A Modern Day Video Flip-Book: Creating a Printable Representation from Time-Based Media

Berna Erol, Jamey Graham, Jonathan J. Hull, and Peter E. Hart
RICOH Innovations, California Research Center
2882 Sand Hill Road, 94025 Menlo Park, CA, USA
{berna_erol, jamey, hull, hart}@rii.ricoh.com

ABSTRACT

In this paper, we describe a method for storing an entire video, animation sequence, or any other media type, on paper. The method is based on printing a key frame from a video on paper along with a barcode that encodes the motion information and other auxiliary information in MPEG-4 format. Unlike other video barcode systems in the prior art, a barcode in our system does not contain a link to the video clip; instead it contains motion information. A client device applies the motion information to an image of the video key frame to obtain full motion video. Therefore, the paper document is a self contained representation of a video clip and access to a server is not required. We modified an MPEG-4 [1] encoder and decoder to implement the video flip-book encoder and decoder. Experiments show that it is possible to encode several seconds of video on paper using our method. This is sufficient to create small animations for some printed materials such as video greeting cards and children's books.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]:Video, animations.

General Terms

Design, human factors, experimentation.

MM'07, September 23–28, 2007, Augsburg, Bavaria, Germany.
Copyright 2007 ACM 978-1-59593-701-8/07/0009...$5.00.

1. INTRODUCTION

Paper is an excellent display medium for photos because of its high resolution, ease of handling, and no power consumption. One obvious method of printing video on paper is the flip-book. However the effort required for constructing, distributing, and rendering video on flip-books is quite high. Also, the large amount of paper used for the construction of these books makes them impractical.

In this paper we describe a method for storing an entire media clip on paper. The method is based on encoding time-based media, such as video, in a still image format that is printable. The still image based representation is obtained by separating a time-based media object into two components: 2D reference information and auxiliary information. For example, if the media object is a video sequence, the reference information is a key frame and the auxiliary information is motion vectors. This is illustrated in Figure 1. A key frame from a video clip is already a 2D still image that is printable. Motion information, which is binary data, is encoded in a barcode format to obtain a still image representation. When these two representations, the key frame and the barcode, are combined, a complete 2D visual representation of the media object is obtained. This can be printed on paper to obtain a still image representation of the time-based media. The decoder on the client device captures this image representation, i.e., key frame and barcode, segments the key frame, decodes the motion information in the barcode, and constructs the video sequence by applying motion to the key frame. The barcode can also include additional binary information such as audio and animations.

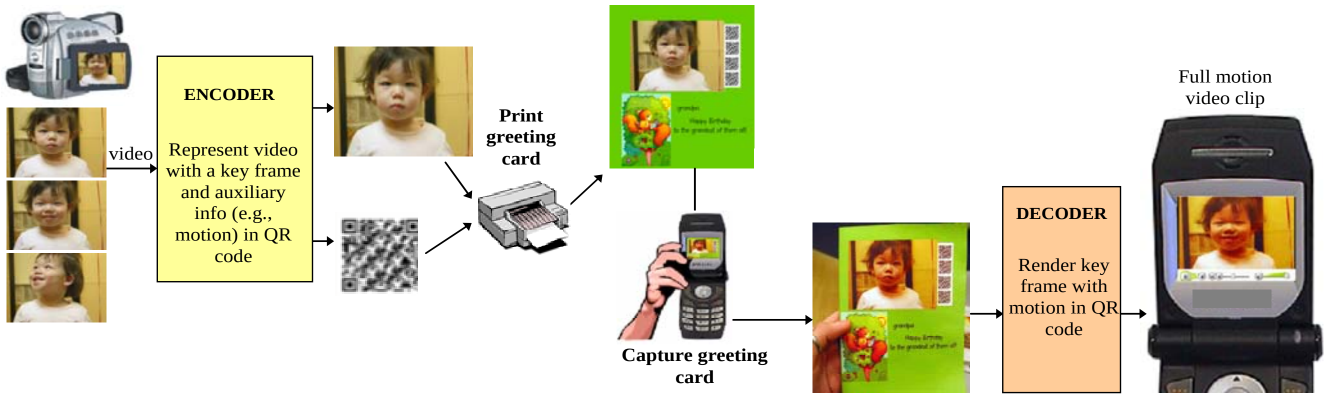
video
ENCODER
Represent video with a key frame and auxiliary info (e.g., motion) in QR code
Print greeting card
Capture greeting card
DECODER
Render key frame with motion in QR code
Full motion video clip

