Signal Processing: Image Communication 60 (2018) 144–159

Contents lists available at ScienceDirect

Signal Processing: Image Communication

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

Light field image coding with jointly estimated self-similarity bi-prediction

Caroline Conti *, Paulo Nunes, Luís Ducla Soares

Instituto Universitário de Lisboa (ISCTE–IUL), Instituto de Telecomunicações, Lisbon, Portugal

ARTICLE INFO

Keywords:
Light field
Plenoptic
Holoscopic
Integral imaging
Image coding
HEVC
Self-similarity
Bi-Prediction

ABSTRACT

This paper proposes an efficient light field image coding (LFC) solution based on High Efficiency Video Coding (HEVC) and a novel Bi-prediction Self-Similarity (Bi-SS) estimation and compensation approach to efficiently explore the inherent non-local spatial correlation of this type of content, where two predictor blocks are jointly estimated from the same search window by using a locally optimal rate constrained algorithm. Moreover, a theoretical analysis of the proposed Bi-SS prediction is also presented, which shows that other non-local spatial prediction schemes proposed in literature are suboptimal in terms of Rate-Distortion (RD) performance and, for this reason, can be considered as restricted cases of the jointly estimated Bi-SS solution proposed here. These theoretical insights are shown to be consistent with the presented experimental results, and demonstrate that the proposed LFC scheme is able to outperform the benchmark solutions with significant gains with respect to HEVC (with up to 61.1% of bit savings) and other state-of-the-art LFC solutions in the literature (with up 16.9% of bit savings).

1. Introduction

Light Field (LF) imaging based on a single-tier camera equipped with a Microlens Array (MLA) – also known as holoscopic, plenoptic, and integral imaging – derives from the fundamentals of light field/radiance sampling [1], where not only the spatial information about the Three Dimensional (3D) scene is represented, but also the angular viewing direction, i.e., the “whole observable” scene.

Recently, LF imaging has become a prospective imaging approach for providing richer content capture, visualization, and manipulation, being applicable in many different areas of research, e.g., 3D television [2,3], biometric recognition [4], and medical imaging [5]. Among the advantages of employing an LF imaging system is the enabling of new degrees of freedom in terms of content production and manipulation, thus supporting functionalities not straightforwardly available in conventional imaging systems, namely, post-production refocusing, changing depth-of-field, and changing viewing perspective.

However, deploying LF image and video applications with its appealing functionalities will require the use of efficient coding schemes to deal with the large amount of data involved in such types of systems. In this context, novel initiatives on LF image and video coding standardization are also emerging. Notably, the Joint Photographic Experts Group (JPEG) committee has recently started the JPEG Pleno standardization initiative [6] that addresses representation and coding of emerging new imaging modalities. In addition, the Moving Picture Experts Group (MPEG) group has recently started a new work item on coded representations for immersive media (MPEG-I) [7].

1.1. Related work

Previous Light Field Coding (LFC) schemes available in the literature can be categorized in three main approaches: (i) based on transform coding [8,9], (ii) based on view extraction [10–17], and (iii) based on non-local spatial prediction [18–22]. Generally, all coding schemes try to take advantage of the particular planar intensity distribution of the LF image. Notably, as a result of the used optical system, the raw LF image corresponds to a 2D array of Micro-Images (MIs), where both light intensity and direction information are recorded.

1.1.1. LFC based on transform coding

Most of the early proposed LFC schemes adopted the transform-based approach by using a Discrete Cosine Transform (DCT) or a Discrete Wavelet Transform (DWT). In [8], a 3D DCT was applied to a stack of MIs to exploit the existing correlation between adjacent MIs, as well as the redundancy within each MI. In [9], the LF content was separated into various viewpoint images by extracting one pixel with the same position from each MI and a 3D DWT was then applied to a stack of them.

---

* Corresponding author.
E-mail address: caroline.conti@lx.it.pt (C. Conti).

https://doi.org/10.1016/j.image.2017.10.006
Received 14 February 2017; Received in revised form 11 October 2017; Accepted 11 October 2017

Afterwards, the lower frequency coefficients were transformed using a Two Dimensional (2D) DWT followed by arithmetic encoding, while the remaining high frequency coefficients were simply quantized and arithmetic encoded. Recently, it has been concluded in the literature that HEVC Main Still Image Profile [23] presents significant compression performance improvements in comparison to previous transform-based still image coding technologies [24,25]—such as JPEG (DCT-based) and JPEG 2000 (DWT-based) standards. Moreover, similar conclusions have been also reached for LF image coding, in [26,27], where HEVC presented significantly better performance than JPEG and JPEG 2000.

1.1.2. LFC based on view extraction

Alternatively, other schemes proposed to extract a set of views from the LF data for coding. In [10–15], MIs or Viewpoint Images (VIs) were extracted from the LF content in order to represent the LF data as a set of views and to use inter-view prediction for achieving compression. In [10,11], these views were then encoded as multiview content using Multiview Video Coding (MVC) [28]. Differently, in [12–15], the views were encoded as a Pseudo Video Sequence (PVS) using a 2D video coding standard, such as H.264/AVC [28], in [12], or High Efficiency Video Coding (HEVC) [23] in [13–15]. Although conceptually different (in terms of coding architecture), both multiview- and PVS-based coding approaches have the same basic purpose of proposing an efficient prediction configuration for better exploiting the correlations between the views. For this, different scanning patterns for ordering the views, as well as different prediction structures have been proposed. In [29], it was shown that the PVS-based coding solution outperformed a transform-based solution (similar to the LF coding solution proposed in [9]) with significant gains, notably, at lower bit rates. In addition, an alternative to the multiview representation based on these low resolution MIs/VIs was proposed in [16,17] using super-resolved rendered views. In this case, the scalable coding architecture proposed in [16] was used, which supported backward compatibility to legacy 2D and 3D multiview displays in the lower layer while the highest layer supports the entire LF content. In [17], the associated disparity information was also encoded and transmitted in the lower layers along with the set of views.

