that can be used for classifying multimedia content, as described later.

The MPEG-7 content organization tools are based on the following key concepts:

- *Collection*: Unordered sets of multimedia content, segments, descriptors, concepts or mixed sets of the above.
- *Model*: Parameterized representation of an instance or class multimedia content, descriptors or collections, as follows:
  - *Probability model*: Associates statistics or probabilities with the attributes of multimedia content, descriptors or collections.
  - *Analytic model*: Associates labels or semantics with multimedia content or collections.
  - *Cluster model*: Associates labels or semantics and statistics or probabilities with multimedia content collections.
  - *Classification model*: Describes information about known collections of multimedia content in terms of labels, semantics and models that can be used to classify unknown multimedia content.

10.2 COLLECTIONS

This section describes MPEG-7 concepts and tools related to collections [1]. As shown in Figure 10.2, the Collection tools include tools for content collections, segment collections, descriptor collections, concept collections, mixed collections and structured collection. The Content Collection tool describes collections of multimedia documents, whereas the Segment Collection tool describes collections of segments of multimedia data, including

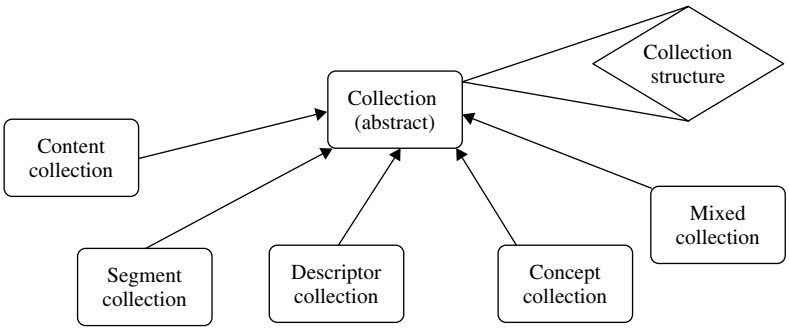

Figure 10.2 Overview of Collection tools – Content Collection, Segment Collection, Descriptor Collection, Concept Collection and Mixed Collection tools are specialized types of Collection tools. The Structured Collection describes the relationships among Collections

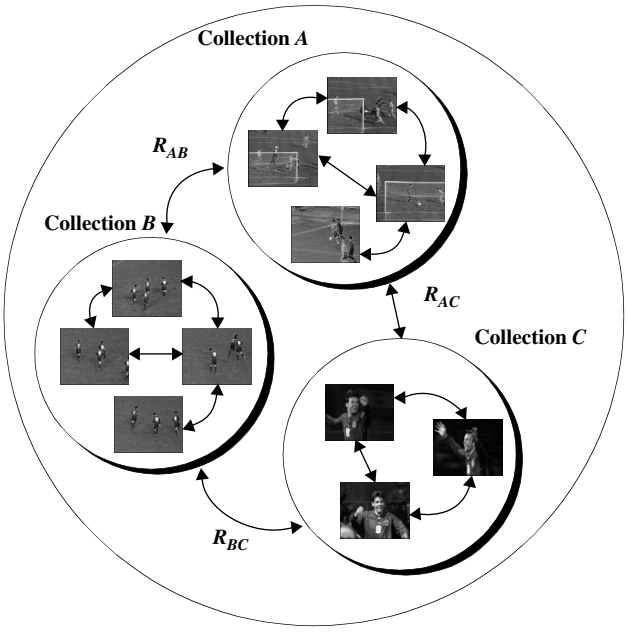

Figure 10.3 The Content Collection tools describe a collection of multimedia content. Collection Structure tool describes the relationships across multiple collection

image regions and video segments. The Descriptor Collection tool describes collections of multimedia content descriptions that are instances of descriptors, such as the MPEG-7 Dominant Color Descriptor (DCD) or Scalable Color Descriptor. The Concept Collection tool describes collections of semantic concepts, such as objects, events, people and places. The Mixed Collection tool describes collections of content, descriptions and concepts. Finally, the Structured Collection tool describes the relationships among collections.

As described above, the Collection tools describe collections related to multimedia content. Figure 10.3 shows an example of a collection structure including multiple collections of images. In this example, each collection consists of a set of images, perhaps with common properties, each depicting similar events in a soccer game. Within each collection, the relationships among the images can be described, such as the degree of similarity of the images in the cluster or spatiotemporal relationships. Across the collections, the degree of similarity of the collections can be described.

