Signal Processing: *Image Communication* 37 (2015) 47–57

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

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

Inter-prediction methods based on linear embedding for video compression

Martin Alain [a,b,*], Christine Guillemot [b], Dominique Thoreau [a], Philippe Guillotel [a]

[a] Technicolor, 975 Avenue des Champs Blancs, 35576 Cesson-Sevigne, France
[b] INRIA, 263 Avenue du General Leclerc, 35042, France

ARTICLE INFO

Article history:
Received 5 January 2015
Received in revised form
24 July 2015
Accepted 24 July 2015
Available online 1 August 2015

Keywords:
Locally linear embedding
Temporal prediction
Video compression

ABSTRACT

This paper considers the problem of temporal prediction for inter-frame coding of video sequences using locally linear embedding (LLE). LLE-based prediction, first considered for intra-frame prediction, computes the predictor as a linear combination of K nearest neighbors (K-NN) searched within one or several reference frames. The paper explores different K-NN search strategies in the context of temporal prediction, leading to several temporal predictor variants. The proposed methods are tested as extra inter-frame prediction modes in an H.264 codec, but the proposed concepts are still valid in HEVC. The results show that significant rate-distortion performance gains are obtained with respect to H.264 (up to 15.31% bit-rate saving).

1. Introduction

Most video coding standards achieve data compression by exploiting similarities within frames (i.e., the spatial redundancy), as well as between the target frame and one or several reference frames (i.e., the temporal redundancy). Intra-frame coding techniques are used to reduce the spatial redundancy within each frame separately, whereas inter-frame coding techniques are used to reduce the temporal redundancy between successive frames of a video sequence.

Intra-prediction is a powerful tool to remove spatial redundancy in intra-frame coding. In the H.264 video compression standard [1], each frame is partitioned into blocks, and for each block to be coded, a predictor block is created by extrapolating previously coded and reconstructed pixels surrounding the target block to be coded. Nine prediction modes have been defined which propagate surrounding pixels along different directions. In HEVC [2], the intra-frame prediction has been extended to support 33 directional prediction modes. The encoder selects the prediction mode which is the best in a rate-distortion (RD) sense, using a Lagrangian optimization technique, and signals the retained mode to the decoder.

Inter-frame predictors are typically obtained by motion estimation and compensation methods that match every block to be coded with a similar block in one or several reference frames, using the so-called block matching (BM) algorithm [1,2]. The position of the best matching block in a reference frame is signaled to the decoder by transmitting a motion vector. The motion vector may locate the best matching block with a fractional pixel (pel) accuracy thanks to fractional positions interpolation in the reference frames.

The motion estimation can also be performed using the so-called template matching (TM) technique [3]. The methods exploit the correlation between the current block and a pre-defined set of neighboring pixels, called the template of the current block. Rather than looking for the most correlated block in the reference frames, one looks

* Corresponding author. Tel.: +33 2 99 84 74 35.
E-mail address: martin.alain@irisa.fr (M. Alain).

http://dx.doi.org/10.1016/j.image.2015.07.011
0923-5965/

for the most correlated template. The block which is adjacent to this template is used as a predictor for the current block. The motion compensation is performed using the exact same process, so no motion information needs to be transmitted to the decoder. This technique efficiency has also been demonstrated for intra-frame prediction [4]. The RD performance of this method can be improved by using a weighted combination of multiple predictors. Initially a simple averaging of the predictors was performed [5,6], but methods using adaptive weights, e.g. using sparse approximation [7,8], were shown to bring significant improvements.

In this paper, we consider an approximation method called locally linear embedding (LLE), introduced in [9] for data dimensionality reduction, which we adapt to the problem of temporal prediction. The LLE technique has already been shown to be very efficient for intra-frame prediction in [10,11]. However, the derivation from intra-frame to inter-frame prediction is not trivial, mainly because the proposed techniques are now in competition with the motion estimation/compensation, which is a more efficient prediction tool than the intra-frame directional modes. The idea is to first search for a representation of the template as a linear combination of K templates (called K-NN templates) taken from a search window denoted as SW. The linear combination coefficients (or weights) are then applied on the blocks adjacent to the K-NN templates to yield the current block predictor. The LLE weights are computed using a least square formulation of the template approximation problem under the constraint that they sum to one.