1.1.3. LFC based on non-local spatial prediction

Schemes based on the non-local spatial predictive approach rely on a non-local prediction techniques that exploit the existing redundancy between MIs in a (spatial) neighborhood to encode the entire raw LF image, being usually integrated (but not necessarily so) on a standard 2D image codec. The idea of exploiting non-local spatial redundancy has been firstly proposed for 2D image and video compression in order to further enhance the performance of H.264/AVC intra prediction [30]. Notably, the intra macroblock compensation technique was proposed in [30] to extend the usage of motion compensated prediction for intra-coded frames.

In the context of LF content coding, previous work of the authors [18,19] showed that further improvements are still possible for LF images with respect to the state-of-the-art for 2D image coding using the HEVC Main Still Picture profile [24,25,31] by using the concept of Self-Similarity (SS) compensated prediction. Similar to the intra macroblock compensation [30], the SS estimation process uses a block-based matching over the previously coded and reconstructed area of the current picture (referred to as SS reference [18]), to find the 'best' predictor for the current block. As a result, the chosen block becomes the candidate predictor and the relative position between the two blocks is signaled by an SS vector. In [19], a novel vector prediction scheme was also proposed to take advantage of the particular characteristics of the SS prediction data and thus increase coding efficiency. Subsequently, in [20], a scheme to extend the SS prediction concept by using HEVC inter B frame bi-prediction was proposed for LF image coding. However, in this case, to guarantee that the two prediction signals came from two different MIs, the search area was proposed to be separated into two non-overlapping parts [20] to perform the prediction estimation as in conventional HEVC bi-prediction. Although not targeting LF image coding, another prediction scheme similar to the SS compensated prediction, known as Intra Block Copy (IntraBC) [32], has been recently proposed in the literature in the context of Screen Content Coding (SCC) [32]. In this case [32], the prediction estimation is performed considering only integer pixel accuracy and the search window is expanded to the entire CB row or column, or to the entire previously coded area of the picture by using a hash-based search [32].

Furthermore, instead of using a block-based matching approach, an alternative prediction scheme based on locally linear embedding was proposed in [21], where a set of nearest neighbor patches were estimated from the same search area and linearly combined to predict the current block. More recently, in [22], a prediction scheme based on Gaussian Process Regression (GPR) was also proposed for LF image coding. In this case, two separate search areas are adopted for finding a set of nearest neighbor patches and the prediction is modeled as a non-linear (Gaussian) process for estimating the predictor block.

1.2. Motivations and contributions

Motivated by the authors' results in [18,19,21], this paper proposes an improved LF image coding solution based on HEVC and a novel Bi-predicted Self-Similarity (Bi-SS) estimation approach using the generic concept of superimposed prediction [33], which allows bi-prediction using samples from the same search area. Therefore, instead of dividing the search are into two non-overlapping parts to derive each predictor block from different MIs (as in [20]), these predictor blocks can be located in the same MI and in overlapped pixel positions. Moreover, instead of simply combining the two (independent) best uni-predicted candidate predictor blocks for bi-prediction (as in [20]), the locally optimal rate-constrained algorithm [34] is used for jointly estimating these two predictor blocks.

In addition to this, a theoretical analysis of the proposed Bi-SS prediction is also presented, which shows that other non-local spatial prediction schemes – such as the IntraBC [32], the preceding uni-predicted SS solution in [19], and the bi-prediction proposed in [20] – are suboptimal in terms of Rate-Distortion (RD) performance and, for this reason, can be considered as restricted cases of the jointly estimated Bi-SS solution proposed here. Furthermore, studies about the influence of the MI cross-correlation and the weighting factors used for bi-prediction on the RD efficiency of the Bi-SS prediction are also presented to experimentally validate the theoretical assumptions used for LF image coding.

Experimental results show that the proposed LFC solution using the jointly estimated Bi-SS prediction – from now on referred to as LFC Bi-SS solution – is able to outperform with significant coding gains various state-of-the-art LFC solutions based on different non-local special predictions.

1.3. Paper outline

The remainder of the paper is organized as follows: Section 2 describes the proposed LFC Bi-SS solution architecture; Section 3 proposes the jointly estimated Bi-SS prediction and presents the theoretical and experimental analyses of its prediction efficiency improvement for LF image coding; Section 4 presents the test conditions and experimental results; and, finally, Section 5 concludes the paper.

2. LFC Bi-SS solution architecture

The proposed LFC Bi-SS solution is not tuned for any particular optical acquisition setup since it does not require any explicit knowledge about it (e.g., microlens size, focal length, and distance of the microlenses to the image sensor). Notice that, although these parameters