RESEARCH   **Open Access**

CrossMark

Scalable kernel-based minimum mean square error estimate for light-field image compression

Zhixiang You[1,2*], Ping An[1,2*] and Deyang Liu[3]

Abstract

Light-field imaging can capture both spatial and angular information of a 3D scene and is considered as a prospective acquisition and display solution to supply a more natural and fatigue-free 3D visualization. However, one problem that occupies an important position to deal with the light-field data is the sheer size of data volume. In this context, efficient coding schemes for such particular type of image are needed. In this paper, we propose a scalable kernel-based minimum mean square error estimation (MMSE) method to further improve the coding efficiency of light-field image and accelerate the prediction process. The whole prediction procedure is decomposed into three layers. By using different prediction method in different layers, the coding efficiency of light-field image is further improved and the computation complexity is reduced both in encoder and decoder side. In addition, we design a layer management mechanism to determine which layers are to be employed to perform the prediction of the coding block by using the high correlation between the coding block and its adjacent known blocks. Experimental results demonstrate the advantage of the proposed compression method in terms of different quality metrics as well as the visual quality of views rendered from decompressed light-field content, compared to the HEVC intra-prediction method and several other prediction methods in this field.

Keywords: Light-field image, Image compression, Scalable kernel-based MMSE estimate, Layer management mechanism, HEVC SCC

1 Introduction

Light-field imaging, also referred to as plenoptic imaging, holoscopic imaging, and integral imaging, can capture both spatial and angular information of a 3D scene and enable new possibilities for digital imaging [1]. Light fields (LFs) captured by the light-field imaging represent the intensity and direction of the light rays emanated from a 3D scene. The full LFs can be represented by a seven-dimensional plenoptic function $L = L(x, y, z, \theta, \phi, \lambda, t)$ introduced by Adelson and Bergen [2], where (x, y, z) is the viewing position, (θ, ϕ) is the light ray directions, λ is the light ray wavelength, and t is the time. The seven-dimensional plenoptic function is further simplified into four dimensions by considering only information

taken in a region free of occlusions at a single time instance [3, 4]. The simplified 4D function $L = L(u, v, x, y)$ can represent a set of light rays by being parametrized as an intersection of rays with two planes, where uv describing the ray position in aperture (object) plane and xy describing the rays position in image plane. Under this point of view, the LFs can be explored from the digital perspective with the advances in computational photography [5].

There are several techniques that can be utilized to capture a light-field image, such as by using coded apertures, multi-camera arrays, and micro-lens arrays. In the micro-lens array-based technique, the most appropriate way to capture the LF image is by using plenoptic camera which is produced by Lytro. The commercially available plenoptic cameras can be divided into two categories: standard plenoptic camera and focused plenoptic camera. The focused plenoptic camera can provide a trade-off between spatial and angular information by putting the focal plane of microlenses away from the

* Correspondence: zhx.you@gmail.com; anping@t.shu.edu.cn
Ping An is an IEEE member
[1]Shanghai Institute for Advanced Communication and Data Science, Shanghai 200444, China
Full list of author information is available at the end of the article

Springer Open

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 2 of 14

image sensor plane. Since the wide range of possible applications and the rapid development of light-field technology, many research groups have considered the standardization of light-field application. The JPEG working group starts a new study, known as JPEG Pleno [6], aiming at richer image capturing, visualization, and manipulation. The MPEG group has also started the third phase of free-viewpoint television (FTV) since 2013, targeting super multi-view, free navigation, and full parallax imaging applications [7].

Since the acquired LF image records both spatial and angular information of a 3D scene, it can naturally provide the benefits of rendering views at different viewpoints and views at different focused planes, which expand its applications. However, a vast amount of data is needed for such a LF image during the acquisition step for its enhanced features. Even though many image coding methods [8–11] have been proposed, they cannot be directly used for LF image. Therefore, efficient compression schemes for such particular type of image are needed for effective transmission and storage.

According to the available techniques to capture and visualize LF images, the compression schemes of LF images can be mainly categorized into two different categories, where the general diagram of workflow for LF image acquisition and visualization is depicted in Fig. 1. The first kind of compression methods, called spatial correlation-based compression method, is to compress the acquired lenslet image directly (see Fig. 1a) based on the fact that the elementary images (EIs) of lenslet image exhibit repetitive patterns and a large amount of redundancy exists between the neighboring EIs, which can be seen in Fig. 2a. By exploiting the inherent non-local spatial redundancy of LF image, a coding method combining the locally linear embedding (LLE) is proposed in [12]. This work is further improved by combining the LLE based-method and self-similarity (SS)-based compensated prediction method in [13]. Paper [14] puts forward a disparity compensation-based light-field image coding algorithm by exploring the high spatial correlation existing in the LF images, which is further improved by using kernel-based minimum mean-square-error estimation prediction [15] and Gaussian process regression-based prediction method [16]. In [17], Conti et al. introduced the SS mode into

HEVC to improve the coding efficiency of light-field image, which is similar to the intra-block copy (IntraBC) incorporated into the HEVC range extension to code the screen contents. To further improve the coding performance, a bi-predicted SS estimation and SS compensation are proposed in [18], where the candidate predictor can be also devised as a linear combination of two blocks within the same search window. A displacement intra-prediction scheme for LF contents is proposed in [19], where more than one hypothesis is used to reduce prediction errors.

The other kind of compression method, called pseudo-sequence-based compression method, considers creation of a 4D LF representation of LF image prior to compression (see Fig. 1b). The pseudo-sequence-based coding methods try to decompose the LF image into multiple views, which can be seen in Fig. 2b. The derived multiple views are then organized into a sequence to make full use of the inter-correlations among various views. In [20], a sub-aperture images streaming scheme is proposed to compress the lenslet images, in which rotation scan mapping is adopted to further improve compression efficiency. A pseudo-sequence-based scheme for LF image compression is proposed in [21], in which the specific coding order of views, prediction structure, and rate allocation have been investigated for encoding the pseudo-sequence. A new LF multi-view video coding prediction structure by extending the inter-view prediction into a two-directional parallel structure is designed in [22] to analyze the relationship of the prediction structure with its coding performance. In [23], a lossless compression method of the rectified LF images is presented to exploit the high similarity existing among the sub-aperture images or view images. A novel pseudo-sequence-based 2D hierarchical reference structure for light-field image compression is proposed in [24], where a 2D hierarchical reference structure is used with the distance-based reference frame selection and spatial-coordinate-based motion vector scaling to better characterize the inter-correlations among the various views decomposed from the light-field image.

Although the pseudo-sequence-based compression method can compress the LF image effectively, the process to derive the 4D LF view images from the raw sensor data strongly depends on the exact acquisition
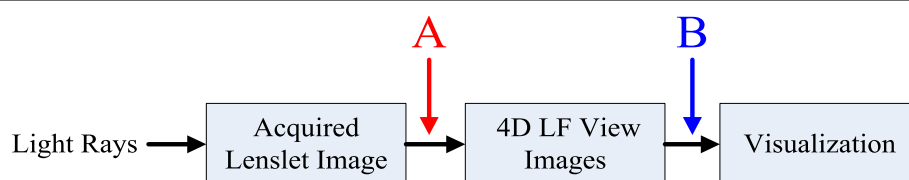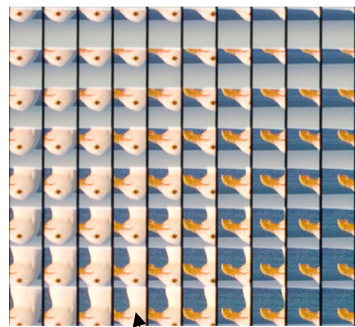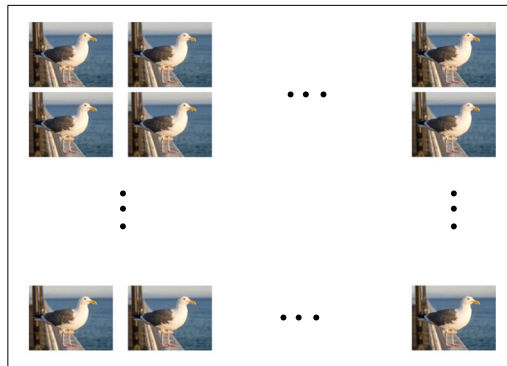

Fig. 1 General acquisition and display pipeline for LF images

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 3 of 14

Fig. 2 Acquired light-field image *seagull*: **a** lenslet image. **b** 4D LF view images

device. In contrast, the spatial correlation-based compression method does not need to extract view images from the LF image, and this kind of method has the potential to achieve a better coding efficiency if we can take full use of the high spatial correlation between the adjacent EIs. Therefore, in this paper, we follow the spatial correlation-based compression method and propose a scalable kernel-based minimum mean square error estimation (MMSE) estimation method to effectively compress the LF image by exploring such high spatial correlation. The contributions of this paper are as follows:

1) Hybrid kernel-based MMSE estimation and intra-block copy for LF image compression. The kernel-based MMSE estimation aims to predict the coding block by using an MMSE estimator where the required probabilities are obtained through kernel density estimation (KDE). Although the kernel-based MMSE estimation method can achieve a high coding efficiency, it does not always necessarily lead to a good prediction for the unknown block to be predicted, especially in non-homogenous texture areas. Fortunately, the unknown blocks located in such areas can be better predicted by using a direct match with the block to be predicted. Therefore, we combine the kernel-based MMSE estimation method with the IntraBC mode to further improve the whole coding efficiency.