10.3 MODELS

This section describes MPEG-7 concepts and tools related to models [1]. The Models are expressed in terms of statistics or probabilities associated with the attributes of collections of multimedia content or are expressed through examples or exemplars of the multimedia content classes. The Model tools describe parameterized models of multimedia content, descriptors or collections. Figure 10.4 shows an overview of the different types of models: probability models, analytic models, cluster models and classification models. Each of these types of Model tools is further broken down into specific types of Model tools. The Probability Model tool describes different statistical functions and probabilistic structures, which can be used to describe samples of multimedia content and classes of descriptors using statistical approximation. The Analytic Model tool describes a collection of examples of multimedia content or clusters of descriptors that are used to provide a model for a particular semantic class. A collection of art images labeled with tag indicating that the paintings are examples of the Impressionist period forms an analytic model. The Cluster Model tool describes analytic models and their associated probability models. This allows the description of the centroid of a collection of descriptor instances

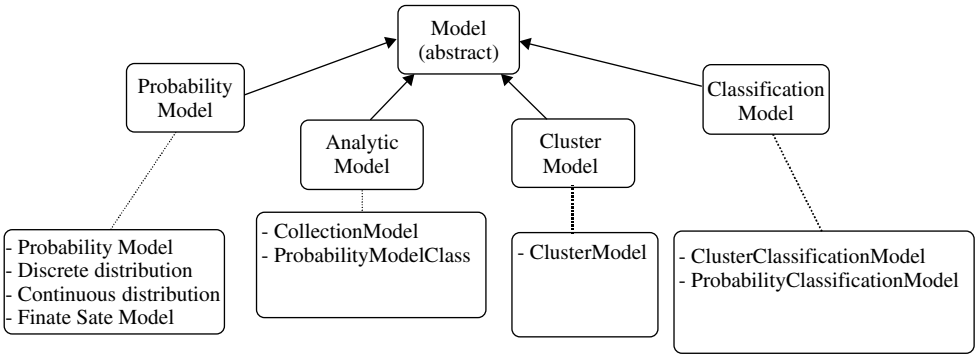

Figure 10.4 Overview of the different types of models: probability models, analytic models, cluster models and classification models

in order to represent a semantic class. Finally, the Classification Model tool describes probabilistic information that can be used for classifying multimedia.

10.3.1 Probability Models

This section describes MPEG-7 concepts and tools related to probability models [1]. The Probability Model tool describes attributes and features of multimedia content using probabilities and statistics and apply to different types of discrete probability distributions, continuous probability distributions and finite-state models.

A probability distribution characterizes the statistical properties of random variables in the form of a distribution. For many purposes, pieces of information suffice for characterizing the distribution. Information about the location (mean) and dispersion (variance) is most frequently used. In general, distributions are discrete or continuous. In the discrete case, a probability mass function is described by specifying its form and parameters (i.e. probability of success, number of trials and so forth). Furthermore, an arbitrary discrete probability mass function is described by specifying the values of the distribution directly using histograms. In the continuous case, a probability density function is likewise described by specifying its form and parameters (i.e. mean, variance, moments, shape and so forth).

Tools for describing probability distributions are provided that cover cases of both discrete and continuous probability distributions. The Probability Distribution tool includes attributes that are common to all distributions, including mean, variance, minimum, maximum, mode, median, moment and so forth. In general, the probability distributions are multidimensional and are described using parameters of the distributions that are vectors. In the case that a distribution is one-dimensional, only the first element of each vector is used, which specifies the corresponding parameter as a scalar quantity. The different forms of discrete probability distributions that are described by the discrete distribution tools include arbitrary histograms and binomial, hypergeometric, Poisson, geometric and uniform distribution types. The different forms of continuous probability distributions that are described by the continuous distribution tools include Gaussian, generalized Gaussian, exponential, gamma, continuous, lognormal and Gaussian mixture model distribution types. For example, the probability density function for a generalized Gaussian distribution given as follows:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{|(x-\mu)|^r}{2\sigma^2}\right],$$

for $-\infty < x < \infty$ and $\sigma^2 > 0$, where μ corresponds to the mean, σ corresponds to the standard deviation and r corresponds to the shape, is described by specifying the values for the mean, standard deviation and shape.

