CONTENT MANAGEMENT IN A CONSUMER DIGITAL VIDEO LIBRARY DESIGN

Chin Chye Koh[*], *Jorge E. Caviedes*[†], *and Sanjit K. Mitra*[*].

[*]Department of Electrical and Computer Engineering,
University of California, Santa Barbara, CA 93106
chinkoh@ece.ucsb.edu, mitra@ece.ucsb.edu

[†]Intel Corporation,
5000 W. Chandler Blvd, Chandler AZ 85226
jorge.e.caviedes@intel.com

ABSTRACT

With the increasing amount of research in digital video processing and video quality, there is a need for effective and timely communication of ideas and results. An area that needs improvement is the use of a standardized pool of test videos that are freely available and well known. In this paper we discuss the need for such a standardized suite of test videos for the explicit purposes of research collaboration and benchmarking. To ensure that a video library with such content is available, we discuss the type of videos that are pertinent; and we have also outlined a framework for an online digital video library that would best help the dissemination of information and encourage the use of the library in collaborative research and benchmarking.

1. INTRODUCTION

The growth of the consumer electronics industry driven by the increasingly knowledgeable consumer has resulted in intense activity in the research and development of digital video processing algorithms and digital video quality metrics. These algorithms and metrics are geared towards enhancing the visual experience for a multiplicity of content types and viewing options. As the consumer's knowledge of video processing expands and their expectations for video quality increases, more advanced processes embedded into consumer products will become necessary. There will also be little room for mistakes in the execution of these products. Thus, for the R&D community, there is an increased emphasis on thorough and efficient testing and validation throughout the product development cycle.

In a R&D environment driven by consumer expectations, the development and integration of metrics and algorithms is defined by a relatively rigid work flow. At each stage of the development, validation is required (be it subjective or objective) to ensure that the system functionally meets specified targets. These validation tests require the use of appropriate video content of which the amount and type required is dependent on the application under evaluation and the specific stage of the workflow.

The selection of appropriate test video sequences is a complex problem and is generally hampered by: (1) lack of the necessary expertise in applying relevant criteria to the selection of test videos, (2) videos selected may not be fully relevant to the video space targeted but only represent the best choice available, and (3) lack of appropriate videos that are copyright free or available for the intended use. These problems create difficulties in the proper benchmarking and comparison of new technologies and products. Driven by the fact that the tried and true sequences from a decade ago have out lived their use a new system must be implemented to propel R&D forward.

The consumer digital video library (CDVL) is a concept and proposal to standardize video content necessary for validation purposes in video processing and video quality research [1]. Categorization on multiple dimensions (i.e. temporal, space, color, and format properties) plus usage criteria will enable users to efficiently benchmark, collaborate, and advance research and development in a coordinated and effective manner. The CDVL concept also allows sustained improvement and evolution of the methodology based on categorization and usage criteria.

All video content would be housed in a central location and interfaced via an online virtual digital library. The library would support three key functions, an online repository of videos containing updated information on video content, provide tools for users to manipulate and process the videos, and provide users the ability to contribute to the library and collaborate in research.

Central to this proposal for a standard suite of videos is the specification, selection, and categorization method for video content recommendations. Thus, there will be an urgent need to bring together and seek consensus among experts in the field. Initially, we will define a simple method for content selection which is based on three criteria (video processes, developmental stage, and test type) and three content descriptors (color richness, spatial complexity, and temporal complexity).

In the following sections we discuss the usage rules that will guide users in selecting content, the content descriptors that span the search space, the content selection process leading to usage recommendations and guidelines. In the last section we present the prototype of the CDVL that has been implemented and close with some concluding remarks.

2. USAGE RULES

The objective of the usage rules is to provide users with a guide in which to carry out video content selection and classification. Once the usage rules are defined, a specific content space can be determined in which videos are selected for specific tasks. In essence, the rules allow filtering a large set of available video contents into a specific set of videos that are relevant to the task at hand. The creation and definition of a set of usage rules requires that several factors are accounted for, namely processing algorithm under test, developmental stage, and test type.

2.1. Video Processing

Video processing can essentially be classified into four broad categories, distribution, corrective, enhancement, and conditioning:

1. Distribution − manipulation of the video for delivery, i.e. compression, transmission.
2. Corrective − deals with noise and artifacts, i.e. analog noise reduction, deblocking.
3. Enhancement − create more appealing videos, i.e. sharpening, color saturation, skin tone correction.
4. Conditioning − allow conversion to different formats, i.e. scaling, de-interlacing.