2) Scalable kernel-based MMSE estimate to accelerate LF image compression. The kernel-based MMSE estimation method is time-consuming both in encoder side and decoder side, and it is much worse for the hybrid prediction method in the encoder side. Therefore, we propose a scalable kernel-based MMSE estimation method to alleviate such shortcomings. In the scalable method, the reconstruction

framework is decomposed into a set of reconstruction layers ranging from the basic layer that produces a rough yet fast estimation to more complex layers yielding high quality results.

3) Adaptive layer management mechanism. We use the prediction mode clue and the gradient information to decide which layer belongs for the current coding block. The adaptive layer management mechanism is designed to select which layer is belong to for a coding block based on how complex its surrounding area is.

Part of this pork has been published in [25]. In this paper, we give more details of theoretical analysis and propose a scalable kernel-based MMSE estimation method to accelerate LF image compression, and we also provide an adaptive layer management mechanism. Experimental results demonstrate the advantage of the proposed scalable compression method in terms of different quality metrics as well as the visual quality of views rendered from decompressed LF content.

The rest of this paper is organized as follows. An overview of the kernel-based MMSE estimation method is introduced in Section 2. The proposed scalable kernel-based MMSE estimate and its different reconstruction layers are described in Section 3. Section 4 gives the details of the layer selection mechanism. Experimental results are presented and analyzed in Section 5, and the concluding remarks are given in Section 6.

2 Kernel-based MMSE estimation method

The kernel-based MMSE estimation method aims to predict the current coding block given its neighbor known context under a kernel-based point of view by constructing a statistical model and calculating a kernel-based MMSE estimation. In order to construct

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 4 of 14

the statistical model, the pixel values in the coding block and its neighbor known context are arranged into a multidimensional formalism. Kernel density estimation (KDE) is used to estimate the probability density function (PDF) of the statistical model with a set of observed vectors. The coding block is then predicted from an MMSE estimator given the PDF.

Let the pixel values in the current coding block be stacked in a column vector $\mathbf{x}_0$, and the pixel values in its neighbor templates with template thickness being T are compacted in a column vector $\mathbf{y}_0$, shown in Fig. 3a. The current coding block and its neighbor templates with template thickness being T are called the prototype region in this paper. Therefore, the main goal of the prediction method can be expressed to derive the MMSE estimate $E[\mathbf{x}|\mathbf{y}_0]$ of vector $\mathbf{x}_0$ given its context $\mathbf{y}_0$. In order to do so, we arrange the vector $\mathbf{x}_0$ and $\mathbf{y}_0$ into a multidimensional vector $\mathbf{z}_0 = (\mathbf{x}_0, \mathbf{y}_0)$, and a random vector variable $\mathbf{z} = (\mathbf{x}, \mathbf{y})$ that has the same configuration as $\mathbf{z}_0$ is considered to obtain the signal statistical behavior. If the PDF of $\mathbf{z}$ is acquired, the MMSE estimator $E[\mathbf{x}|\mathbf{y}_0]$ can be derived from it. In the proposed scheme, we propose to utilize KDE to estimate the PDF of random variable $\mathbf{z}$ given a set of observed vectors $\{\mathbf{z}_k| k = 1, …, K\}$. Here, the observed vectors are composed of K-NN patches, the top K closest templates that have the same configuration as the prototype region in terms of Euclidean distance, derived within the specified horizontal and vertical search windows, shown in Fig. 3b. Since the coding block in prototype region is unknown in the K-NN patches searching procedure, its neighbor blocks are used for searching. We set the template thickness to T_1

which equals to the size of the current coding block to increase the searching accuracy, shown in Fig. 3b.

Given the observed vectors, the estimator of the PDF of $\mathbf{z}$ using KDE with a Gaussian kernel $K(\mathbf{u}) = \exp(-\mathbf{u}^{\mathrm{T}}\mathbf{u}/2)/\sqrt{2\pi}$ can be defined by [26],

$$p(\mathbf{z}) = \frac{1}{KH} \sum_{k=1}^{K} K\left(\frac{\mathbf{z}-\mathbf{z}_k}{H}\right) = \frac{1}{K} \sum_{k=1}^{K} K_Z^{(k)}(\mathbf{z}) \tag{1}$$

where matrix H is called bandwidth, controlling the smoothness of the resulting PDF. $K_Z^{(k)}(\mathbf{z})$ can be considered as a multivariate Gaussian with mean $\mathbf{z}_k$ and covariance matrix $\mathbf{H} = HH^{\mathrm{T}}$. The covariance matrix $\mathbf{H}$ is also called as bandwidth for simplicity, which can be decomposed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{XX} & \mathbf{H}_{XY} \\ \mathbf{H}_{YX} & \mathbf{H}_{YY} \end{bmatrix} \tag{2}$$

With the knowledge of $p(\mathbf{z})$, we can calculate the MMSE estimator $E[\mathbf{x}|\mathbf{y}_0]$ of vector $\mathbf{x}_0$. We find from Eq. (1) that $p(\mathbf{z})$ has the form of Gaussian mixture model (GMM) with a priori probabilities $1/K$ and covariance matrix $\mathbf{H}$. Therefore, it is reasonable to utilize the expressions of MMSE estimator under GMM model [26, 27]. The MMSE estimator of the coding block can be expressed as,

$$\hat{\mathbf{x}}_0 = E[\mathbf{x}|\mathbf{y}_0] = \sum_{k=1}^{K} \omega_k(\mathbf{y}_0)\boldsymbol{\mu}_{X|Y}^{(k)}(\mathbf{y}_0) \tag{3}$$
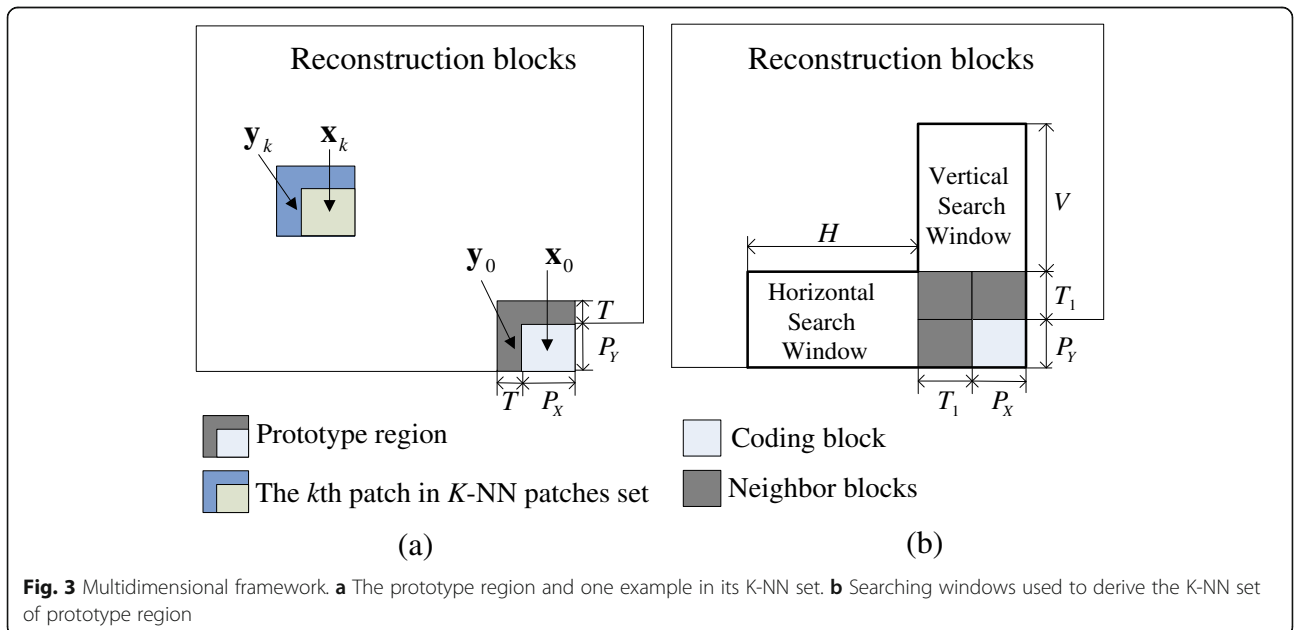