The K-NN search strategy has a strong impact on the predictor quality. In fact, the TM technique efficiency relies on the hypothesis that the template and its adjacent block are well correlated. First, we proposed a direct derivation of the TM technique, where the K-NN can be found by computing distances between the template of the current block and those of candidate blocks in the reference frames. This method is denoted as template matching LLE (TM-LLE) and, as for the TM method, no side information (i.e., no motion vector) needs to be sent to the decoder. A variant of this method is introduced where the first neighbor is searched by template matching (as in TM-LLE), but the remaining $(K-1)$-NN are found by computing a distance between the complete patch formed by the template and adjacent block of the first neighbor and the candidate patches in the search window. The method is denoted as improved template matching LLE (ITM-LLE).

Second, to further improve the K-NN search, we introduce a method enforcing the correlation between the templates and their adjacent blocks, but requiring the transmission of side information to the decoder. Thus, we propose a method where the K-NN search is initialized with a block-matching algorithm. This implies that a motion vector is sent to the decoder. We then find the remaining $(K-1)$-NN as in ITM-LLE. This method is named as block-matching LLE (BM-LLE).

Finally, we propose an improved variant of the ITM-LLE method, denoted as optimized ITM-LLE (oITM-LLE). In this method, we basically obtain L predictors by running L times the ITM-LLE method. The best iteration in a RD sense is retained, and its index is sent to the decoder.

The experiments and their analysis focus on RD performance evaluations of the proposed prediction methods against the standard reference techniques: directional and motion estimated/compensated prediction modes of H.264 and template matching averaging (TM-A). This analysis is carried out using a legacy H.264 implementation, but note that the proposed techniques are still applicable in HEVC, since the inter-frame prediction tool in HEVC follows the same principles as those used in H.264. Simulation results show that significant RD performance improvements are achieved compared to the reference prediction methods. The performed analysis includes elements of complexity in terms of execution times measured at the encoder.

The rest of the paper is organized as follows. Section 2 reviews background on video compression methods (H.264, HEVC), as well as the TM-based prediction methods. Section 3 describes the proposed LLE-based temporal prediction techniques. Section 4 explains how the proposed prediction methods have been used in an H.264 codec and they could be used in HEVC. We then give the PSNR-rate performance gains compared to the reference H.264 codec. Section 5 sums up how the proposed techniques can be integrated in the HEVC codec and why they are still valid.

2. Temporal prediction: background

This section first summarizes the relevant features of the temporal prediction methods in the H.264 video compression scheme [1] and the corresponding techniques in HEVC [2]. It then briefly revises state-of-the-art temporal prediction methods based on template matching.

2.1. Motion-compensated inter-frame prediction in H.264 and HEVC

A motion vector (MV) is estimated for each block of the possible partitions (4×4 to 16×16, 16×8 in H.264, 8×8 to 64×64 in HEVC), usually using a BM algorithm. The MV establishes a correspondence between the current block and a block in one of the reference frames. This block is used as the predictor. The reference frames are stored in two buffers called L_0 and L_1 lists. The L_0 list contains reference frames from the past while the L_1 list contains reference frames from the future. For each block, the motion vector, the reference frame index and the list index are coded and sent to the decoder. The coding efficiency of the MVs relies on their predictive coding, under the assumption that the motion vector field is continuous (at least locally). Thus, in H.264, a motion vector predictor (MVP) is computed as the median of available neighboring MVs. Only the difference between the current MV and the MVP is then coded. In HEVC, the MVs coding efficiency is even improved by using the Adaptive Motion Vector Prediction list or the Merge list. The improved efficiency compared to H.264 for inter-frame prediction thus comes more from the optimization of the side information coding than from the prediction quality. In both H.264 and HEVC,