Figure 1. Process of creating a self contained still image representation for video and playback without accessing a server.

Our experiments showed that it is possible to encode several seconds of video on paper without needing to access a server for playback. Such video clips can be used on greeting cards, children's books, and video albums. Since maintaining a video server is not required, users are able to playback video sequences as long as they have the physical paper and the decoder. Moreover, since the users must have the paper to access to the video, the system is inherently secure.

2. RELEVANT RESEARCH

There has been a lot of effort on representing time-based media in a printable static image form. Representation with key frames that summarize the salient content of video has been investigated extensively, resulting in multi-layer representations such as Video Manga [2]. Using such a technique, key frames are printed on paper as a representation of video. However, in this case the motion information is lost. Printing several key frames and morphing between key frames [3] is also possible in order to obtain a video effect, but many key frames need to be printed and other binary information such as audio cannot be represented as a static image.

Another attempt at utilizing paper for video display is Video Paper [4][5]. Video Paper contains key frames from video clips and provides random access to video with barcodes. In Video Paper, the barcode stores only the links to the video data and does not attempt to store an entire video sequence on paper. Therefore, access to a server is needed to play back the actual video sequence, where in our method it is not needed.

There is also some relevant work in the video compression area, mostly in error recovery of reference frames where all the bits for reference frames are not available. In most of those cases, some part of a reference frame is available and used for reconstructing the missing parts of the reference frame [6]. In multiple description coding video schemes, there could be more than one video stream representing the same video content. If the reference frame is missing from one of the streams, the other video streams can be used for replacing the missing reference frame [7]. However, these error recovery methods reconstruct reference frames from the bit stream, where as in our method a missing reference frame is constructed solely from a picture of a printed key frame.

3. SYSTEM DESCRIPTION

Many different types of time-based media objects, such as video, audio, face, text, and other animations can be converted into a static representation. In this section we describe the encoding and decoding process when the media object type is a video sequence.

Figure 2 shows an overview of the process for generating a static representation for video. Our implementation is based on an MPEG-4 codec, however, any compression scheme that uses reference frames and predictive coding can be employed. First, a video sequence is MPEG-4 coded such that there is only one reference (I-) frame. The reference frame is already a 2D representation, so it can be printed. The MPEG-4 stream is then processed to take out the bits representing the I-frame, which we refer to here as an MPEG-4 bitstream*.

In order to represent the MPEG-4 bitstream*, we employ QR codes [8] that can encode up to 2,953 bytes of binary data. Depending on the barcode size and the maximum number of barcodes that can be printed, only a limited number of bits, BBITS, can be used to represent binary data. The barcode contains HEADER information besides MPEG-4 bits. The HEADER, which is of length HBITS, includes an identifier which is recognized by our application and other information such as a chroma key for segmentation and relative location of the reference frame. If the sum of MPEG-4 bitstream* bits and HBITS is larger than BBITS, MPEG-4 encoder parameters are set such that there is a larger quantization value for P frames and a reduced frame rate. The video clip is re-encoded until the desired MPEG-4 bitstream* length is achieved. At the end, the HEADER and MPEG-4 bitstream* are encoded in the QR code[8].