Fig. 3 Multidimensional framework. **a** The prototype region and one example in its K-NN set. **b** Searching windows used to derive the K-NN set of prototype region

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 5 of 14

$$\omega_k(\mathbf{y}_0) = \frac{K_Y^{(k)}(\mathbf{y}_0)}{\sum_{k=1}^{K} K_Y^{(k)}(\mathbf{y}_0)} \tag{4}$$

$$\boldsymbol{\mu}_{X|Y}^{(k)}(\mathbf{y}_0) = E[\mathbf{x}|\mathbf{y}_0, \mathbf{y}_k] = \mathbf{x}_k + \mathbf{H}_{XY}\mathbf{H}_{YY}^{-1}(\mathbf{y}_0 - \mathbf{y}_k) \tag{5}$$

where $K_Y^{(k)}(\mathbf{y})$ is the marginal kernel for $\mathbf{y}$, with mean $\mathbf{y}_k$ and covariance matrix $\mathbf{H}_{YY}$. According to Eqs. (3)–(5), we can obtain the prediction of the coding block and the estimation method is referred as the kernel-based MMSE (K-MMSE) estimation method. For simplicity, the K-MMSE estimation can be rewritten as

$$\hat{\mathbf{x}}_0 = \tilde{\mathbf{x}}_0 + \mathbf{H}_{XY}\mathbf{H}_{YY}^{-1}(\mathbf{y}_0 - \tilde{\mathbf{y}}_0) \tag{6}$$

where $\tilde{\mathbf{x}}_0$ and $\tilde{\mathbf{y}}_0$ express the linear predictions of $\mathbf{x}_0$ and $\mathbf{y}_0$ from the sets of vectors $\mathbf{z}_k(k = 1, ..., K)$ by

$$\tilde{\mathbf{x}}_0 = \sum_{k=1}^{K} \omega_k(\mathbf{y}_0)\mathbf{x}_k, \quad \tilde{\mathbf{y}}_0 = \sum_{k=1}^{K} \omega_k(\mathbf{y}_0)\mathbf{y}_k \tag{7}$$

There are two issues that should be tackled in K-MMSE estimation method. One is to derive the weight vector $\omega_k(\mathbf{y}_0)$. The other is to estimate the kernel bandwidth matrix $\mathbf{H}$.

In order to derive the weight vector, we adopt a more direct way. That is to minimize the residual energy $\varepsilon(\boldsymbol{\omega})$ by solving a squared error function, where $\varepsilon(\boldsymbol{\omega})$ is given by

$$\varepsilon(\boldsymbol{\omega}) = \left\| \mathbf{y}_0 - \sum_{k=1}^{K} \omega_k(\mathbf{y}_0)\mathbf{y}_k \right\| \tag{8}$$

In order to estimate the kernel bandwidth matrix, we propose a new BE method which is based on the analysis of physical interpretation of the K-MMSE estimation method. From Eq. (6), we find that the K-MMSE estimator consists of two parts. The first part is a linear prediction of vector $\mathbf{x}_0$, and the second part is a correction vector representing the unpredictable part of $\mathbf{y}_0$ is transformed into subspace $\mathbf{x}$ [26]. The physical interpretation of the second part can be understood as the vector $\mathbf{x}_0$ being close to vector $\mathbf{y}_0$ is likely to have a similar unpredictable part as vector $\mathbf{y}_0$. The matrix $\mathbf{H}_{XY}\mathbf{H}_{YY}^{-1}$ can be regarded as a transfer matrix used to transfer the unpredictable part of $\mathbf{y}_0$ to subspace $\mathbf{x}$. It is reasonable to infer that the bandwidth matrix $\mathbf{H}$ in K-MMSE estimation is used to measure the similarity of the subspace $\mathbf{x}$ and subspace $\mathbf{y}$. Therefore, we propose to utilize Eq. (9) to estimate the bandwidth matrix $\mathbf{H}$ approximately.

$$\mathbf{H} = \eta^2(1 + \mathbf{x}^T\mathbf{y}) \tag{9}$$

where η is a hyper parameter and is specified to 1.0 in the proposed system.

3 Scalable kernel-based MMSE (SK-MMSE) estimation

The kernel-based MMSE estimation method is a powerful prediction method and can achieve a better prediction accuracy for LF contents in the homogenous texture areas. However, some shortcomings still exist. Firstly, since such method predicts the coding blocks given their neighbor known contexts, it does not always necessarily lead to a good prediction for the unknown block in some non-homogenous texture areas. Secondly, the kernel-based MMSE estimation method is time-consuming, especially in the decoder side. Thirdly, for some visually flat regions in homogenous texture areas, applying kernel-based MMSE estimation method would be an overkill since similar reconstruction quality could be achieved by simpler (and, therefore, faster) estimators when dealing with relatively simple structures. To this end, this paper proposes a scalable kernel-based MMSE estimation method, also called SK-MMSE, that aims at further improving the coding efficiency and accelerating the prediction process by decomposing the prediction procedure into different layers. The proposed SK-MMSE algorithm is comprised by three prediction layers. The layer division is based on the contents of the previously encoded blocks adjacent to the coding block. The higher the layer within the scalable hierarchy, the higher the computational complexity. The scalable layers are introduced in the next subsections, and the layer management mechanism will be illustrated in the next section.

3.1 Hybrid prediction layer (HPL)

We have mentioned above that the K-MMSE estimation method does not always necessarily lead to a good prediction for the unknown blocks in the non-homogenous texture areas. In order to improve the prediction accuracy for the blocks in such texture areas, we propose to use the hybrid kernel-based MMSE estimation and IntraBC method (hybrid prediction method) to predict the coding blocks and the coding blocks in such texture areas construct the hybrid prediction layer. The hybrid prediction method is based on the HEVC screen content coding (HEVC-SCC) framework. In the hybrid prediction method, the K-MMSE estimation method, IntraBC prediction, and the intra-directional prediction are all used as the competing prediction modes. The proposed hybrid prediction explores the idea of using the IntraBC scheme or intra-directional prediction to find the best prediction of the coding block $\hat{\mathbf{x}}_0^{HPL}$ when K-MMSE estimation method fails based on the rate-distortion optimization (RDO) procedure. In other word, the hybrid prediction

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 6 of 14

method uses the "try all then select best" intra-mode decision method to find the best prediction mode and optimal depth for each coding block.

It is worth to notice that the K-MMSE estimation method is introduced into HEVC SCC by replacing one of the existing 35 intra-directional prediction modes in order to avoid the modification of bit stream structure, which means the prediction samples that generated by the K-MMSE estimation method will replace the outputs produced by the substituted intra-directional prediction mode.

3.2 K-MMSE prediction layer (KPL)
Other than the non-homogenous texture areas, in many cases, we are dealing with the homogenous texture areas. In these areas, the coding block and its adjacent reconstructed blocks have the similar texture structure, which means that the coding block and its adjacent reconstructed blocks have a high correlation. For such areas, we propose to use the K-MMSE estimation method to predict the coding blocks, as described in Section 2, and the coding blocks in such texture areas construct the K-MMSE prediction layer. Since the high correlation is existed between the subspace $\mathbf{x}$ and subspace $\mathbf{y}$, the current coding block is likely to have a similar unpredictable part as its adjacent reconstructed blocks. Therefore, we can achieve a higher prediction accuracy by using the K-MMSE estimation method for the coding blocks in KPL. The estimator of the coding blocks in the homogenous texture areas can be expressed as

$$\hat{\mathbf{x}}_0^{KPL} = \tilde{\mathbf{x}}_0 + \mathbf{H}_{XY}\mathbf{H}_{YY}^{-1}(\mathbf{y}_0 - \tilde{\mathbf{y}}_0) \tag{10}$$

As mentioned earlier, the K-MMSE estimation method is also implemented in the HEVC-SCC framework. For coding blocks in the HPL, the "try all then select best" intra-mode decision method is used to find the best prediction mode and optimal depth. However, for the coding blocks in KPL, we will skip the IntraBC mode and only to derive the best prediction mode and optimal depth among the K-MMSE mode (K-MMSE estimation method) and the other 34 intra-directional prediction modes to reduce the computation complexity. Moreover, since the LF image is composed of numerous EIs and the texture-homogeneous areas hardly prevail in the LF image, the coding unit size 64×64 is seldom chosen as the optimal block size. Consequently, for K-MMSE estimation method, we only choose four coding block sizes ranging from 32×32 to 4×4.