Since each of these processes target different aspects of videos, no individual video or test pattern exists that is appropriate for the test of all processing types. For example, color enhancement targets the user's perception of the presented colors and their richness and requires that the video content selection consist of a wide variety of deep colors. Scaling affects the sharpness of videos and consequently requires video content that contain a variety of spatial and temporal complexity. A video appropriate for the validation of color enhancement may not necessarily be suitable for rescaling. Furthermore, targeted video material allows for in-depth investigation of a given algorithm.

2.2. Developmental Stage

The various stages in the development of a product or algorithm can be broadly classified into the following stages, exploration, prototyping, integration, and production:

1. Exploration − quick tests of various concepts.
2. Prototyping − longer tests to verify proof of concept.
3. Integration − longer tests to ensure that algorithm works as part of a larger system.
4. Production − tests required to verify that system is ready for deployment.

Each of these stages may require different types and numbers of videos. For example, during the exploratory stage, only a small number of critical videos are necessary as the time frame is short and the validation requirements are less stringent. In the integration phase, a large number of videos are required to benchmark the system to ensure that all functions operate as expected. The length of videos used at each stage may also vary since in the exploration and prototyping stages processing might be cumbersome and time consuming, while at the integration and production stages real time processing would allow for sequences that are longer and test the stability of the processing.

2.3. Test Type

The two main types of validation tests are the (1) subjective and (2) objective tests. The subjective test is essentially a psychophysical experiment that consists of human subjects making subjective evaluations on the effectiveness of an algorithm. Psychophysical experiments are extremely popular in the development of video metrics and algorithms as they provide important data from the ultimate source, the consumer. However, because of time and cost issues, care must be exercised in the selection of test videos. Most developers have access to expert assessors and their skills can be leveraged at any of the stages of development. Experts are able to accurately determine a real user's reaction without the cost or statistical reliability. Non-Expert testing is usually implemented at two if not one stage of product development before release. Objective tests are metric based tests where processed videos are measured using a metric and benchmarked. These objective metrics may consist of mature metrics such as PSNR [2,3] that do not perform as well as metrics that have been developed to correspond closely to the human visual system. However, they still play an important role in the automatic detection of catastrophic failures, i.e. buffer overflow in video encoding resulting in missing blocks and frame freezes.

3. CONTENT DESCRIPTORS

Once a set of content descriptors are defined that adequately describes the entire video content space, it becomes a simple task to reduce the entire space to a smaller content space that is appropriate for the validation tasks. We have defined three content descriptors for content selection: color richness, spatial complexity, and temporal complexity. Each of these descriptors is associated with a small number of statistical parameters extracted from the video.

Note that the descriptors that are presented in the following sections were not the product of rigorous tests, i.e. no psychophysical experiments and statistical analysis were carried out. They were developed to aid in the conceptualization, development, implementation and demonstration of the digital library. Thus, in the future it would be necessary to pursue the development of these descriptors pragmatically and rigorously.

3.1. Color Richness

Color richness in the context of the Digital Library is defined as the variety and saturation of the colors present in the digital video. We used simple statistical measures extracted from the pixel distributions from an appropriate color–space. The color–space used in the computation of color statistics was the CIE 1976 ($L^*u^*v^*$) color–space (*CIELUV*) [4]. The *CIELUV* color–space is approximately perceptually uniform and consists of the three perceptual axes L^*, u^*, and v^*. Here, L^* specifies brightness of colors while u^* and v^* specifies color location.

The color richness statistic, C, used was computed as the product of the standard deviations along the three perceptual axes, L^*, u^*, and v^*. The statistic was computed for each frame and averaged across all frames to obtain a single color descriptor, C_{ave}. This descriptor essentially represents a three dimensional volume describing the extent of the average color variation in the video. Thus, larger values of C_{ave} indicate high color richness. Note that the assumption here is that that the color content of the videos remains relatively stable throughout the sequence of frames, and this should be visually checked before final selection.

3.2. Spatial Complexity

Spatial complexity is defined here as the variability of pixel intensities, the texture complexity, and the strength and prevalence of edges contained in the video's spatial content. There exists many metrics and algorithms that are able to extract spatial complexity information from an image or video. Most of these measures operate on the luminance component of an image as it has been found that using additional information from the color components do not provide significantly improved estimates of spatial complexity.

We have continued along this line and propose to use a combination of three metrics that in totality provide a reasonable estimate of an image's spatial complexity. The individual statistics are entropy, a blurriness metric, and a spatial activity metric.