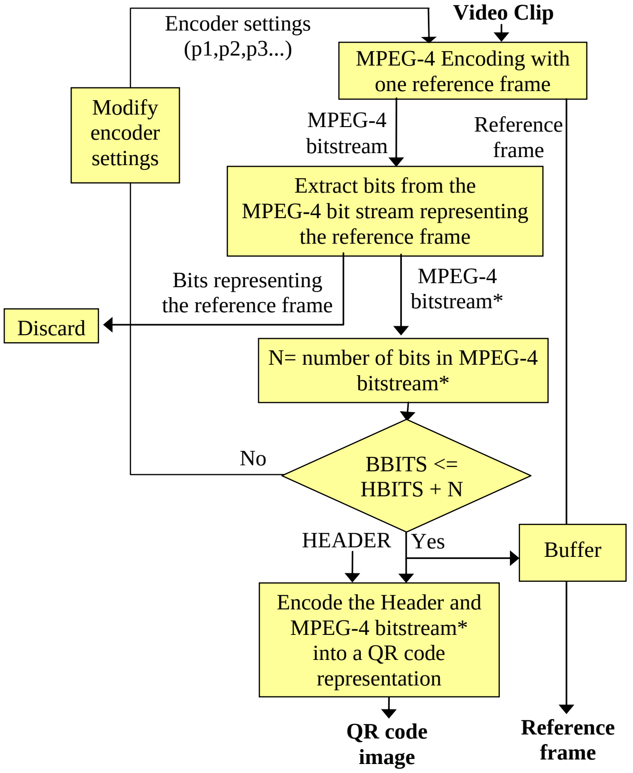
Video Clip
Encoder settings
(p1,p2,p3...)
MPEG-4 Encoding with one reference frame
Modify encoder settings
MPEG-4 bitstream
Reference frame
Extract bits from the MPEG-4 bit stream representing the reference frame
Bits representing the reference frame
MPEG-4 bitstream*
Discard
N= number of bits in MPEG-4 bitstream*
No
BBITS <= HBITS + N
HEADER
Yes
Buffer
Encode the Header and MPEG-4 bitstream* into a QR code representation
QR code image
Reference frame

Figure 2. Creation of a stand alone image-based representation from video.

An overview of the decoding and reconstruction process at the client device is presented in Figure 3. First, an image of the QR code and the reference frame is captured. Then the QR codes are located and decoded. If the QR code has an invalid header then the decoding is terminated, otherwise the MPEG-4 bitstream* headers are decoded to obtain video resolution.

Next, the reference frame is segmented from the captured image. Since all the motion and prediction error information is computed based on the reference frame, it is critical to correctly register the reference frame for good playback accuracy. Several methods can be employed for segmenting the reference frame. One method is to print distinct markers at the corners of the key frame as shown in Figure 4.a. Another method is to perform chroma keying by printing a unique color around the reference frame, as illustrated in Figure 4.b. Alternatively, as shown in Figure 4.c QR codes can be placed around the reference frame and the reference frame location is obtained from the barcode decoding software. The HEADER indicates the method of segmentation as well as a chroma key value, the shape of markers and other information required for segmentation.

The reference key frame is segmented based on the method indicated in the HEADER, scaled to the size of the video resolution and dewarped in order to obtain the best representation for the reference (I-) frame. The I-frame and MPEG-4 bitstream* is then passed to a modified MPEG-4 decoder. The modified decoder is implemented based on the MPEG-4 specification with the difference being that the bits encoding the first reference frame are not decoded from the bitstream but obtained from another source. As a result, motion vectors and prediction errors are applied onto the image captured from paper to obtain full motion video.