3.3 Linear prediction layer (LPL)
There is a special case in the KPL, that is the visually flat regions, such as the skies and walls. In such flat regions, the luminance information of the coding block and its

adjacent reconstructed blocks is simple. In the K-MMSE estimation method, the coding block is predicted by using two terms, shown in Eq. (6). The first term is a linear prediction of vector $\mathbf{x}_0$, and the second term is a correction vector representing the unpredictable part of $\mathbf{y}_0$ is transformed into subspace $\mathbf{x}$ [26]. Since the luminance information in the flat regions is simple, the unpredictable part of $\mathbf{y}_0$ can be neglected with negligible effect to the prediction accuracy. This means that we can predict the coding blocks in such flat regions by directly using a linear prediction and do not need to compute the correction vector. Likewise, the coding blocks in the visually flat regions construct the LPL and the estimator of the coding blocks in such regions can be derived by

$$\hat{\mathbf{x}}_0^{LPL} = \tilde{\mathbf{x}}_0 = \sum_{k=1}^{K} \omega_k(\mathbf{y}_0)\mathbf{x}_k \tag{11}$$

The weight vector $\omega_k(\mathbf{y}_0)$ is achieved by using Eq. (8). The used linear prediction method is implemented in the HEVC-SCC framework by using the same way as the K-MMSE estimation method.

In this section, we have introduced three prediction layers according to the contents of the coding blocks and their adjacent reconstructed blocks, which comprise the proposed SK-MMSE estimation method. The main idea is to further improve the coding efficiency and accelerate the prediction process. The three prediction layers are summarized as follows.

1) *HPL* consists of the non-homogenous texture areas and hybrid prediction method is used to predict the coding blocks in such layer.
2) *KPL* consists of the homogenous texture areas where the texture information is abundant. For KPL, the K-MMSE estimation method is utilized to predict the coding blocks.
3) *LPL* consists of the visually flat regions in homogenous texture areas, and a linear prediction method is used to predict the coding blocks which is a simplified form of the K-MMSE estimation method by discarding the correction vector.

4 Layer switching and management mechanism
The goal of the proposed SK-MMSE is to further improve the coding efficiency and accelerate the prediction process. In order to do so, we propose a content adaptive layer selection scheme. Since the current coding block is unknown, we propose to utilize the contents of the neighbor known blocks adjacent to the coding block to decide which layer is belonged to for the current coding block. Since the contents of the coding block and its neighbor known blocks are

Table 1 The accuracy of $PM_C = PM_L = PM_U = PM_{UL}$ if $PM_L = PM_U = PM_{UL}$ in each depth level

Images	Depth 0 (%)	Depth 1 (%)	Depth 2 (%)	Depth 3 (%)	Average (%)
Bike	69.3	91.2	84.2	71.8	79.1
Fountain	67.4	79.6	82.8	81.2	77.8
Laura	72.6	85.9	85.6	84.0	82.0
Seagull	74.8	84.7	82.5	85.0	81.8
I04_Stone_Pillars_Outside	81.4	88.3	83.5	71.2	81.1
I06_Ankylosaurus_&_Diplodocus_1	95.4	95.4	95.4	88.9	93.8
I08_Magnets_1	91.9	97.5	92.7	84.4	91.6
I11_Color_Chart_1	91.0	98.0	94.3	88.4	92.9

closely linked, it is feasible to use the contents of the neighbor known blocks to determine which layers are to be employed to perform the prediction of the coding block. In order to achieve the goal, two assumptions are taken into consideration:
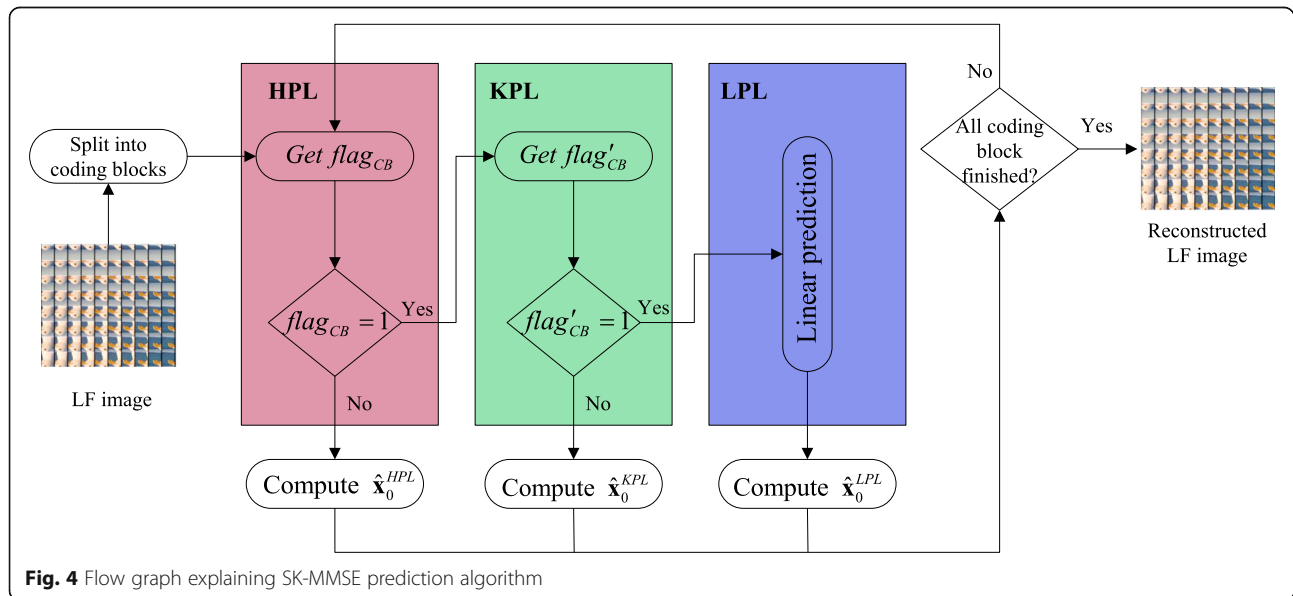
1) Given that the HPL is used to improve the whole coding efficiency, to decide accurately which coding blocks belong to this layer is of great importance. The decision criterion should take into account the contents of the coding blocks.

2) In order to accelerate the prediction process, the LPL is expected to perform well for the visually flat regions in homogenous texture areas. The visual flatness should be considered as the decision criterion, and the decision criterion should be simple and fast.

In order to consider the two assumptions, we propose to use the prediction mode correlation to decide which blocks belong to HPL and use the gradient information to measure the visual flatness. The following will introduce the two layer switching mechanism.

For homogenous texture areas, the coding blocks and their neighboring blocks are closed linked. In most cases, they have the similar texture information and structural characteristics. As a results, the prediction modes of the coding blocks should be similar to their neighboring known blocks in the homogenous texture areas. Consequently, we can apply the prediction mode information to decide whether the coding block belongs to the homogenous texture areas and further determine whether the coding block belongs to HPL or KPL. Since the prediction mode information of the current coding block is unknown, the prediction mode information of its neighboring known blocks are utilized in the decision criterion.

Suppose the optimal prediction modes of the current coding block and its left neighboring block, up neighboring block, and up-left neighboring block are denoted by
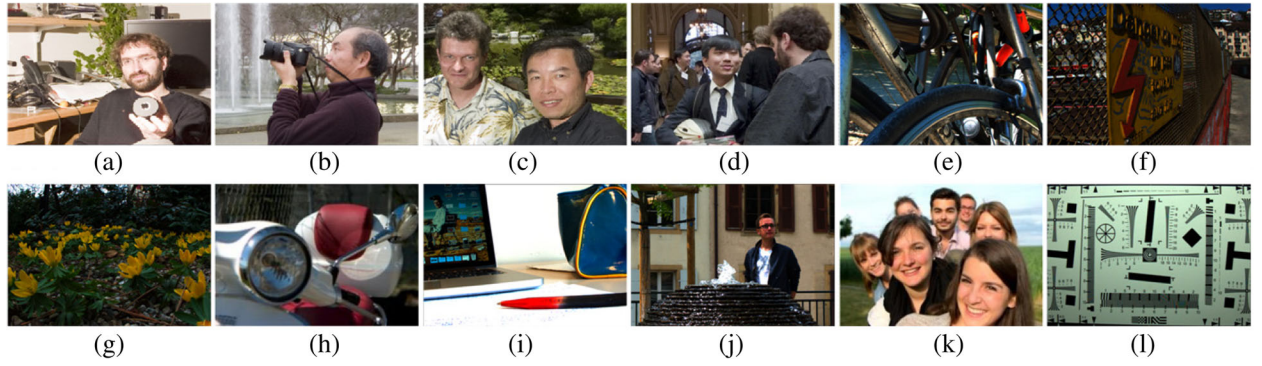