The Probability Model tools also apply to finite-state model types, which correspond to models of Markov sources, or equivalently, probabilistic functions of Markov chains. The underlying assumption in using finite-state models is that the system is described at any time as being in one of a finite set of distinct states. The states may correspond to the different program topics of a news programs, that is, local news, weather,

sports, commercials and so forth. At discrete times, the system undergoes a change of state, which is possibly back to the same state, according to a set of probabilities associated with the state. A probability is associated with the likelihood of the transition from sports to weather in news programs. In the special case of discrete-time, first-order Markov chains, the probabilistic dependence between states is truncated such that the probability of the current state depends only on the preceding state. Furthermore, it is assumed that the probabilities are independent of time, which leads to a set of state-transition probabilities.

The Finite State Model tools describe finite-state models, state-transition models and Hidden Markov Models (HMMs). The assumption made for State Transition Models is that each state corresponds to a deterministically observable event, that is, the output of each source being in any given state is not random. Since this model is often too restrictive, the HMM tool includes the case in which the observation is a probabilistic function of the state. The Finite State Model tool defines the abstract base class for state-transition models and discrete HMMs. The finite-state model tool includes an attribute that indicates the number of states, which applies to both state-transition models and discrete HMMs. The state-transition model tool specifies each of the states and gives both the state-transition probabilities as a matrix and the initial state probabilities as a vector. The discrete HMM tool extends the state-transition model tool and provides additional attributes and elements that describe the probabilistic nature of the different observations of each state. The discrete HMM tool includes an attribute that indicates the number of observations per state. It also specifies a sequence that defines the observation symbols and the observation symbol probability distribution of each state [2].

10.3.2 Analytic Models and Cluster Models

This section describes MPEG-7 concepts and tools related to analytic models [1]. The Analytic Model tools describe the association of labels or semantics with collections or classes of multimedia content. The analytic models may be specified by enumerating the items contained in the collection, by describing parameterized representations of the collections in the form of probability models, or both. This allows the analytic models to characterize different semantic concepts by describing the following:

- clusters of multimedia data representing those concepts;
- example collections of multimedia content descriptions; and
- probability models.

The Analytic Model tool is the base type of the tools for describing collection models, probability model classes and cluster models. The Collection Model tool describes the association of semantics with a collection. The Collection Model tool allows the association of a particular semantics with multiple collections. The following describes a collection model consisting of a content collection of three images. The collection of images describes the concept of ‘soccer shots on goal’.

```
<CollectionModel confidence="0.75" reliability="0.5" function="described">
        <Label><Name>Soccer shots on goal</Name></Label>
        <Collection xsi:type="ContentCollectionType">
                <Content xsi:type="ImageType">
```

```
                        <Image>
                                <MediaLocator xsi:type="ImageLocatorType">
                                        <MediaUri>soccer1.jpg</MediaUri>
                                </MediaLocator>
                        </Image>
                </Content>
                <Content xsi:type="ImageType">
                        <Image>
                                <MediaLocator xsi:type="ImageLocatorType">
                                        <MediaUri>soccer2.jpg</MediaUri>
                                </MediaLocator>
                        </Image>
                </Content>
                <Content xsi:type="ImageType">
                        <Image>
                                <MediaLocator xsi:type="ImageLocatorType">
                                        <MediaUri>soccer3.jpg</MediaUri>
                                </MediaLocator>
                        </Image>
                </Content>
        </Collection>
</CollectionModel>
```

The Probability Model Class tool describes the association of semantics with classes of descriptions of multimedia content, which are represented using probability models. The probability models provide an aggregate statistical representation of the descriptions in a class. The Probability Model Class tool allows the association of the same semantics with multiple probability models.

10.3.3 Classification Models

This section describes MPEG-7 concepts and tools related to classification models [1]. The Classification Model tool describes information about known multimedia content that can be used to classify unknown multimedia content. The output of the classifiers is the association of labels or semantics with the unknown content. The classifier uses a classification model that describes the association of labels or semantics with known content. The classification model consists of analytic models, which can be content collection models, descriptor collection models or probability class models. In each of these cases, using the group of models, the classifier is able to analyze the unknown content and choose which analytic model best fits the unknown content.

