HOLOGRAPHIC INFORMATION CODING

José Peixeiro
Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal
jose.peixeiro@ist.utl.pt

Abstract— Holographic representations are the most loyal representations of the real world, as changing the user orientation to the hologram also changes the perspective of the scene/object in a very smooth and transparent way without the use of any glasses. With the explosion of digital technologies, also holograms migrated to digital representation and models. This step forward led to the development of many processes allowing to create and reproduce holograms only with the help of a computer, thus overcoming all the limitations and difficulties associated to the physical creation and reproduction of holograms. However, the visual richness of holograms is naturally associated to large amounts of data. Thus, without compression, the storage and transmission of holographic data would require a too large bandwidth; as a consequence, compression is a must for holographic data in order reasonable transmission and storage rates after coding are achieved. This work has the goal of developing an efficient coding solution for holographic information. In order to improve the holographic information coding, the adopted solution uses previously optimized transforms to allow a better adaptation to the input data characteristics. After, these optimized transforms are implemented and tested in the HEVC codec. Using the optimized transforms to perform natural images coding, the obtained results shows a performance improvement comparing with the standard HEVC coding performance. On the other hand, it is not consistently verified when coding the holographic data. The qualitatively result of the coding performance when using optimized transforms depends on the component and the holographic data format being coded.

Index Terms— Holography, hologram, digital holography, hologram creation, hologram reconstruction, hologram compression, optimized transforms.

I. INTRODUCTION

The world of multimedia has always been a very exciting one for our society and multimedia content takes an important role in many different areas, notably education, communication, art, science and entertainment. Due to this growing interest, more realistic ways of representing the world around us have been developed and improved over the years. Naturally, to improve the user experiences, three-dimensional (3D) representations assume a special role and have become more and more popular, generating great excitement. 3D imaging solutions display the visual content by using the same dimensions of the real world and so 3D representations take an increasingly important role in many application scenarios. From the various possible 3D representation solutions, holography emerges as the most realist method of representing objects and scenes as it faithfully reproduces the light field associated to a scene. Despite its importance, 3D stereo displays suffer from the so-called vergence-accommodation conflict as the eyes do not converge and focus at the same spatial point, thus eventually causing headache, nausea or visual fatigue in watching people. In holographic systems, the convergence and focus points are the same as in real life, which is a big advantage of holography regarding more conventional 3D systems. Holography is the science or practice of producing a hologram, which may be defined as a 3D imaging technique [1]. A hologram is a 3D representation of an object or scene, also known as a 3D photograph. A hologram records and reproduces everything that is visible to the human eyes, notably depth, size, shape, texture and relative position [2]. This technique records all the relevant information about the original object or scene and so, under optimal conditions, there should be no differences between the hologram and the real object or scene being represented. Holography must not be confused with other 3D imaging and display techniques which use conventional lens imaging, like lenticular and parallax barrier stereo and autostereoscopic 3D displays.

Although it may sound unexpected, holograms are not a very recent invention. The holographic theory has been invented by the Hungarian scientist Dennis Gabor, in 1948, while trying to improve the resolution of electron microscopes [3]. The development of computer technology has also created the need and the possibility to have holograms in a digital format and also to record and reproduce holograms by simply using computer resources, thus leading to the progress of digital holography. Digital holography refers to the methods used to reconstruct holographic images from physically recorded holograms and also the methods used to create a hologram from a real or virtual image simply requiring computational resources [4]. The remainder of this paper is organized as follows: Section 2 review the reference system and the holographic representation formats. Section 3 presents the adopted coding solution used to improve the holographic information coding and Section 4 assesses the adopted solution performance when coding holographic data and natural images. Finally, Section 5 presents the final conclusions and will conclude with the plan for the future work.

II. HOLOGRAPHIC MAIN CONCEPTS AND REPRESENTATION FORMATS

After many years of research and development, generating a hologram by means of the practical physical process has found interesting and easier alternatives, notably the generation of digital holograms using computers. In fact, the physical process may be complex since associated to several practical constraints and limitations, this is what makes the so-called Computed Generated Holograms (CGH) increasingly popular. To appropriately and efficiently represent and code the holograms, the data defining to the holograms must adopt a certain representation format; the relevant formats will be reviewed at the end of this section.

