INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO-IEC/JTC1/SC29/WG11
CODED REPRESENTATION OF PICTURE AND AUDIO INFORMATION

ISO-IEC/JTC1/SG29/WG11
MPEG 93/516
AVC-509
July 1993

Title: **Flexible Object Oriented Decoding Using Slice Identification**

Source: Thomas Sikora
Monash University, Melbourne, Australia,
on behalf of Australian "UVC" Consortium

Purpose: Proposal

Abstract:

This contribution proposes to extend the MPEG 2 slice header with a "slice_id" to allow identification and separate decoding of "objects" at the receiver. An "object" in this context, flexible in shape and location, is formed by parts of a number of consecutive slices having an identical identification number. This functionality is particularly useful for interactive Multimedia applications but should also have advantages for other applications such as Multipoint Videoconferencing, intelligent search through databases and related topics.

1. Object Oriented Decoding

It is generally expected, that many future video applications will allow a high degree of interactivity between the user and the video source (independent on whether the video source is another Multipoint Videoconferencing participant or whether the video is stored on CD-ROM, interactive CD or on a database in general). For the user it is of interest to have a certain degree of control over the depth or detail of information presented to him. This functionality can be assisted through coding techniques which allow the separate decoding of parts of images (objects of interest) under the control of the user.

The possibility for object oriented decoding can be readily incorporated into an MPEG2 coding scheme by assigning a slice identification number (slice_id) to each slice belonging to a particular object. Since the slice size can be varied with every slice it is possible to define "objects" which may have any shape and even the shape and the location of the object can vary from frame to frame. The idea is outlined in Figures 1, 2 and 3 for different applications envisaged.

Figure 1 depicts a possible Multipoint Videoconferencing situation with four participants and four objects defined. In this application the objects have the shape of rectangular windows which do not vary its size or location from one frame to another. Each object consists of slices with identical slice_id and it is assumed that the video is encoded in a way that allows the separate and complete decoding of each object individually. The receiver has the choice to decode the entire full resolution video or to decode just one or more objects at a time dependent on decoder and/or display capability or personal preference without decoding the full bit stream. The receiver retains control over the information he or she prefers.

Figure 2 outlines an application with two objects, object 1 being the background and object 2 a window of fixed size but with variable location from one frame to another. In this application object 2 may be seen as a window which provides a pan within the original frame to track a particular movement (leaping jaguar, escaping individual, movie in a movie, etc.). The viewer may only want to decode object 2 for particular purposes (to save access costs to a database, to only store object 2 on its own database to save storage space etc.). In combination with a database the viewer may request from a server only the parts of a sequence belonging to object 2 to quickly search for particular patterns or individuals.

Figure 3 depicts an application where both the shape and the location of particular objects may vary from frame to frame.

For the encoder the complexity of the task to identify or define particular objects varies and is dependent on the application. While the scenario outlined in the Figures may be readily achieved with a standard MPEG2 encoder, the applications in Figures 2 and 3 may require an encoder of higher complexity. However, the simplicity of the decoder is retained. It should be emphasized, that for some applications, i.e. database Multimedia retrieval applications, the encoder complexity is of secondary concern and even non real-time encoding may be tolerated to achieve a gain in functionality.

2. Proposed MPEG2 Syntax Extension

The functionality to decode separate objects can be provided be assigning a slice_id number to each slice. We propose the following slice header extension:

	No. of bits	Mnemonics
slice() {		
slice_start_code	32	bslf
slice_id	4	uimsbf ||***Proposed Extension***
quantizer_scale	5	uimsbf
.....		
.....		
}		

Slice_id can vary from 1,2,... 15. Note, that in the "normal" case where the entire frame belongs to one object all slices have slice_id = 1.

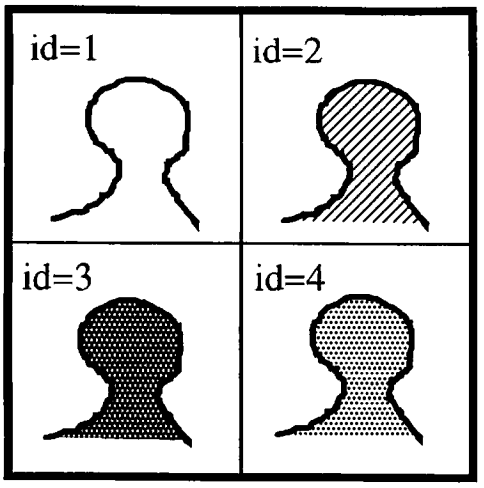

Figure 1: 4 objects,const. in shape and constant location over time

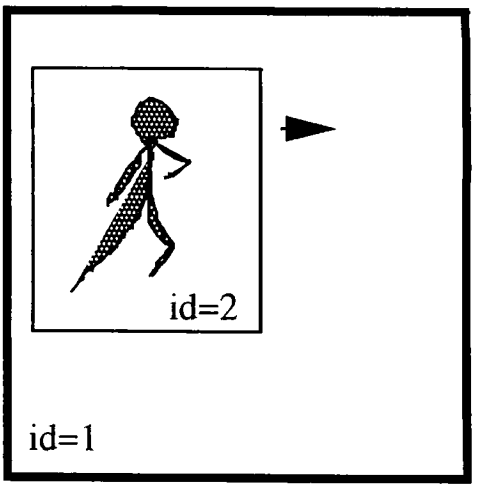

Figure 2: 2 objects,object 2 const. in shape but variable location over time

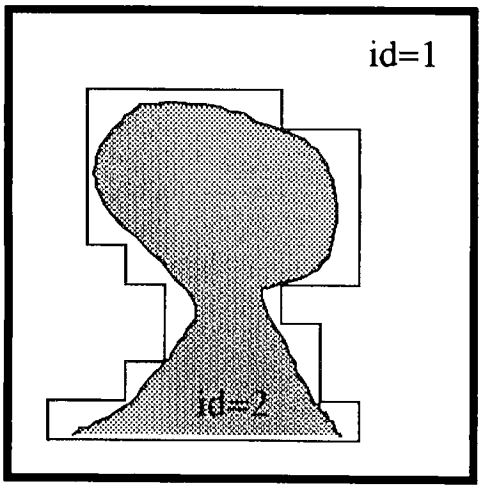

Figure 3: 2 objects,object 2 variable. in shape and variable location over time