J. Vis. Commun. Image R. 22 (2011) 332–344

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

MIJ2K: Enhanced video transmission based on conditional replenishment of JPEG2000 tiles with motion compensation

Alvaro Luis Bustamante *, José M. Molina López, Miguel A. Patricio

Univ. Carlos III de Madrid, Avda. Univ. Carlos III, 22, 28270 Colmenarejo, Madrid, Spain

ARTICLE INFO

Article history:
Received 26 March 2010
Accepted 15 February 2011
Available online 19 February 2011

Keywords:
JPEG2000
Streaming
rtp
Real-time
Motion
Tile
Interframe
Compensation

ABSTRACT

A video compressed as a sequence of JPEG2000 images can achieve the scalability, flexibility, and accessibility that is lacking in current predictive motion-compensated video coding standards. However, streaming JPEG2000-based sequences would consume considerably more bandwidth. With the aim of solving this problem, this paper describes a new patent pending method, called MIJ2K. MIJ2K reduces the inter-frame redundancy present in common JPEG2000 sequences (also called MJP2). We apply a real-time motion detection system to perform conditional tile replenishment. This will significantly reduce the bit rate necessary to transmit JPEG2000 video sequences, also improving their quality.

The MIJ2K technique can be used both to improve JPEG2000-based real-time video streaming services or as a new codec for video storage. MIJ2K relies on a fast motion compensation technique, especially designed for real-time video streaming purposes. In particular, we propose transmitting only the tiles that change in each JPEG2000 frame. This paper describes and evaluates the method proposed for real-time tile change detection, as well as the overall MIJ2K architecture.

We compare MIJ2K against other intra-frame codecs, like standard Motion JPEG2000, Motion JPEG, and the latest H.264-Intra, comparing performance in terms of compression ratio and video quality, measured by standard peak signal-to-noise ratio, structural similarity and visual quality metric metrics.

1. Introduction

Video communication over IP-based networks is becoming increasingly popular, and it has emerged as one of the most important applications using Internet technology. Motion JPEG2000 (MJP2) [14] is a video coding standard based on the JPEG2000 (also called J2K) image codec compression [13,26]. JPEG2000 employs an intra-frame coding technique that avoids the use of motion compensation adopted by most of the previous standards, such as MPEG-2 [12], or MPEG-4 [21]. Thus, the compression delay achieved by the MJP2 codec could be slightly shorter than such motion compensation-based techniques.

MPEG-2 and MPEG-4 or even Motion JPEG (MJPEG) are the most used and tested codecs. However, the MJP2 codec is now being integrated into new video surveillance devices and systems [4,9]. Compared to MPEG-based systems, the MJP2 codec can take advantage of JPEG2000s unequaled number of features. This standard provides error resilience, regions of interest (ROIs) definition, as well as spatial, component, resolution and quality scalability [5,22].

The bit-stream can be easily parsed and adapted in real-time in each of these scalabilities without having to decode frames. MJP2 is also the leading digital cinema standard currently supported by Digital Cinema Initiatives [7] (a consortium of most major studios and vendors) for the storage, distribution and exhibition of high-definition motion pictures. It is an open ISO standard and an advanced update of MJPEG, which was based on the legacy JPEG format [20].

So, it is expected to provide a better solution for applications that are required to stream high-quality and high-resolution videos over IP-based networks [8,25,16]. The application areas include: digital cinema, PC-based video capturing, remote surveillance, high-resolution medical, satellite imaging and so on.

Also, the MJP2 codec should be considered in real-time transmission systems because it does not employ any motion compensation or inter-frame compression. Instead, each frame is an independent entity encoded with JPEG2000 [13]. This feature will hugely reduce compression and transmission process delay as, contrary to motion compensation techniques, it can be transmitted immediately after an individual frame is compressed.

However, this low delay is achieved at the cost of increasing bandwidth requirements, since it does not reduce any temporal redundancy in videos (which is the main goal of motion compensation-based techniques).

There is not much work related to the optimization of real-time transmissions of JPEG2000 video sequences, at least from our point of view. Some researchers have tried to perform scene analysis to

* Corresponding author.
E-mail addresses: aluis@inf.uc3m.es (A.L. Bustamante), molina@ia.uc3m.es (J.M. Molina López), mpatrici@inf.uc3m.es (M.A. Patricio).

1047-3203/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jvcir.2011.02.002

reduce Motion JPEG2000 video surveillance delivery bandwidth and complexity [17]. This work separates foreground objects from the background, which it compresses at different qualities, selecting better qualities for the important objects in the sequence. The tests run in this case are insufficient, and the research paper does not set out a detailed process of how to detect background/foreground objects and how it is used in JPEG2000 sequences.

Other research, closer to the process proposed in this paper, is described in [18,19,6]. They perform a conditional replenishment of JPEG2000 code-blocks with motion compensation, but this technique appears to be too complex to be applied in real-time environments, since they work with a low-level code stream. Clients and servers would also have to be purposely designed for use with these techniques, and standard protocols such as described in RFC 5371 [10] designed for JPEG2000 transmission could not be used. Neither do they test the delay introduced by the techniques that they describe, and all results appear to be simulated, and not tested on a real implementation.