A. Reference System

To better organize this paper, the digital holography reference system presented in Fig. 1 will be considered.

Fig. 1: Reference system for digital holography.

The reference system considers three main phases:

- **Acquisition** – This is the phase where the holographic data is generated using a large variety of processes. While optical generation is possible with a variety of processes and sensors, e.g.
direct measurement of the fringes associated to the interference patterns or indirect creation through point clouds, computer generated digital holograms are increasingly more common and there the objects don’t even need to have physical existence. The acquisition process provides the digital holographic data in a variety of representation formats to be presented in the next section.

- **Coding** – Here the huge amounts of digital holographic raw data represented using several alternative representation formats must be coded with an appropriate coding solution. While available coding solutions may be used, taking the holographic data as regular ‘luminance’ data, it is naturally possible to develop specific coding solutions which exploit the intrinsic characteristics of the holographic data. It is also possible to extend available standard coding solutions to better consider the holographic data features, e.g. through some additional coding modes.

- **Reconstruction** – Finally, the holographic data needs to be decoded to allow the reconstruction/rendering of specific views to be consumed by the user. This may happen using optical processes involving the illumination of a physical hologram where the interference pattern diffracts incident light to reproduce the original light field, thus recreating the initial physical objects and a realistic user experience involving visual depth cues. Alternatively, the view reconstruction may happen computationally using appropriate reconstruction models which replicate the optical process using some holographic 3D display. Finally, 2D reconstruction is also possible where views with different perspectives and focus are created for a regular 2D display or to 3D stereo/autostereoscopic displays.

**B. Hologram Representation Formats**

The holographic data represents the interference fringes of the reference and object wavefields. To be able to reproduce good quality holograms, a clean complex object field at the hologram plane has to be obtained. With optical creation, this typically involves the use of the so-called phase-shifting holographic method which samples and records three times the interference wavefield intensity with different phase shifts applied to the reference wavefield, creating the so-called interferograms. With CGH, the complex field is directly computed and thus only two components, amplitude and phase, are needed. In this context, the various representation format options are presented in the following:

- **Intensity Information or Interferograms:** The first format corresponds to the three interferograms which directly represent the intensity of the complex wavefield for three different phase shifts of the reference wave, this means $I_1$, $I_2$ and $I_3$ with phase shifts of 0, $\pi/2$ and $\pi$, respectively.

- **Phase Shifted Distances:** As the three interferograms above correspond to a huge amount of data if high resolutions are used and since the complex wavefield has only two components, amplitude and phase, the three interferograms may lead to a representation using only two components, the so-called phase shifted distances format, $D_1$ and $D_2$. These components are expressed as:

$$D_1(x, y) = I_1 - I_2$$
$$D_2(x, y) = I_3 - I_1$$

- **Real/Imaginary:** This format represents the interference wavefield as a complex number using a Cartesian coordinate system this means using two components: the Real and Imaginary parts of the complex wavefield.

- **Amplitude/Phase:** This format represents the same complex interference wavefield using a polar coordinate system where the two components are now the Amplitude and Phase.

Each representation format component was converted into 8 bit representation, and so one is represent with 8 bits per sample.

**III. CODING MODE DEPENDENT TRANSFORMS: ADOPTED SOLUTION**

Analyzing the holographic data components illustrated in Figure 2, some directionality are clearly observed. This higher correlation along certain directions is mainly observed in the background of each holographic data component, with the background occupying a rather significant percentage of the holographic component area.