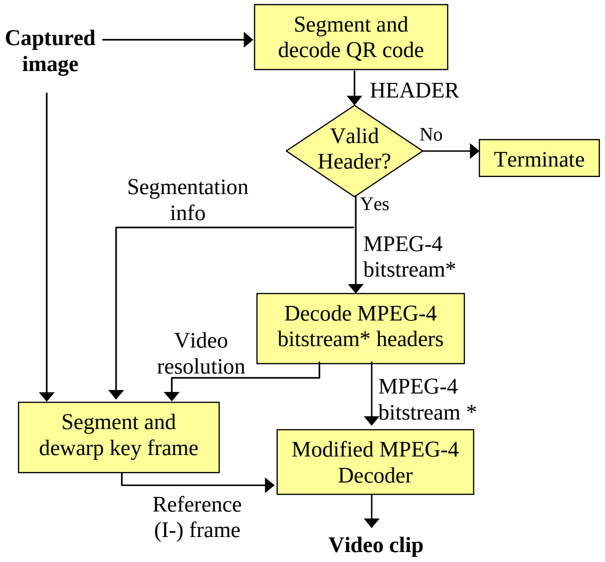

Figure 3. Decoding video via rendering of printed key frame.

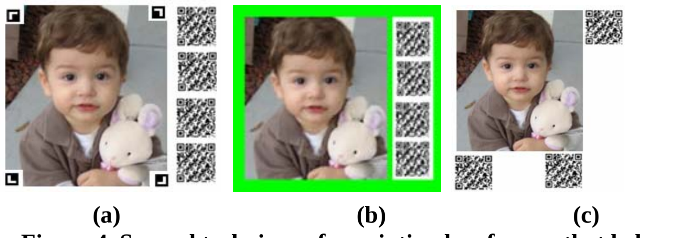

Figure 4. Several techniques for printing key frames that help registering the reference frame accurately: (a) Placing markers to the corner of the printed frame, (b) framing image with a chroma-key color, and (c) positioning QR codes at the edges of the frame.

4. EXPERIMENTS

We encoded two video clips, Akiyo and Smile, with an MPEG-4 video encoder. The video clips contained slow motion, mostly facial expressions. We envision that these types of video sequences, where most information is in the first frame, are more suitable for our technique compared to high motion video sequences, where most information content is in the motion and the prediction error.

The first two seconds of Akiyo sequence, shown in Figure 5, is coded at 180×120 resolution, at 3fps. When the MPEG-4 bitstream is analyzed, it is seen that 6 Kbytes is used for encoding the first frame, 0.3 Kbytes is used for encoding motion vectors and 2.5 Kbytes is used for encoding the prediction error. Using our method, the bits used for encoding the first frame can be removed and the video sequence can be represented with 0.3+2.5=2.8 Kbytes, which is only 30% of the coded video bits. This information can be encoded in a single QR code Version 40, which can encode up to 2,953 bytes of binary data [8].

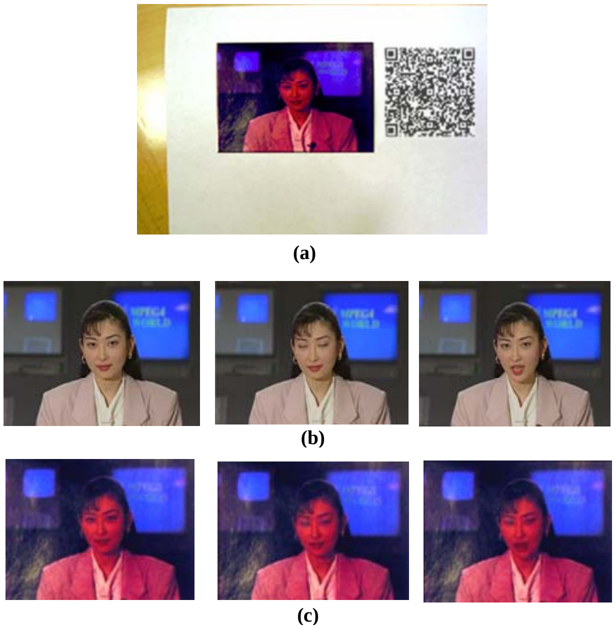

Figure 5. (a) Key frame printed on paper, (b) original Akiyo sequence, and (c) reconstructed sequence using printed key frame and motion information from the QR code.

