Abstract. Cosmos-7 is an application that can create and filter MPEG-7 semantic content models with regards to objects and events, both spatially and temporally. The results are presented as numerous video segments that are all relevant to the user’s consumption criteria. These results are not ranked to the user’s ranking of relevancy, which means the user must now laboriously sift through them. Using self organizing networks we rank the segments to the user’s preferences by applying the knowledge gained from similar users’ experience and use content similarity for new segments to derive a relative ranking.

1 Introduction

Filtering multimedia content is complex because the medium is transient both spatially and temporally. Therefore the content itself has different semantic meaning both spatially and temporally in relation to objects and events, respectively. In order to be able to filter multimedia content we require; 1) A content model that describes the content in terms of spatial and temporal semantic relationships of object and events, 2) A filter that sifts relevant information from the content model based on the user’s information requirements.

COSMOS-7 [1] is an MPEG-7 compliant application that reduces the complexity of creating such a content model and filter. It exclusively uses part 5 of the MPEG-7 [2] standard (Multimedia Description Schemes) that semantically describes objects and events and there relationships both temporally and spatially. Unlike other multimedia content modeling systems [3] it does not use low level (syntactic) features, only high level (semantic features) that are meaningful to the user. Using the COSMOS-7 filtering manager a filter is created that can exploit the rich detail captured in the content model by allowing a user to filter out undesirable content.

On examination of the results after filtering it was found there were numerous entries returned that fitted the filter criteria. These results were not ranked by the relevancy to the context of the user’s information requirements. This is achievable by understanding the importance a user attaches to high level features. Using collaborative ranking, which is similar to filtering but without exclusion of items, we can predict the user’s preference for content by extracting similar users and using...
their preference ranking for content. In this paper we examine the use of two self organizing neural networks to 1) collaboratively filter users into similar clusters and then rank the segments for the user using the previous experience of the peer group 2) Use content based video similarity measures to rank segments outside the peer groups experience in order to find a relative ranking based on experience of similar content by the peer group.

In section 2 an overview is provided of what semantic aspects COSMOS-7 models and what MPEG-7 tools it uses to encapsulate these concepts. Section 3 describes the filtering process of COSMOS-7 to extract a video summary of user preferred content. Section 4 describes the two self organizing neural networks used to personalize the results to the user’s taste. The final section is our conclusion and future research.

2 Overview of COSMOS-7

In this section we begin by specifying the attributes that need to be modeled to provide a fully inclusive description of the semantic content which is both concise and flexible. We then describe the COSMOS-7 System, from two particular angles; 1) How COSMOS-7 is modeled using specific MPEG-7 tools to produce such a rich and multi faceted content model. 2) The COSMOS-7 filtering manager that creates and manages filters for extracting content to the user’s consumption requirement.

2.1 Modeling Semantic Content Attributes

The semantic content model has to be tightly integrated with the video streams using referencing strategies. In this way, the filtering process may determine the meaning conveyed by the media within the archive so as to compare against the specified filtering criteria. Previously [1] we have identified four key aspects;

Events - Events within the semantic content model represent the context for objects that are present within the video stream at various structural levels. Events are therefore occurrences within the video stream that divide it into shorter semantically continuous content segments involving specific objects, and thus can frequently serve to represent object behaviour.

Temporal Relationships - Temporal relationships between events enable the semantic content model to express the dynamism in content that is apparent at these higher levels, thereby enabling filtering of non-static semantic content which is more in line with “What will or did happen here?” Again, this may occur on both a general and a specific level.

Objects - The expression of objects and object properties within the semantic content model provides for filtering with regards to objects that are readily identifiable within the video content stream. The term ‘object’ refers to any visible or hidden object depicted within a video frame at any necessary level of detail, from entire objects and groups of objects to the bottom-most component objects. Objects may themselves exist within a structural hierarchy thereby enabling inheritance of properties from higher level objects.