![Figure 2: Bunny’s holographic data components: Left) Amplitude component; Right) Phase component.](image)

HEVC uses variants of the DCT and DST, which are data-independent transforms, this means transforms that are not adapted to the data to be coded, and so their compression performance is naturally penalized by their non-adaptability properties. Regarding the coding solutions available in the literature, the JPEG 2000 extension using Discrete Adaptive Directional Wavelet Transforms to exploit the holographic data directionality [6] shows a very good compression performance, notably outperforming the JPEG 2000 standard. Taking this evidence and the holographic data characteristics into account, it may be concluded that one possible approach to improve the holographic data compression performance is to use directional dependent transforms adapted to the HEVC Intra prediction coding modes which are associated to different prediction directions. Thus, the goal is to adopt directional transforms adapted to the directionality of each Intra prediction mode instead of a single (DCT or DST) transform which is agnostic to the coding mode directionality.

There are many possible ways to design directional transforms with different properties. The ideal one would be a low complexity process producing a transform providing a good compression performance this means a good trade-off between rate and distortion. Regarding the computational complexity, orthogonal transforms are easier to implement as they only require a matrix multiplication. Moreover, non-separable transforms perform better than separable transforms in terms of compression performance. Based on these elements, it was decided that this work aims improving the compression performance for holographic data by designing directional transforms which should be both orthogonal and non-separable. Despite their higher computational complexity, non-separable transforms may reach a better compression performance. Since the adopted solution will adopt non-separable transforms, for an $M\times M$ residuals block, an $M^2\times M^2$ optimized transform will be designed, where $M$ is the transform size. After reviewing the literature, it was also decided that the adopted coding mode dependent transform solution should design orthonormal mode-dependent transforms using a rate-distortion optimization criterion. The adopted solution is based on a solution proposed by Osman Sezer in his PhD Thesis, “Data-driven Transform Optimization for Next Generation Multimedia Applications” [8] and also adopted later by Adria Arrufat in his PhD Thesis “Multiple Transforms for Video Coding” [7]. These two references are fundamental for the solution developed and implemented in this Thesis.

To reach a more efficient holographic data codec, the designed coding
mode dependent transforms will be integrated in the HEVC Reference Software as this codec is the current state-of-the-art on image and video coding. As static holographic data is to be coded, the developed coding solution is based on the HEVC Intra coding modes and no Inter coding modes are used.

The development and implementation of the improved holographic data codec involves three main steps, notably Residuals Extraction, Rate-Distortion Optimized Transform Calculation and HEVC Codec RDOT Integration. Each one of these steps is precisely described in the following sections.

A. Residuals Extraction

As explained above, the developed coding solution adopts optimized transforms depending on the HEVC Intra prediction mode and transform size. Since the transform is applied to the Intra prediction residuals, to learn the best transforms (from a compression performance perspective) a set of prediction residuals for each Intra prediction mode and each transform unit size was obtained by running the HEVC Reference Software for some pre-selected training data. Most Intra prediction modes correspond to a specific Intra prediction direction, these residuals contain information depending on the directionalities of the input content, in this case, information about the directionalities of the holographic data. As the directional dependent transforms are obtained from these residuals a representative set of these residuals for each HEVC transform unit size and Intra prediction mode must be extracted from appropriate training data obtained through the HEVC Reference Software, in this case using the HEVC Reference Software. As the transform computation process should be as faithful as possible to the original residuals data, the extracted residuals are those calculated before the quantization process.

Residuals from five different holographic data elements were extracted in this stage: three from the ParisTech dataset (Bunny, Luigi and Girl) and two from the Interfere-I database (3D Multi and 3D Venus). Each holographic data element was coded using six different quantization parameters, notably 12, 17, 22, 27, 32 and 37. Thus, each holographic data element has six sets of residuals coming from six different quantization conditions. The quantization parameters were selected by simply adopting the quantization parameters recommended in the HEVC test conditions (22, 27, 32 and 37) and adding two lower quantization parameters (12 and 17) to obtain a more representative set of smaller transform block sizes.