The entropy of an image is a measure of the variability of the pixel intensities in an image and is often used to describe the textures of images [5]. The blurriness metric we have used was developed by Marziliano et.al [6] which was shown to correlate well with perceived blur [7]. The blurriness metric is based on the average strength of edges located by a Sobel edge detector. The spatial activity metric we have used was developed by Wolf and Pinson which essentially measures the standard deviation of an edge enhanced image with particular attention paid to diagonal orientations [8]. These three metrics were computed for individual frames and multiplied together to form the frame spatial complexity metric, S. Finally, as in the case of color richness, the individual S values were averaged across all frames to yield a single spatial complexity metric, S_{ave}.

3.3. Temporal Complexity

There are many simple statistical measures that can be used to describe the motion content present in digital videos. However, the simplest forms are based on computing statistics from the motion vectors of the videos. In the context of the Digital Library, we have used the following definition of temporal complexity: "the degree of difficulty in tracking the motion content in the video." Thus, relatively high motion may not seem complex if the motion is smooth, whereas, slightly lower intensity motion may seem complex if there are multiple objects traveling at various speeds.

To account for the definition of motion complexity used, we defined a simple temporal complexity statistic, T. The statistic was computed as the product of the four statistical measures, mean, median, standard deviation, and maximum value after removing the top 5% obtained from the magnitude of the videos motion vectors as tested by Peker and Divakaran [9]. The motion vectors were computed using the 4-step method published by Po and Ma [10]. As in the case of color richness, the individual T values were averaged across all frames to yield a single temporal complexity metric, T_{ave}

Each of these four statistics provides a different intuitive meaning. For example, the mean value represents the intensity of the overall motion activity in the video assuming that high motion intensity correlates with high perceived motion activity. The median value provides a sense of the motion intensity of the middle level of motion in the video, essentially filtering out any spurious motion or noise. The standard deviation provides a sense of the variety of the motion intensity in the video where it is assumed that non–homogeneous motion would appear to be more complex. Finally, the maximum value provides a sense of the intensity of objects with in the video content that are moving the fastest. Removing the top 5% of motion vector magnitudes is an attempt to attenuate the effect of noise or very small objects moving at a high rate of speed.

3.4. Threshold Selection

As stated earlier, the content descriptors described were not selected from a formal procedure. However, it was still necessary to determine levels within the descriptors. Thus, we informally selected thresholds based on a group of 50 videos. Three coarse ranges within each descriptor were selected and labeled as: low, medium, and high.

4. CONTENT SELECTION PROECDURE

Growth of the Digital Library will depend on an efficient and effective content selection method. The method for content selection is to be fully developed through a collaborative effort by experts in the field. The plan of attack is to use both an objective and a subjective selection procedure in tandem. The objective scheme serves to reduce a large set of videos into a smaller set of videos which are then presented to a panel of experts that select the video contents through a group decision process.

4.1. Objective Procedure

The objective procedure for video content categorization would be based on the use of a fixed number of descriptors presented before, each describing a specific attribute. These descriptors would be used to visualize the video content space in a multi–dimensional space. The descriptors would be binned into several ranges that are either coarse or fine. This space would then be evenly populated with video contents to ensure that the content space enjoys appropriate coverage. Based on the target applications, video content would be selected from the space using rules defined by experts.

4.2. Subjective Procedure

The purpose of the subjective procedure is to develop the usage rules and criteria employed in the recommendation of video contents for specific applications. Knowledge and expertise comes into play in content selection that addresses the following needs:

1. Necessary tests that any algorithm must pass to be considered useful in its processing category.
2. Stress tests which allow establishing the level of performance of an algorithm.
3. Tests that allow ruling out undesired side effects of the processing.
4. The subjective criteria will be documented and available so that users will be able to verify and provide feedback on effectiveness and usefulness.

4.3. Content Space: An Example

Let us assume that the process to be validated is spatial up–scaling, the developmental stage is prototyping, and the test type is to be subjective. Thus the requirement dictates that a small number of critical videos are selected. Usage rules require that for the validation of spatial scaling algorithms, video content that have high spatial and temporal complexity are used. On a one dimensional scale, the videos would fall within the range indicated in by the gray bars in Figure 1.

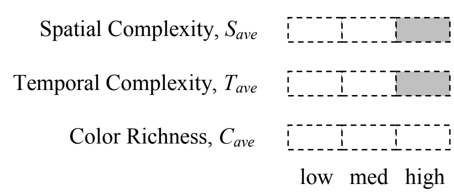


Figure 1: One dimensional representation of the range of values that would be targeted in content selection.