Fig. 4 Flow graph explaining SK-MMSE prediction algorithm

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 8 of 14

Fig. 5 The central rendered views from each LF test image: **a** Fredo, **b** Jeff, **c** Sergio, **d** Zhengyun1, **e** I01_Bikes, **f** I02_Danger_de_Mort, **g** I03_Flowers, **h** I05_Vespa, **i** I07_Desktop, **j** I09_Fountain_&_Vincent_2, **k** I10_Friends_1, **l** I12_ISO_Chart_12

PM_C, PM_L, PM_U, PM_{UL}. We define a flag flag$_{CB}$ used to determine whether the coding block belongs to HPL or KPL, which is defined as

$$\text{flag}_{CB} = \begin{cases} 1 & \text{if } (PM_L = PM_U = PM_{UL}) \\ 0 & \text{Otherwise} \end{cases} \quad (12)$$

From Eq. (12), we see that if $PM_L = PM_U = PM_{UL}$, the flag$_{CB}$ is set to 1 and the current coding block is determined to belong to KPL. Otherwise, flag$_{CB}$ is set to 0 and the current coding block is determined to belong to HPL. The main reason is that if $PM_L = PM_U = PM_{UL}$, the current coding block is likely to have the same prediction mode as its neighboring blocks (left neighboring block, up neighboring block, and up-left neighboring block), which indicates that the current coding block and its neighboring blocks have a similar texture information and structural characteristics and they belong to the homogenous texture areas. Therefore, the current coding block is grouped into KPL. If the flag$_{CB}$ is equal to 0, which means that the current coding block is quite different from its neighboring blocks, the current coding block is grouped into HPL.

In order to verify the decision accuracy, Table 1 shows the accuracy of $PM_C = PM_L = PM_U = PM_{UL}$ if $PM_L = PM_U = PM_{UL}$ in each depth level. The accuracy means that the probability of $PM_C = PM_L = PM_U = PM_{UL}$ across all the test QPs when $PM_L = PM_U = PM_{UL}$. From Table 1, we see that the average accuracy of $PM_C = PM_L = PM_U = PM_{UL}$ if $PM_L = PM_U = PM_{UL}$ in each depth level across all the depth levels is from 77.8 to 93.8%, 85.0% on average. This means that if $PM_L = PM_U = PM_{UL}$, the prediction mode of the current coding block can be considered to be the same as its neighboring blocks. Therefore, we can use such prediction mode information to decide whether the current coding block is located at

the homogenous texture areas. From Table 1, we also find that the average accuracy is lower for rectangular lens LF images (e.g., bike, fountain, Laura, and seagull) for depth 0 than other depth level. The reason is that the size of EI image in rectangular lens LF image is approximate to the coding block size for depth 0. Since the EIs exhibit repetitive patterns, the homogenous texture areas hardly prevail in such depth level. Fortunately, the average accuracy across all the depth level is approximate to 80% and it is feasible to use the prediction mode information of the neighboring known blocks to decide whether the current coding block is located at the homogenous texture areas.

We have mentioned above that there is a special case in the KPL, that is the visually flat regions. If we can find

Table 2 Y-rate-distortion gains of three coding methods over HEVC

Images	HEVC-SCC		K-MMSE		SK-MMSE	
	BD-PSNR (dB)	BD rate (%)	BD-PSNR (dB)	BD rate (%)	BD-PSNR (dB)	BD rate (%)
Fredo	2.73	− 40.40	3.11	− 43.07	3.52	− 48.59
Jeff	1.58	− 24.22	2.31	− 32.92	2.47	− 35.76
Sergio	2.36	− 32.99	2.94	− 38.61	3.18	− 41.99
Zhengyun1	1.70	− 29.56	2.24	− 36.24	2.41	− 39.63
I01_LF	0.47	− 10.34	0.64	− 13.94	0.77	− 16.60
I02_LF	0.36	− 8.56	0.43	− 10.07	0.59	− 13.56
I03_LF	0.13	− 3.02	0.17	− 4.04	0.23	− 5.42
I05_LF	0.55	− 16.41	0.89	− 26.03	1.00	− 29.72
I07_LF	0.65	− 17.07	0.64	− 17.07	0.83	− 21.72
I09_LF	1.26	− 24.30	2.03	− 36.63	2.13	− 38.40
I10_LF	0.17	− 5.92	0.28	− 9.90	0.31	− 10.73
I12_LF	0.97	− 26.52	1.55	− 40.00	1.72	− 44.14
Average	1.08	− 19.94	1.44	− 25.71	1.61	− 28.85

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52
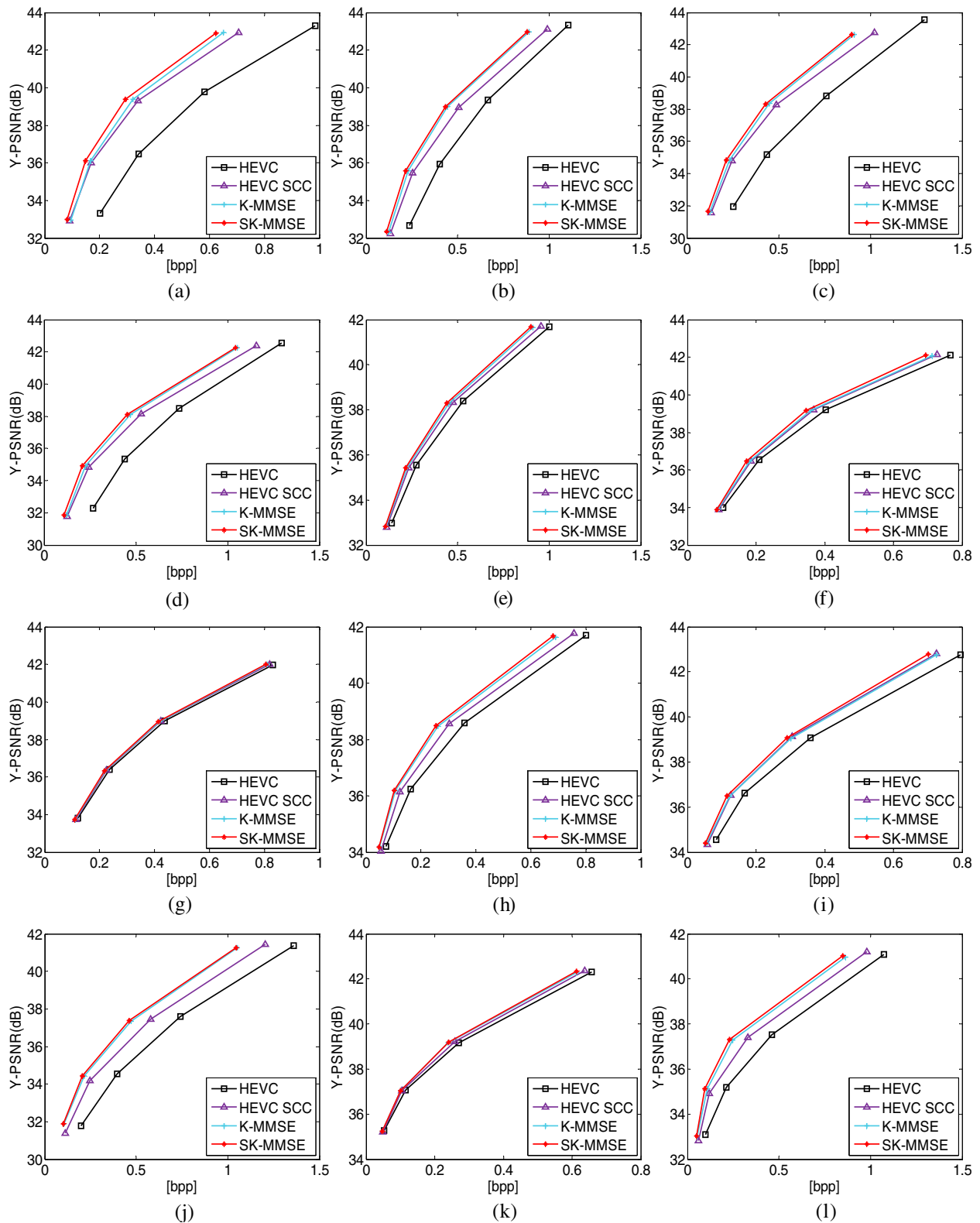
Page 9 of 14