B. Rate-Distortion Optimized Transform Calculation

This step computes the optimized transforms based on the previously collected residuals data. This method performs an algebraic transform optimization able to exploit the correlation between the directional edges in the residuals data to increase the final compression efficiency [8], thus, reaching a better trade-off between the rate and distortion. The main goal is to find the best orthonormal (orthogonal and with unitary norm) transform minimizing a Lagrangian cost that expresses the trade-off between the rate and distortion while adopting some rate and distortion estimation metrics. Note that rate and distortion are estimated since real values are not easily available (for example, the rate would have to be determined using the HEVC entropy coding) [8]. The initial transform is a non-separable transform computed using the Kronecker product and using a separable DCT similar to the one used in the HEVC codec. This initial transform is iteratively refined taking into consideration the appropriate set of residuals extracted in the training process and the selected Lagrangian cost function that will be presented below. The idea is to deviate from the initial transform only if the residuals are more efficiently coded using another set of basis functions.

The transform optimization is an iterative process in which first the optimized coefficients for a given transform are calculated and after the optimized transform is calculated accordingly to the optimized coefficients. The optimization process is divided into three main steps: Initialization, RDOT Calculation Loop and Stopping Criteria Checking; each of these main steps contains various smaller steps. Figure 3 shows a fluxogram with the smaller steps of the RDOT calculation; these steps are applied to a specific residuals set corresponding to a certain Intra prediction mode and transform size. Since to compute a precise estimation for the rate is not easy (it would require using the HEVC entropy coding), the rate is estimated using a nonlinear approximation, the $L_0$ norm, $(||c||_{L_0})$ which measures the sparsity of the transformed coefficients. This norm counts the number of non-zero elements in one matrix/vector [8]. Regarding the distortion, it is estimated using the Euclidean norm [8]. The HEVC Reference Software distortion metric is not used because this method is adapted to the Euclidean norm, particularly the optimized transform expression is based on estimating the distortion using the Euclidean norm. Also, the Lagrangian multiplier expression is obtained based on the Euclidean and $L_0$ norm for estimating distortion and rate, respectively.

More specifically, the following expression represents the overall transform optimization process [8]:

$$G_{opt} = \arg \min_{G} \sum_{c_j} \min_{x_i} \left( ||x_i - G^Tc_j||^2 + \lambda ||c_j||_0 \right)$$

where $G_{opt}$ is the optimized transform, $x_i$ the residuals block for a specific holographic component, $x'_i$ the reconstructed residuals block using the optimized transform and $c_j$ the quantized transform coefficients. While the first term expresses the distortion estimation, the second term expresses the rate estimation.

Figure 3: Fluxogram with the main steps of the RDOT optimization.

In detail, each step of the transform optimization process works as follows:

**Initialization:**

1. **Initial Transform Coefficients Calculation:** The optimization process starts by calculating the initial transform coefficients using the initial transforms, which are, in this case, the result of the Kronecker product in order to obtain an $M^2 \times M^2$ non-separable transform from two $M \times M$ separable transforms. The initial transform starts the iterative process, which will finally converge to the optimized transform. The expression used to calculate the initial coefficients, $c_{i, initial}$, is:

$$c_{i, initial} = G_{initial} \cdot x_i$$

where $G_{initial}$ is the result of the Kronecker product calculated with the floating-point value DCT-II or DST-VI, depending on the TU size. The DCT and DST are used in HEVC because in general they provide a good RD trade-off, also non-separable transforms allow a better compression performance than separable transforms.
2. **Initial Transform Coefficients Hard-Thresholding:** This method aims to find an orthonormal transform producing sparse coefficients offering the minimum distortion for a given quantization step \([7][8]\). This sparseness is achieved in this step by hard-thresholding the transformed coefficients using a Lagrangian multiplier, \(\lambda\) \([7][8]\):

\[
c_{i,\text{initial}}[n] = \begin{cases} c_{i,\text{initial}}[n] & |c_{i,\text{initial}}[n]| \geq \sqrt{\lambda} \\ 0 & \text{otherwise} \end{cases}
\]

(1.4)

where \(c_{i,\text{initial}}[n]\) represents the \(n\)th transform coefficient. Regarding the specific (and critical) value to be taken by the Lagrangian multiplier, many approaches can be followed. This work will use the approach suggested in \([7]\) where the \(\lambda\) value depends on the quantization step \(\Delta\) as:

\[
\lambda = \frac{\Delta^2}{4}
\]

(1.5)

As shown in \([7]\), this Lagrangian multiplier value obtains the optimal balance between the distortion and rate constraints.