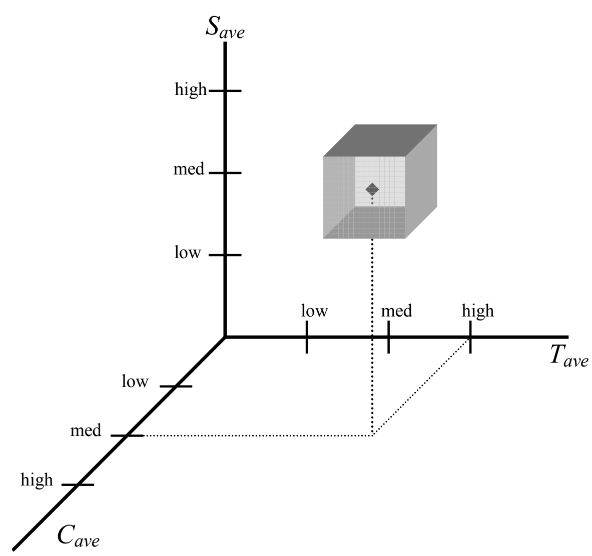


Figure 2: Three dimensional visualization of the space where content is to be selected from.

In a 3–dimensional space, where each of the content descriptors represents an axis, visualization of the content space is shown in Figure 2. From the available videos in the database, we would select a pre–determined number of videos that fall in the space described by the gray cubes in Figure 2. If a small number of videos are required, the space occupied by the cube can be shrunk to form a smaller cube containing the appropriate number of videos. Note that since color richness is not a factor in this scenario, we have taken a mid–level of color richness. Format and length constraints may further narrow down the search. Depending on the target performance, content for advanced stress tests may be left out.

5. PROTOTYPE

As a proof of concept, we developed a prototype of the digital library that can be accessed online at the address http://www.cdvl.chinkoh.com. The implementation of the prototype was based on a database model (Microsoft Access 2003) to allow the video content to be organized according to multiple criteria or categories, arranged in hierarchies or conceptual structures. The user interface, built using Microsoft FrontPage 2003, supports search and browse operations, an interface to add and modify content, and links to tools that users may use in processing or extracting statistics from videos. Snapshots of the user interface are provided in Appendix A.

The key component of the library was the Access database as it contained all pertinent information available on the video records. Information that was stored in the database included text descriptions, intrinsic properties, content descriptors, process history, source and rights, contributor information, and location. The records in the database were used to provide information to the user via a user interface built on server query language.

The user interface was developed as an online interactive web that comprised of multiple sections. The design of the user interface was based on three main guidelines, navigational simplicity, flexibility in browsing and searching for video content, and opportunity for users to contribute.

6. DYNAMIC LIBRARY

As an alternative to static content covering a multiplicity of cases, one could provide multiple processed versions from a small set of source contents in the form of a dynamic library.

In a dynamic library, users would specify the type of processed videos to be downloaded and all processing will be carried out within the library and sent as a stream. The implication of such a scheme is that more control over the type of processes would be exerted resulting in standardized test videos. Furthermore, with a small number of video contents, the dynamic library would be able to provide a much larger number of test videos than in the static case. The drawback is the need to have a complete and powerful set of processing tools and resources.

There are many caveats to such a system, and the biggest problem would be bandwidth considerations. With the static library, memory is the main problem as content is expected to grow over time. Bandwidth is also an important consideration as large formats and long sequences will have to be provided as well.

With the vast number of video processing algorithms and compression codecs available, supporting all possible combinations would be difficult. Additionally, the maintenance of the library would also be a problem as there would be a need to constantly update the system with the latest tools and settings to ensure that a comprehensive set of videos are available.

Many of these problems can be avoided by simply defining practical and manageable goals for the system and user base while insuring that the benefits will be attainable. Perhaps, the most important benefit of a dynamic digital library is that the videos and test conditions would be standardized allowing users to better communicate research and developmental results.

7. CONCLUDING REMARKS

In this paper, we have discussed the need for and concept of a digital video library to support consumer video research and development. In the initial stages of the creation of the digital video library, our role would be that of a facilitator where the concepts of the digital library are first drafted. The concept along with a rudimentary set of usage rules would then be further developed through a collaborative effort with experts in all aspects of video. Our near term goals are to further develop the content descriptors described in Section 2 and to populate the digital library with a small but highly relevant number of videos including the related usage rules. Our long term goal is to push on towards standardization of test videos and the implied methodology.