Fig. 6 Rate-distortion results for the LF test image set with Y-PSNR as the objective quality metric. **a** Fredo, **b** Jeff, **c** Sergio, **d** Zhengyun1, **e** I01_Bikes, **f** I02_Danger_de_Mort, **g** I03_Flowers, **h** I05_Vespa, **i** I07_Desktop, **j** I09_Fountain_&_Vincent_2, **k** I10_Friends_1, **l** I12_ISO_Chart_12

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 10 of 14

Table 3 YUV-rate-distortion gains of three coding methods over HEVC

Images	HEVC-SCC		K-MMSE		SK-MMSE	
	BD-PSNR (dB)	BD rate (%)	BD-PSNR (dB)	BD rate (%)	BD-PSNR (dB)	BD rate (%)
Fredo	2.54	− 40.52	2.90	− 43.22	3.29	− 48.78
Jeff	1.42	− 23.91	2.16	− 33.35	2.31	− 36.17
Sergio	2.13	− 32.93	2.73	− 39.29	2.95	− 42.73
Zhengyun1	1.55	− 29.16	2.14	− 36.86	2.29	− 40.18
I01_LF	0.36	− 9.24	0.52	− 13.13	0.63	− 15.64
I02_LF	0.27	− 7.51	0.33	− 9.19	0.46	− 12.38
I03_LF	0.10	− 2.54	0.12	− 3.25	0.17	− 4.52
I05_LF	0.43	− 14.93	0.74	− 25.11	0.82	− 27.67
I07_LF	0.50	− 15.09	0.51	− 15.69	0.67	− 20.15
I09_LF	1.01	− 23.10	1.69	− 36.17	1.78	− 37.97
I10_LF	0.12	− 4.50	0.23	− 9.07	0.25	− 9.56
I12_LF	0.76	− 24.62	1.28	− 39.18	1.43	− 43.42
Average	0.93	− 19.00	1.28	− 25.29	1.42	− 28.26

the coding blocks that belong to these regions and use a simpler and faster method to predict these coding blocks with negligible effect to the prediction accuracy, the computational complexity can be reduced, especially in decoder side. To this end, we utilize the gradient information of the top nearest patch of the prototype region in K-NN patch set to judge whether the coding block belong to the LPL. Let $\mathbf{z}_1 = (\mathbf{x}_1, \mathbf{y}_1)$ be the vector form of the top nearest patch of the prototype region in K-NN patch set. Suppose G_{z1}, G_{x1}, and G_{y1} represent the gradient of vector $\mathbf{z}_1$, $\mathbf{x}_1$, and $\mathbf{y}_1$, respectively, we also define a flag flag'_{CB}. If $\|G_{x1} - G_{y1}\| < G_{z1}$, the coding block and its neighboring blocks are considered to be located at the visually flat regions and the flag'_{CB} is set to 1. Otherwise, flag'_{CB} is set to 0, and the coding block is considered to belong to the KPL. The flag flag'_{CB} is defined by

$$\text{flag}'_{CB} = \begin{cases} 1, & \text{if } \left\|G_{x1} - G_{y1}\right\| < G_{z1} \\ 0, & \text{Otherwise} \end{cases} \quad (13)$$

The proposed algorithm is summarized by the flow graph in Fig. 4. It is shown that the SK-MMSE algorithm is a scalable LF image coding method, where the whole coding blocks are divided into three layers. The HPL is used to improve the prediction accuracy for coding blocks in the non-homogenous texture areas while the KPL is applied to ensure the coding efficiency for the coding blocks in the homogenous texture areas where the texture information is abundant. Regarding to the LPL, a simplified prediction method is adopted to further reduce the whole computational complexity with negligible effect to the prediction accuracy, especially for the decoder side.

Note that, the proposed framework can also be used to predict chrominance blocks.

5 Experimental results and discussion

In order to validate the efficiency of the proposed method, 12 LF test images including eight circular lens LF test images provided by the ICME 2016 grand challenge in light-field image compression [28] and four rectangular lens LF test images provided by Dr. T. Georgiev [29] are used in the test set. The used LF test images are all captured by the focused plenoptic camera. The resolution of the circular lens LF test images is 7728×5368. The original resolution of the rectangular lens LF test images is 7240×5432, and we cut the four test images into size of 3840×2160 for simplicity. The size of each EI in rectangular lens LF image is 75×75. All the LF test images are transformed into YUV 4:2:0 format. The central rendered views from each LF test image are shown in Fig. 5.

The HEVC SCC reference software SCM-3.0 [30] is modified for the proposed hybrid codec architecture. The coding configurations were set as the "All Intra," which is defined in [31]. Four tested quantization parameters 22, 27, 32, and 37 are used. The proposed hybrid prediction method (referred to as SK-MMSE) is compared with three prediction schemes: the original HEVC (referred to as HEVC), the screen content coding extension Ver. 3.0 to HEVC (referred to as HEVC-SCC), and the kernel-based minimum mean-square-error estimation method [15] (referred to as K-MMSE). The K-MMSE method is realized in the HEVC SCC reference software SCM-3.0. The Y-PSNR and YUV-PSNR between the original LF image and decoded LF image shown in [28] are used as the objective quality metric.

The template thickness T in Fig. 3a is set to 4. The dimensions of the searching windows used in the SK-MMSE and K-MMSE method is given by $V = 128$, $H = 128$, shown in Fig. 3b. We have mentioned that the K-MMSE method is integrated into the HEVC SCC standard by replacing one of the 35 intra-directional prediction modes in SK-MMSE method. In our experiment, the intra-prediction mode "4" is replaced and K is set to 6.

Table 2 gives the rate-distortion gains of the three prediction methods over HEVC intra-standard with Y-PSNR as the objective quality metric. From Table 2, we see that the proposed SK-MMSE is clearly superior to the other methods. An average gain of up to 1.61 dB has been achieved by SK-MMSE to HEVC intra-standard. Compared to HEVC-SCC, around 30.9% BD-rate can be saved in average by using SK-MMSE. This is because integrating the K-MMSE method to the HEVC-SCC standard can effectively improve the prediction accuracy. When compared to
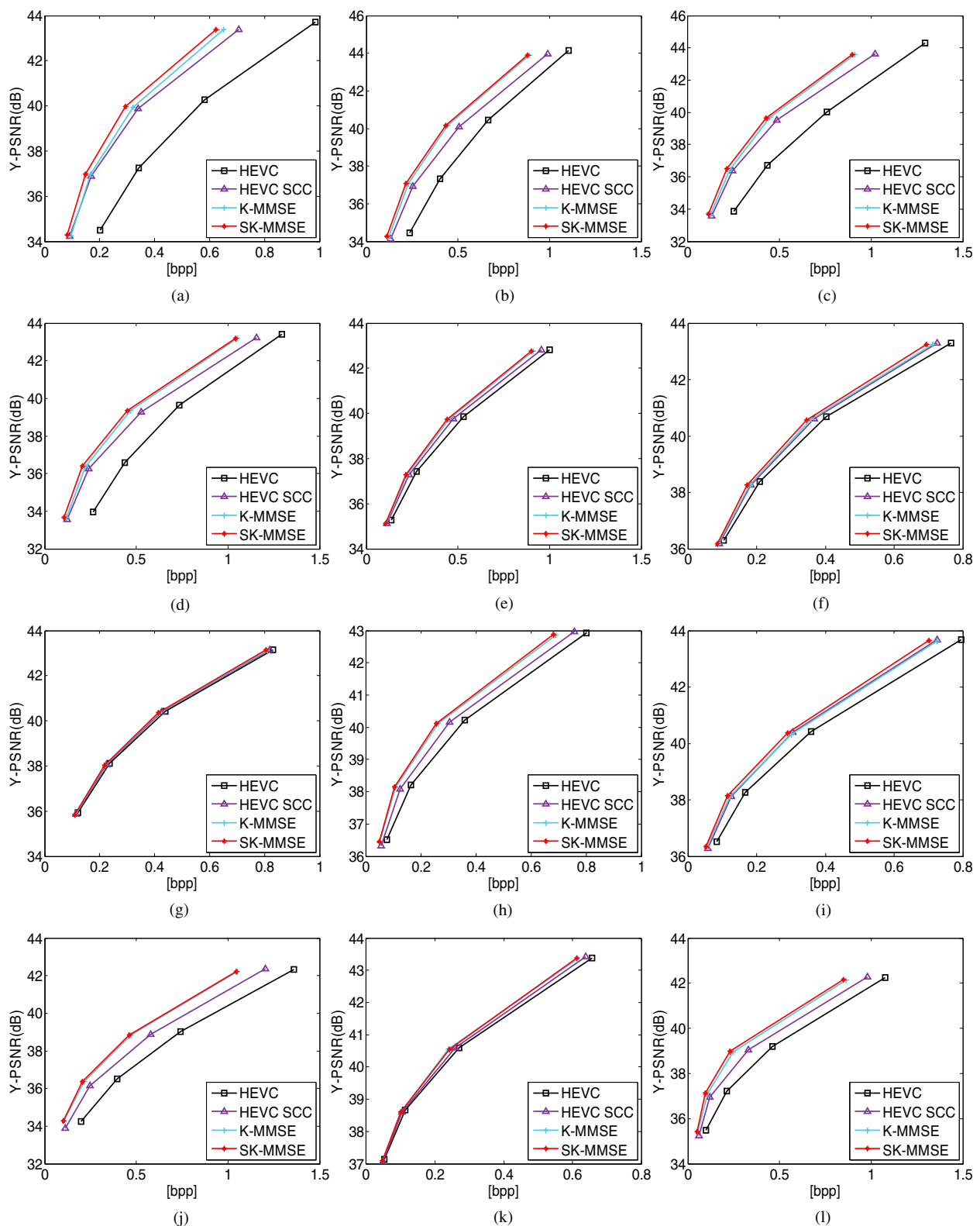
You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 11 of 14