10.4 EXAMPLES OF USAGE

The Cluster Model tool can be used to describe clusters of images. Consider the relevance feedback search scenario illustrated in Figure 10.5. Initially, the user issues a query to an image database. The image database then retrieves images matching the query and returns them to the user. As shown on the left-hand side in Figure 10.5, the user marks which images are relevant to the user's information need, call these positive examples (+), and the remainder are considered as not relevant. This information is then used by

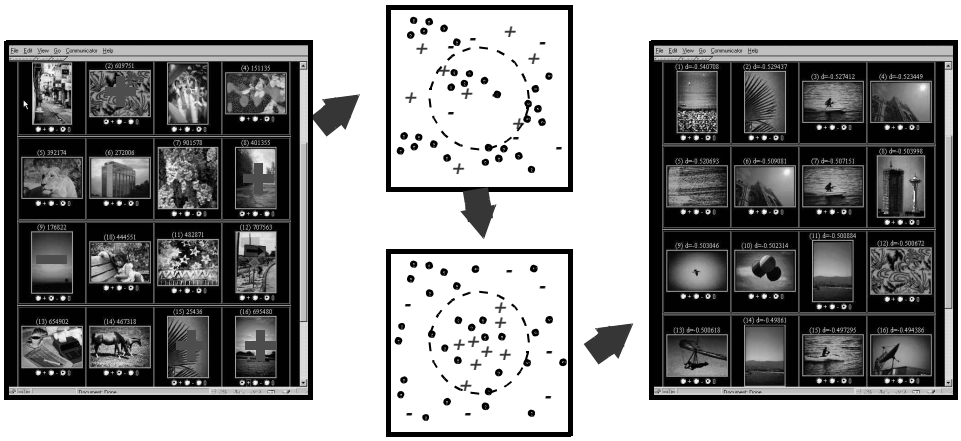

Figure 10.5 Example relevance feedback search scenario for image databases. The clusters of positive and negative examples of images are described using Cluster Model tool

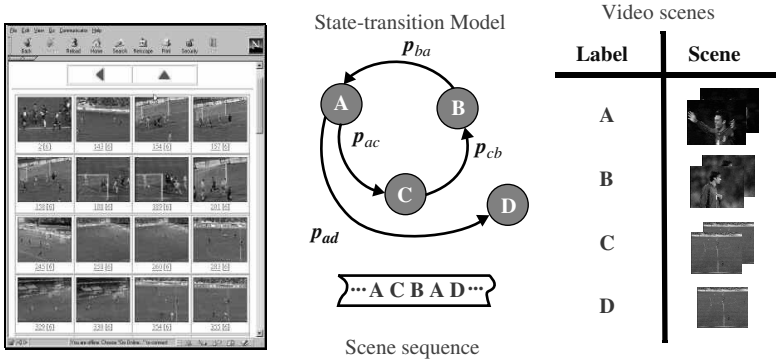

Figure 10.6 Example modeling of soccer video sequence using State Transition Model tool

the retrieval system to transform the metric space that measures similarity of images such that relevant images are clustered together and are returned to the user, as shown on the right-hand side of Figure 10.5 [3]. In this example, the description of the two clusters of images – positive and negative examples, respectively – are described using the Cluster Model tool.

The state-transition model can be used to summarize the transition of events along the temporal dimension of a video sequence. Consider the example in Figure 10.6, which involves a video of a soccer game. In the game, certain types of scenes, such as 'shots on goal', 'goal scores', 'great passes', and 'corner kicks', recur throughout. Each of these scene types can be described using the Model State tool (Analytic Model) that associates semantics with the content. Then, the State Transition Model tool groups together the recurring states and describes the probability of transition between the scenes, as shown in the center and right-hand side in Figure 10.6. This allows the State Transition Model to characterize the entire soccer game video. The State-Transition Model description can

then be compared with other descriptions in order to retrieve similar video sequences or classify the video sequence [4].

REFERENCES

[1] ISO/IEC 15938 (2001), Information Technology – Multimedia Content Description Interface, Singapore.

[2] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, Englewood Cliffs, NJ, 1993.

[3] C.-S. Li, J. R. Smith and V. Castelli, "S-STIR: similarity search through iterative refinement," *IS&T/SPIE Symposium on Electronic Imaging: Science and Technology – Storage & Retrieval for Image and Video Databases VI*, San Jose, Calif., 1998, pp. 250–258.

[4] J. R. Smith and C.-S. Li, "Image classification and querying using composite region templates," *Journal of Computer Vision and Pattern Recognition*, **75**(1/2), 165–174 (1999).