8. ACKNOWLEDGEMENT

This work was supported by a University of California MICRO grant with matching support from Intel Corporation. We are also grateful to Philip Corriveau (Intel Corporation) for his constructive remarks in finalizing the paper.

9. REFERENCES

[1] C.C. Koh and S.K. Mitra, "A proposed framework for a digital video library for research applications," *Proc. of the VPQM*, Scottsdale, Az, January 2006.

[2] Z. Wang, H.R. Sheikh, and A.C. Bovik, "Objective video quality assessment," *The Handbook of Video Databases: Design and Applications*, B. Furht and O. Marqure, ed., CRC Press, pp. 1041-1078, Boca Raton, FL, September 2003.

[3] B. Girod, "What's wrong with mean-squared error", *Digital Images and Human Vision*, A.B. Watson (Ed.), Pp. 207-220, MIT Press, Cambridge, MA, 1993.

[4] G. Wyszecki and W.S. Styles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., John Wiley & Sons, Inc., New York, 1982.

[5] R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 2nd Edition, Prentice Hall, Upper Saddle River, NJ, 2002.

[6] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "Perceptual blur and ringing metrics: application to JPEG2000," *Signal Processing: Image Communication*, vol. 19, iss. 2 , pp. 163-172, February 2004,

[7] A. Leontaris and A.R. Reibman, "Comparison of blocking and blurring metrics for video compression," *IEEE Proc.: ICASSP*, vol. 2, pp. 585–588, March 2005.

[8] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system," *SPIE Proc.: Int. Symp. on Voice, Video, and Data Communication*, Sep. 1999.

[9] K.A. Peker, A. Divakaran, T.V. Papathomas, "Automatic measurement of intensity of motion activity of video segments," *SPIE Proc.: Storage and Retrieval for Media Databases*, Vol. 4315, pp. 341-351, San Jose, CA, 2001.

[10] L-M. Po, W-C. Ma "A Novel Four Step Search Algorithm For Fast Block Motion Estimation" *IEEE Trans.: Circuits and Systems for Video Technology*, vol. 6, no.3, pp 313-7, June 1996.

APPENDIX A. PROTOTYPE CDVL

The prototype CDVL consists of several sections and includes: Home, Browse, Search, Definitions, Documents, Tools, Contribute, Contact, and Admin. Most of these sections are simple HTML based pages and are simple to build. Other pages are based on ASP code to allow access to the database of records. In the following, we provide snapshots of a number of the pages.

In Figure 3, we have the entry point into the CDVL. The purpose of this page is to provide a quick introduction to the purpose of the CDVL and navigation to other sections of the library.

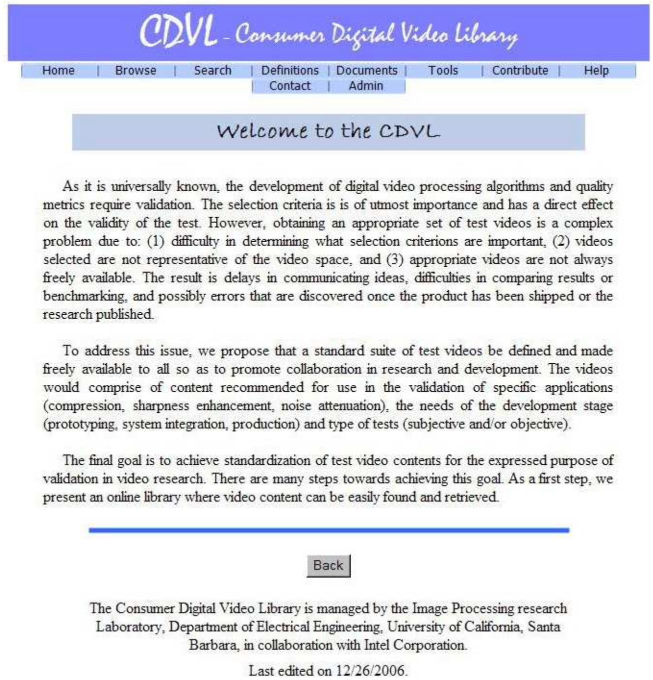


Figure 3: The introductory page for the CDVL.