Fig. 7 Rate-distortion results for the LF test image set with YUV-PSNR as the objective quality metric. **a** Fredo, **b** Jeff, **c** Sergio, **d** Zhengyun1, **e** I01_Bikes, **f** I02_Danger_de_Mort, **g** I03_Flowers, **h** I05_Vespa, **i** I07_Desktop, **j** I09_Fountain_&_Vincent_2, **k** I10_Friends_1, **l** I12_ISO_Chart_12

You *et al. EURASIP Journal on Image and Video Processing* (2018) 2018:52

Page 12 of 14

K-MMSE, the proposed SK-MMSE can achieve about 0.17 dB average gains. The main reason is that the K-MMSE do not work well for the blocks in non-homogenous texture area. By combining the IBC mode, the proposed SK-MMSE can achieve a better prediction of the coding block in such non-homogenous texture area. From Table 2, we can also observe that the K-MMSE method allows an average of 22.4% rate saving compared to the HEVC-SCC, which means that K-MMSE mode can obtain a better prediction and is selected as the best prediction mode in most cases. Figure 6 shows the rate-distortion curve of the test image set using different coding schemes with Y-PSNR as the objective quality metric, which further confirms that the proposed SK-MMSE performs better than other prediction methods.

The rate-distortion gains of the three prediction methods over HEVC intra-standard with YUV-PSNR as the objective quality metric are given in Table 3. From Table 3, we can achieve a consistent conclusion as Table 2. Compared to HEVC intra-standard, an average gain of up to 1.42 dB has been achieved by the proposed SK-MMSE method. Likewise, around 10.5 and 32.8% BD rate can be saved in average by using SK-MMSE when compared to K-MMSE and HEVC-SCC method, respectively. This also validates that the proposed SK-MMSE architecture can effectively compress the LF data. Figure 7 gives the rate-distortion curve of the test image set using different coding schemes with YUV-PSNR as the objective quality metric, which further proves the validity of the proposed SK-MMSE.

Table 4 shows the execution time ratios of three coding methods to HEVC intra-standard both in encoder side and decoder side. From Table 4, we observe that the K-MMSE method requires the most execution time both

in encoder side and decoder side. The main reason is that the calculation of kernel bandwidth matrix **H** is time-consuming for all the coding blocks. Since the decoder side has to do the same prediction procedure, the K-MMSE method needs 49.1 times execution time to the HEVC intra-standard. In order to reduce the computation complexity, we propose the SK-MMSE method. Table 4 also shows the effectiveness of the proposed SK-MMSE in computation complexity. From Table 4, we see that proposed SK-MMSE achieves 16.4 and 70.5% average coding time saving when compared to K-MMSE method in encoder side and decoder side, respectively. The main reason mainly lies in two aspects. One is that the coding blocks is divided into three layers, and by using a simpler and faster prediction method to LPL with negligible effect to the prediction accuracy, the computation complexity is reduced in the encoder side. The other is by dividing the coding blocks into three layers, the IntraBC mode and linear prediction mode (linear prediction method) are selected as the optimal prediction mode by many coding blocks. These two modes cost much less than the K-MMSE mode, especially in the decoder side. Although the computation complexity of the SK-MMSE is less than the K-MMSE, it still needs around two times execution time to HEVC-SCC in the encoder side.

Since LF image captures both spatial and angular information of a scene, view images can be rendered from the LF image data. In order to further validate the validity of the proposed coding scheme, we give a visual quality investigation of rendered view image from the decoded LF image in Fig. 8. As shown in Fig. 8, the proposed SK-MMSE can obtain a better visual quality, especially in some texture regions. The main reason mainly lies in two aspects. One is that

Table 4 Encoding and decoding time ratio to HEVC

Images	HEVC-SCC		K-MMSE		SK-MMSE	
	Encoder side	Decoder side	Encoder side	Decoder side	Encoder side	Decoder side
Fredo	4.1	0.84	18.9	76.6	14.8	24.0
Jeff	4.8	0.91	18.7	75.2	16.5	28.9
Sergio	4.8	0.74	17.8	67.3	16.0	25.2
Zhengyun1	4.7	0.64	18.2	52.4	16.7	19.4
I01_LF	8.1	0.92	17.4	46.9	14.5	10.2
I02_LF	7.7	1.00	16.1	44.8	13.8	9.3
I03_LF	7.0	0.62	13.3	16.4	12.3	3.7
I05_LF	4.8	0.81	19.3	51.4	10.4	10.5
I07_LF	5.2	0.60	15.5	23.2	10.3	4.6
I09_LF	20.3	0.68	20.7	45.2	27.6	18.1
I10_LF	4.2	0.84	12.0	25.4	7.0	4.1
I12_LF	11.5	0.69	23.9	64.0	17.4	15.7
Average	7.3	0.77	17.7	49.1	14.8	14.5

You et al. EURASIP Journal on Image and Video Processing (2018) 2018:52
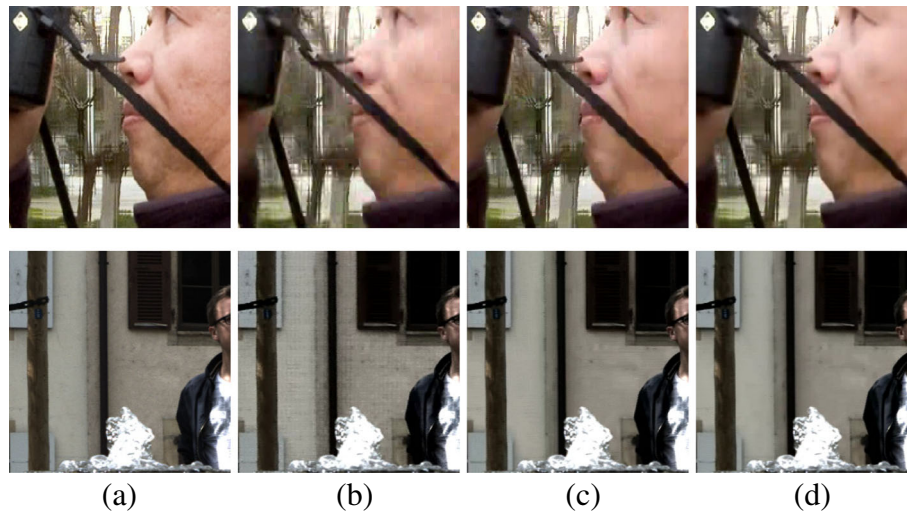
Page 13 of 14

Fig. 8 Visual rendering views from the decoded LF images at a similar bit-rate: **a** the original image, **b** HEVC intra-standard, **c** HEVC-SCC standard, and **d** the proposed SK-MMSE prediction method. The bit-rate for Jeff is 0.11 bpp and the bit-rate for I09_Fountain_&_Vincent_2 is 0.102 bpp

the proposed scheme can achieve a better coding efficiency compared to other coding methods. The other is that the proposed SK-MMSE prediction method can keep the detail information of EIs in the prediction process effectively.

6 Conclusions

In this paper, we propose a scalable kernel-based MMSE estimation method to effectively compress the LF image. The coding blocks are divided into three layers. In the HPL, a hybrid kernel-based MMSE estimation and IntraBC method are proposed to predict the coding blocks to improve the prediction accuracy of coding blocks in non-homogenous texture area, which explores the idea of using the IntraBC scheme or intra-directional prediction to find the best prediction of the coding block when K-MMSE estimation method fails based on the rate-distortion optimization (RDO) procedure. In the KPL, we propose to use the K-MMSE estimation method to predict the coding blocks to ensure the coding efficiency for homogenous texture area. In the LPL, we propose to predict the coding blocks by directly using a linear prediction method and do not compute the correction vector. The linear prediction method can be seen as a simplified form of the K-MMSE estimation method. In order to decide which layer is belonged to for the current coding block accurately, we propose to use the prediction mode correlation to decide which blocks belong to HPL and use the gradient information to measure the visual flatness.