3. **Initial Transform Rate and Distortion Estimation:** Using the transform coefficients after the hard-thresholding, the residuals are reconstructed using the following equation:

\[
x'_i = \text{round}(G_{\text{initial}}^T c_{i,\text{initial}})
\]

(1.6)

where round() represents the rounding to the nearest integer. While the distortion is estimated using the Euclidean norm between the reconstructed and the original residuals, the rate is estimated using the \(L_0\) norm over the coefficients after performing the hard-threshold as shown in Equation (1.4). Replacing the rate and distortion estimations, the initial transform Lagrangian cost is calculated as:

\[
J_{\text{initial}}(\lambda; G_{\text{initial}}, x_i) = D(G_{\text{initial}}, x_i) + \lambda R(G_{\text{initial}}, x_i)
\]

(1.7)

where \(D(G_{\text{initial}}, x_i)\) and \(R(G_{\text{initial}}, x_i)\) are, respectively, the distortion and rate estimations using the transform \(G_{\text{initial}}\) for the residual block \(x_i\) and \(\lambda\) is the Lagrangian multiplier controlling the rate versus distortion trade-off.

**RDOT Calculation Loop:**

4. **Optimized Transform Definition:** This step has now the goal to compute the optimized mode-dependent transform. For this, a sparse orthonormal transform is calculated using the coefficients calculated previously in the step 2; while the first optimization iteration uses the initial transform coefficients, the other iterations already use the optimized coefficients corresponding to the optimized mode-dependent transform under construction. The \(M^2\times M^2\) optimized transform is obtained as \([8]\):

\[
G_{\text{opt}} = UV^T
\]

(1.8)

where \(U\) and \(V\) are obtained by applying a singular value decomposition to \(Y\) which is defined in Equation (1.9). The singular value decomposition allows to decompose one matrix into three different ones as shown in Equation (1.10). This matrix \(Y\) is calculated using the available optimized coefficients and the residuals blocks set by \([8]\):

\[
Y = \sum_{i=0}^{N} c_{i}x_i^T
\]

(1.9)

The singular value decomposition of \(Y\) is expressed as \([8]\):

\[
Y = USV^T
\]

(1.10)

The detailed explanation of this step may be found in \([8]\).

5. **Optimized Transform Coefficients Calculation:** Here, the optimized transform coefficients, \(c_{i,\text{opt}}\), are calculated using the same procedure applied in Step 1 but using the optimized transform calculated in the previous step.

6. **Optimized Transform Coefficients Hard-Thresholding:** This hard-thresholding is performed in the same way as in Step 2. As the Lagrangian multiplier value is independent of the used transform, it simply takes the same value as above.

7. **Optimized Transform Rate and Distortion Estimation:** Both the rate and distortion are estimated here in the same way as in Step 3 while using now the most recent optimized transform and the optimized coefficients to calculate the reconstructed residuals using as before the Euclidian and \(L_0\) norms for the distortion and the rate, respectively.

**Stopping Criteria Checking:**

8. **Convergence Checking:** As mentioned above, this optimization process is an iterative process and thus some stopping criterion is needed to avoid an infinite number of iterations. When convergence is reached, the optimization process stops. In this work, convergence is declared when the difference between the Lagrangian cost of two consecutive iterations is, in module, lower than 0.001, this means:

\[
\Delta J = |J_{\text{iteration}} - J_{\text{iteration-1}}| < 0.001
\]

(1.11)

where \(J_{\text{iteration}}\) and \(J_{\text{iteration-1}}\) represents the Lagrangian cost of the current and the previous iterations, respectively. If \(\Delta J\) is lower or equal to 0.001, convergence is reached and the process stops as the cost is not significantly changing anymore with further iterations; if \(\Delta J\) is higher than 0.001, the process continues to the next step.