To solve this problem, we propose the Motion Inter-Frame JPEG2000 (MIJ2K) method. As described in this paper, it applies a real-time motion compensation technique to the MJP2 sequences before transmission. It will lead to a significant reduction in bandwidth requirements without adding extra delay to the compression and transmission process.

In previous research conducted by A.L. Bustamante et al. in [3] a system for real-time transmission of JPEG2000 live video streams over a real-time transport protocol (RTP) was applied in compliance with RFC 5371 [10]. It takes advantage of the JPEG2000 code-stream structure to perform intelligent transmission. In this paper, we also proposed a pre-transmission optimization of sequences, using a low-complexity motion detection system. MIJ2K codec will add a real-time inter-frame compression technique in a native intra-frame system. It combines the benefits of inter-frame techniques (greater compression ratio and higher quality) with the advantages of using an intra-frame method (low complexity compression and fast delivery).

Added to a common JPEG2000 video streaming system, the MIJ2K technique will hugely improve transmission, saving bandwidth, and achieving better qualities for the same bit-rate, as shown throughout this paper. This paper fully describes MIJ2K, including a real-time motion detection and compensation system. For streaming purposes, we use it over the J2K Streaming RTP developed in [3]. We compare MIJ2K against other standard codecs, like standard MJP2, MJPEG, and H.264-Intra. To evaluate the quality of MIJ2K, we will test the codec in terms of delivery delay in seconds; quality of video sequences measured by peak signal-to-noise ratio (PSNR), structural similarity (SSIM) and the visual quality metric (VQM); and compression ratio, measuring the average bit-rate for transmission against other codecs.

The proposed MIJ2K is patent pending and can be easily applied to the previous work done in [3]. Also it is RFC 5371 compliant.

The remainder of the paper is organized as follows. Section 2 provides background on JPEG2000 code streams. Section 3 introduces some fundamentals about video compression evaluation. Section 4 fully describes the proposed architecture, including the adopted real-time motion compensation technique. Section 5 is used to evaluate the quality of MIJ2K compared with other standard codecs. Finally, we present some conclusions about the proposed method.

2. JPEG2000 code stream

JPEG2000 is a wavelet-based [1] image compression standard, created by the Joint Photographic Experts Group committee in the year 2000, with the intention of superseding their original discrete cosine transform-based JPEG standard (dating from 1992).

Although JPEG2000 offers a modest increase in compression performance compared with JPEG, its main benefit is significant code-stream flexibility. The code stream obtained after compression of an image with JPEG2000 is scalable, meaning that it can be decoded in a number of ways. For instance, by truncating the code stream at any point, we can get a representation of the image at a lower resolution or signal-to-noise ratio. By ordering the code stream in various ways, applications can achieve significant performance increases.

Some of the characteristics of JPEG2000 images are:

- Superior compression performance: At high bit rates, where artifacts become nearly imperceptible, JPEG2000 has a small machine-measured fidelity advantage over JPEG. At lower bit rates (e.g., less than 0.25 bits/pixel for gray-scale images), JPEG 2000 has a much more significant advantage over certain JPEG modes: artifacts are less visible and there is almost no blocking. The compression gains over JPEG are attributed to the use of DWT and a more sophisticated entropy encoding scheme [23].
- Multiple resolution representation: JPEG2000 decomposes the image into a multiple resolution representation in the course of its compression process. This representation can be put to use for other image rendering purposes beyond compression as such [13].
- Progressive transmission by pixel and resolution accuracy, commonly referred to as progressive decoding and signal-to-noise ratio (SNR) scalability: JPEG2000 provides efficient code-stream organizations which are progressive by pixel accuracy and by image resolution (or by image size). This way, after a small part of the whole file has been received, the viewer can see a lower quality version of the final picture. The quality then improves progressively as more data bits are downloaded from the source. The 1991 JPEG standard also has a progressive transmission feature, but it is rarely used.
- Lossless and lossy compression: like JPEG 1991 [28], the JPEG2000 standard provides both lossless and lossy compression in a single compression architecture. Lossless compression is provided by the use of a reversible integer wavelet transform in JPEG2000.
- Random code-stream access and processing: JPEG2000 code streams offer several mechanisms to support spatial random access or region of interest access at varying degrees of granularity. This feature is achieved in part by the concept of tiling introduced in JPEG2000, where an image is split into so-called tiles, rectangular regions of the image that are transformed and encoded separately. For each encoded tile there are also other random-access mechanisms such as the concept of precincts.
- Error resilience: like JPEG 1991, JPEG2000 is robust to bit errors introduced by noisy communication channels because data is coded in relatively small independent blocks.
- Side channel spatial information: It fully supports transparency and alpha planes.

Apart from the above features, the main benefit of using JPEG2000 for video streaming is that, unlike other video compressors including MPEG-4, compression could be done in real-time [3] because it is an intra-frame codec. Thanks to this feature, JPEG2000 can be used in events that require real-time transmission, like video surveillance.

As far as our proposal is concerned, however, the main advantage is that the image can, optionally, be partitioned into smaller independent non-overlapped rectangular blocks called *tiles* [13]. We will exploit this exceptional feature, provided by this compressor alone, to perform real-time inter-frame compression using the proposed conditional tile replenishment method. We will employ a