The experimental results demonstrate that the proposed SK-MMSE method can compress the light-field image efficiency. It outperforms the HEVC intra-standard with 1.61 and 1.42 dB average quality improvements with Y-PSNR and YUV-PSNR as the objective quality metric, respectively. With regard to the computation complexity, the proposed SK-MMSE method can save around 16.4 and 70.5% average coding time than the K-MMSE estimation method in encoder side and decoder side, respectively.

Future work will include complexity reduction and how to further improve the prediction accuracy for texture and edge regions.

Authors' contributions
PA designed and conceived the research. ZY performed the simulated experiments and DL analyzed the experimental results. ZY wrote the manuscript. PA and DL edited the manuscript. All authors read and approved the final manuscript.

Authors' information
Zhixiang You received his B.S. degree from Nanjing University in 2005, and M.S. degree from Tsinghua University in 2008. He is currently pursuing the Ph.D. degree in communication and information systems from Shanghai University, Shanghai, China. His research interests include algorithms and systems for 3D and VR imaging, processing, analysis, and quality assessment.

Ping An received her B.S. and M.S. degrees from Hefei University of Technology, Hefei, China, in 1990 and 1993, respectively, and the Ph.D. degree in communication and information systems from Shanghai University, Shanghai, China, in 2002. She is currently a professor in School of Communication and Information Engineering, Shanghai University. Her research interests include stereoscopic and three-dimensional vision analysis and image and video processing, coding, and application.
Deyang Liu received his BS degree from Anqing Normal University, Anqing, China, in 2011, and his MS degree Ph.D. degree in Signal and Information Processing from Shanghai University, Shanghai, China, in 2014 and 2017. He is currently a lecturer in School of Computer and Information, Anqing Normal University. His research interests include 3D video processing, light-field image coding, and scalable light-field video coding.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Shanghai Institute for Advanced Communication and Data Science, Shanghai 200444, China. [2]School of Communication and Information Engineering, Shanghai University, Shanghai 200444, China. [3]School of Computer and Information, Anqing Normal University, Anqing 246000, China.

References

1. R Yang, X Huang, S Li, C Jaynes, Toward the light field display: auto stereoscopic rendering via a cluster of projectors. IEEE Trans. Vis. Comput. Graphics **14**(1), 84–96 (2008)
2. EH Adelson, JR Bergen, in *Computational Models of Visual Processing*. The plenoptic function and the elements of early vision (MIT Press, Cambridge, 1991), pp. 3–20
3. M Levoy, P Hanrahan, in *Proc. 23rd Annu. Conf. Comput. Graph. Interact. Techn*. Light field rendering (1996), pp. 31–42
4. F Liu, G Hou, Z Sun, T Tan, High quality depth map estimation of object surface from light-field images. Neurocomputing **252**, 3–16 (2017)
5. M Levoy, Light fields and computational imaging. Computer **39**, 46–55 (2006)
6. T. Ebrahimi, JPEG PLENO Abstract and Executive Summary, ISO/IEC JTC 1/SC 29/WG1 N6922, Sydney, Australia, 2015.
7. M. P. Tehrani, S. Shimizu, G. Lafruit, T. Senoh, T. Fujii, A. Vetro, et al., Use cases and requirements on free-viewpoint television (FTV), ISO/IEC JTC1/SC29/WG11 MPEG N14104, Geneva, Switzer-land, 2013.
8. C Yan, H Xie, D Yang, J Yin, Y Zhang, Q Dai, Supervised hash coding with deep neural network for environment perception of intelligent vehicles. IEEE Trans. Intell. Transp. Syst. **19**(1), 284–295 (2018)
9. C Yan, H Xie, S Liu, J Yan, Y Zhang, Q Dai, Effective Uyghur language text detection in complex background images for traffic prompt identification. IEEE Trans. Intell. Transp. Syst. **19**(1), 220–229 (2018)
10. C Yan, Y Zhang, J Xu, F Dai, L Li, Q Dai, F Wu, A highly parallel framework for HEVC coding unit partitioning tree decision on many-core processors. IEEE Signal Processing Letters **21**(5), 573–576 (2014)
11. C Yan, Y Zhang, J Xu, F Dai, J Zhang, Q Dai, F Wu, Efficient parallel framework for HEVC motion estimation on many-core processors. IEEE Transactions on Circuits and Systems for Video Technology **24**(12), 2077–2089 (2014)
12. LFR Lucas, C Conti, P Nunes, LD Soares, NMM Rodrigues, CL Pagliari, EAB da Silva, SMM de Faria, in *2014 Proceedings of the 22nd European Signal Processing Conference (EUSIPCO)*. Locally linear embedding-based prediction for 3D holoscopic image coding using HEVC (2014), pp. 11,15,1–11,15,5
13. R. Monteiro et al., Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 1–4, 2016.
14. D Liu, P An, R Ma, L Shen, in *2015 IEEE China Summit & Int. Conf. Signal and Information Processing (ChinaSIP)*. Disparity compensation based 3D holoscopic image coding using HEVC (2015), pp. 201–205
15. D Liu, P An, R Ma, C Yang, L Shen, K Li, Three-dimensional holoscopic image coding scheme using high-efficiency video coding with kernel-based minimum mean-square-error estimation. J. Electron. Imaging. **25**(4), 043015–1–043015–9 (2016)
16. D Liu, P An, R Ma, C Yang, L Shen, 3D holoscopic image coding scheme using HEVC with Gaussian process regression. Signal Process. Image Commun. **47**, 438–451 (2016)
17. C Conti, LD Soares, P Nunes, HEVC-based 3D holoscopic videocoding using self-similarity compensated prediction. Signal Process.Image Commun. **42**, 59–78 (2016)
18. C Conti, P Nunes, LD Soares, *HEVC-based light field image coding with bi-predicted self-similarity compensation*, 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (2016), pp. 1–4
19. Y Li, M Sjostrom, R Olsson, U Jennehag, in *IEEE Transactions on Circuits and Systems for Video Technology*. Coding of focused plenoptic contents by displacement intra prediction, vol 26, no. 7 (2016), pp. 1308–1319
20. F Dai, J Zhang, Y Ma, Y Zhang, *Lenselet image compression scheme based on subaperture images streaming*, 2015 IEEE International Conference on Image Processing (ICIP), Quebec City, QC (2015), pp. 4733–4737
21. D Liu, L Wang, L Li, FW ZhiweiXiong, W Zeng, *Pseudo-sequence-based light field image compression*, 2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW) (2016), pp. 1–4
22. G Wang, W Xiang, M Pickering, CW Chen, in *IEEE Transactions on Image Processing*. Light field multi-view video coding with two-directional parallel inter-view prediction, vol 25, No. 11 (2016), pp. 5104–5117
23. P Helin, P Astola, B Rao, I Tabus, *Sparse modelling and predictive coding of subaperture images for lossless plenoptic image compression*, 2016 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON) (2016), pp. 1–4
24. L Li, Z Li, B Li, D Liu, H Li, *Pseudo sequence based 2-D hierarchical coding structure for light-field image compression*, 2017 Data Compression Conference (DCC) (2017), pp. 131–140
25. D Liu, P An, R Ma, X Huang, L Shen, in *2017 Pacific-Rim Conference on Multimedia (PCM)*. Hybrid kernel-based template prediction and intra block copy for light field image coding (2017)
26. J Koloda, AM Peinado, V Sanchez, Kernel-based MMSE multimedia signal reconstruction and its application to spatial error concealment. IEEE Trans. on Multimedia **16**(6), 1729–1738 (2014)
27. D Persson, T Eriksson, P Hedelin, Packet video error concealment with Gaussian mixture models. IEEE Trans. Image Process. **17**(2), 145–154 (2008)
28. M. Rerabek, T. Bruylants, T. Ebrahimi, F. Pereira, and P. Schelkens, Call for proposals and evaluation procedure. ICME 2016 grand challenge: light-field image compression, Seattle, USA pp. 1–8, 2016.
29. T. Georgiev, 2013 (Online), Available: http://www.tgeorgiev.net, Website (Online). Accessed 1 July 2017.
30. HEVC SCC Reference Software Ver. 3.0 (SCM-3.0). [Online]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/. Accessed 31 Aug 2017.
31. H. Yu, R. Cohen, K. Rapaka, and J. Xu, Common test conditions for screen content coding, document JCTVC-X1015, 2016.