Another sequence we encoded was 2-second Smile video, shown in Figure 6, that shows a baby smiling. The encoding was done at 4 fps at 320×240 resolution. The encoded MPEG-4 stream contained 6.6 Kbytes of texture bits for the first frame, 1.8 Kbytes of motion bits, and 0.9 Kbytes for the prediction error bits. If the key frame is encoded in the bitstream, 9.4 Kbytes is required for this data. When the first frame is not coded, then 2.7 Kbytes of side information is sufficient for representing the rest of the frames. Again, 2.7 Kbytes is easily representable with a single large QR code or several small QR codes.

The video sequences were printed on paper using our method and captured with a 2 MegaPixel digital camera as shown in Figure 5.a, Figure 6.a, and Figure 6.b. The MPEG-4 decoder is modified such that it uses the reference frame segmented from captured image as an I-frame. The original video sequences for Akiyo and Smile are shown in Figure 5.b and Figure 6.c, respectively. Figure 5.c shows the reconstructed frames of Akiyo sequence using our method. Figure 6.d and Figure 6.e shows the videos for Smile sequence reconstructed from two different captured images with varying lighting conditions. As can be seen, the Akiyo sequence has fine motion around the mouth and eye areas, and motion artifacts are noticeable. On the other hand, the reconstruction artifacts in the Smile clip are less apparent. These videos are also available for online viewing at [9].

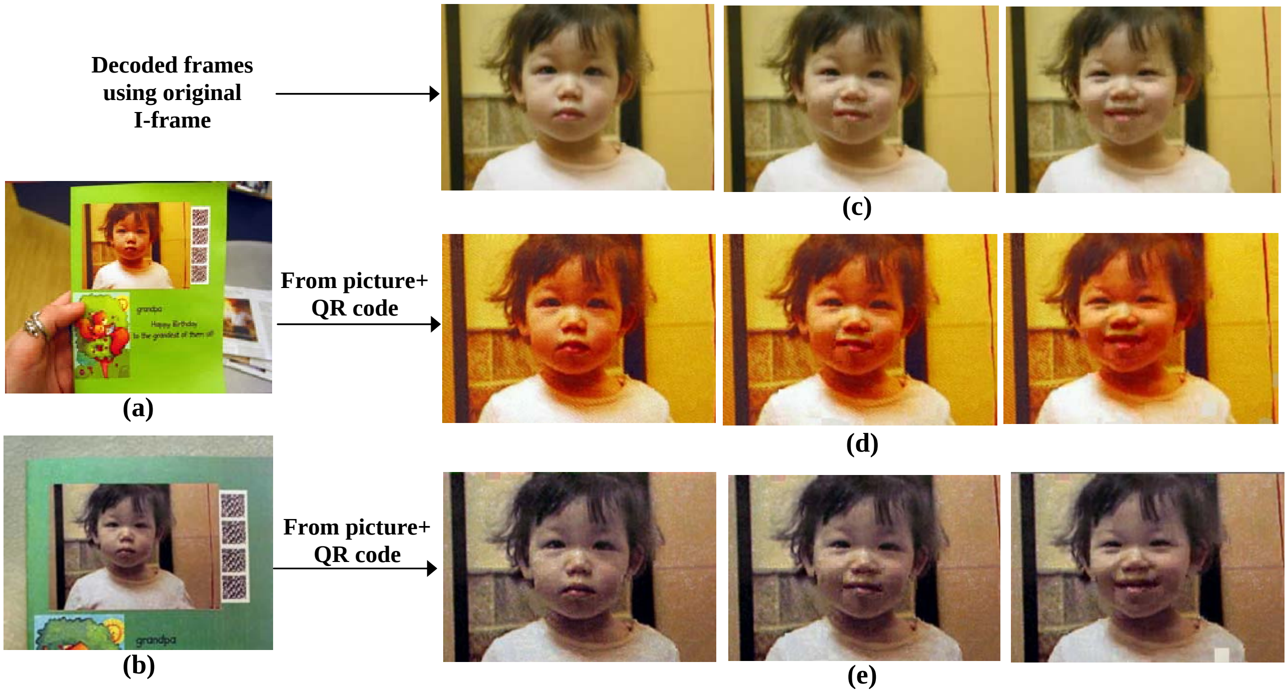
Decoded frames using original I-frame
From picture+ QR code
From picture+ QR code
grandpa
Happy Birthday to the grandest of them all!
grandpa
(a)
(b)
(c)
(d)
(e)