The search section of the CDVL consists of two parts, a search form (Figure 4) and the results table (Figure 5). The search form allows users to quickly filter the available videos in the database using a set of user defined parameters. The results table provides a list of records that meets the user specified parameters. Additionally, the table contains a small amount of information, including thumbnails of the videos, available for each video record.

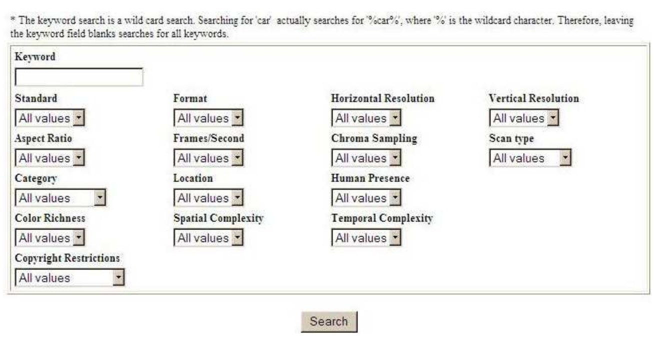
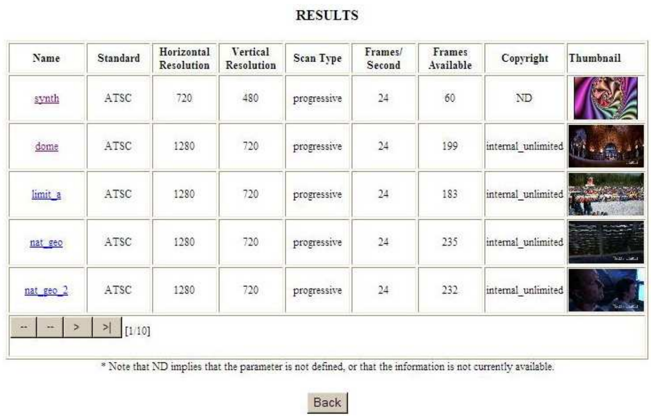


Figure 4: The form used to search the database of video records.



RESULTS

Name	Standard	Horizontal Resolution	Vertical Resolution	Scan Type	Frames/ Second	Frames Available	Copyright	Thumbnail
synth	ATSC	720	480	progressive	24	60	ND	
dome	ATSC	1280	720	progressive	24	199	internal_unlimited	
limit_a	ATSC	1280	720	progressive	24	183	internal_unlimited	
nat_geo	ATSC	1280	720	progressive	24	235	internal_unlimited	
nat_geo_2	ATSC	1280	720	progressive	24	232	internal_unlimited	

-- -- > >| [1/10]

* Note that ND implies that the parameter is not defined, or that the information is not currently available.

Back

Figure 5: The returned results for a search example entered into the form of Figure 4.

In Figure 6, we have a page containing all known information for a specific video. This page is linked directly from hyperlinks in the results table of Figure 5. At the top of the page we are presented with a frame from the video sequence. Directly below the image are links to a low resolution compressed version of the video sequence and the full resolution unprocessed video. The page is further divided into nine sections defining all known attributes of the video.

Available information for selected video

PREVIEW | VIDEO LOCATION

Description

Name	dome
Title	Inside a mosque
Description	Zooming in from a distance onto a gorup of four people inside a ornately decorated mosque. The interior is dark and the color theme is focused on gold. There is alot of high frequency content and the contrast is low. There is a WMV HD logo on the lower right hand corner.
Category	human
Location	indoor
Human Presence	few

Properties

Standard	ATSC
Format	720p
Horizontal resolution	1280
Vertical resolution	720
Aspect ratio	ND
Scan type	progressive
Chroma Sampling	ND
Frames per second	24
Available frames	199

Content Descriptors

Color richness	low
Spatial complexity	med
Temporal complexity	low

Process History

Distributive	WMV9
Corrective	ND
Enhancement	ND
Conversion	ND

Recommended Processes

Process 1	
Process 2	

Source and Rights

Source	WMV HD Content Showcase
Copyright	internal_unlimited

Contributor Information

Name	Chin Koh
Email	chin.chye.koh@intel.com
Organization	Intel Chandler
Comments	ND (this entry is probably not necessary and should be deleted)

Location Information

Sample Frame	\\cev-mvms001\Arch-Content\HD\720p\dome\dome_org\dome.JPG
Preview Clip	\\cev-mvms001\Arch-Content\HD\720p\dome\dome_org\dome.avi
Video	\\cev-mvms001\Arch-Content\HD\720p\dome\dome_org

Release Status

Release	released

Figure 6: Video record with all known information.