Figure 6. (a) (b) Key frames printed on paper, (c) original Smile sequence, and (d) (e) reconstructed sequences using printed key frames and motion information from the QR codes.

5. SUMMARY AND OUTLOOK

We presented a novel method for creating a stand-alone representation of time-based media in a printable still image format. An analog representation is used for the most information intensive part of the time-based media, e.g., the first key frame, and motion vectors and prediction errors are encoded in binary format and represented with a QR code. Our experiments showed that several seconds of video can be encoded on paper this way using a modified MPEG-4 codec. Only a flip-book decoder software is needed on the client device for the playback.

Such short video clips can be used to personalize greeting cards, create video albums, play small animations on books, and can be printed on products to demonstrate usage. Long video clips can be obtained by printing several key frames and larger QR codes. Instead of using the first frame as the reference frame, a key frame could be used that provides the minimum prediction.

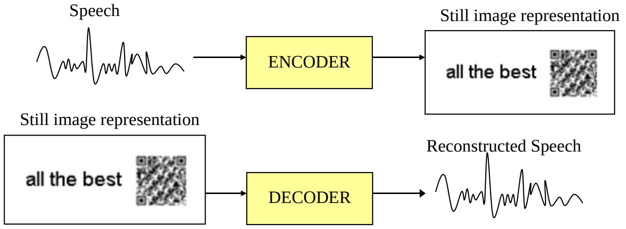

Figure 7. Speech to static image representation

The ideas presented in this paper can be extended to represent other time-based media such as music, speech, and animations. An audio message can be encoded in the QR code. Alternatively, speech can be printed in text and the user's vocal information (e.g., pitch, pronunciation) can be represented in the QR code as shown in Figure 7. At the decoder, OCR would recover the text and speech would be synthesized using the vocal information in the QR code. Music can be printed as musical notes, which can be recognized at the decoder, and the QR code can represent instrument and other audio information to synthesize music at the decoder.

6. REFERENCES

[1] ISO/IEC 14496-2, "'Information technology—Coding of audiovisual objects—Part 2: Visual", 2000.

[2] Uchihashi, S., Foote, J., Girhensohn, A., & Boreczky, J. "Video Manga: Generating Semantically Meaningful Video Summaries", ACM Multimedia Conf., pp. 383-392, 1999.

[3] T. Stich, M. Magnor, "Keyframe Animation from Video", Proc. IEEE ICIP, pp.2713-2716, 2006.

[4] Graham, J, Erol, B., Hull, J.J. and Lee, D.S., "The VideoPaper Multimedia Playback System", ACM Multimedia Conference, pp.94-94, 2003.

[5] Klemmer, S.R., Graham, J., Wolff, G.J., Landay, J.A., "Books with voices: paper transcripts as a physical interface to oral histories", ACM CHI, pp. 89-96, 2003.

[6] Bansal, P., Narendran, M.R., and Murali, M.N.K., "Improved error detection and localization techniques for MPEG-4 video", IEEE ICIP, 2002.

[7] Gallant, M. Shirani, S. Kossentini, F. , "Standard-compliant multiple description video coding", IEEE ICIP, 2001.

[8] ISO/IEC 18004, "Information Technology AIDC Techniques Bar code symbology QR Code," 2000.

[9] Examples at http://rii.ricoh.com/~berna/videoflipbook.html