9. **Iteration Number Checking:** Sometimes the optimization process may take too long to reach convergence. As it is not desirable to increase too much the computational time of the whole optimization process, a second stopping criterion is used in this work. This stopping criterion limits the number of optimization iterations; if the iteration limit is reached, the optimization process stops and the transform providing the lowest cost from all tested is considered the optimized transform. If the number of iterations limit has not been reached, the process goes back to Step 4 for further optimization. After several experiments, it was considered that 200 is a good value to limit the number of iterations while offering a good trade-off between computational time and RD performance.

10. **Optimized Transform Scaling:** After calculating the optimized transform for a specific set of residuals, and since the optimized transforms are learned as floating-point matrices, they need to be scaled to fit the dynamic range used in HEVC DCT and DST default transforms. This is important in order the optimized transform may substitute the default transform in the HEVC Reference Software without scaling problems. Therefore, the last step of the RDOT design process is to scale the floating-point optimized transform. The scaling performed is different from the one performed for the floating-point HEVC DCT and DST. In this case, a scaling factor of \(2^{64N}\) will be used, where \(N = \log_2(M)\); this scaling will be explained in the following section.

At this stage, the RDOT optimization process is finished and the set of optimized \(M^2 \times M^2\) transforms, allowing a better RD trade-off for each Intra prediction mode and transform unit size, may be integrated in the HEVC Reference Software as described in the next section.

**C. HEVC Codec RDOT Integration**

After the previous step, the RDOT, \(G_{\text{opt}}\), this means the optimized transforms are available and thus only the last codec integration step is missing; at this stage, \(G_{\text{opt}}\) needs to be integrated in the HEVC Reference Software code. To perform the RDOT integration, some changes were made to the HEVC Reference Software, notably:

1. **Transforms Calculation Approach:** HEVC uses (DCT and DST)
separable transforms; thus, in the HEVC Reference Software the transform process is implemented using the separable approach. As the designed optimized transform solution follows a non-separable approach, it is necessary to modify the HEVC Reference Software to support these new type of transforms. The non-separable approach calculates the transform coefficients and the reconstructed residuals using the following equations:

\[
\begin{align*}
    c_{nt} &= \text{round}(G_{nt}x) \\
    x' &= \text{round}(G^T_{nt}c_{quant,nt})
\end{align*}
\]  

where \(G_{nt}\) is the non-separable transform, \(c_{nt}\) are the non-separable transformed coefficients before quantization, \(c_{quant,nt}\) are the non-separable transformed coefficients after quantization, \(x\) is the original residuals block and \(x'\) is the reconstructed residuals block.

2. Transforms Scaling Factor: The HEVC Reference Software scaling factors, shown are adapted to the DCT and DST and thus to separable transforms. The first scaling factor to modify is the transform scaling. As referred in the previous section, in the last step of the RDOT definition, the optimized transform is scaled using a scaling factor equal to \(2^{6N}\). The transform scaling factor was changed because non-separable transforms are applied once, unlike separable transforms that are applied twice (to the horizontal and vertical directions). As these scaling factors are applied as shifts on the HEVC Reference Software, it is only possible to use shifts equal to integer values. Thus, to avoid shifts equal to floating point values, the new transform scaling factor is \(2^{6N}\) which maintains the same scaling factor of a forward/inverse transform total scaling, that depends on N/2.

3. Forward and Inverse Transform Scaling Factors: As non-separable transforms are applied in a single step, another difference between the non-separable and separable approaches is that the first only needs one scaling factor for the forward and inverse transforms (while the second transform approach needs two). Beside the transforms scaling factors, the HEVC Reference Software applies scaling factors in the quantization and dequantization processes. These two scaling factors depend, respectively, on the scaling factors performed in the forward transform process and the inverse transform process. To maintain the same global forward and inverse transforms scaling, the scaling factors were adapted. In summary, the multiplication of the new transform scaling factor and \(S_{T1}\) must be equal to \(2^{(15-B-N)}\), on the other hand the multiplication of the new transform scaling factor and \(S_{T12}\) sum must be equal to \(2^{(15-B-N)}\). Table 1 summarizes the new scaling factors, where \(S_{T1}\) and \(S_{T11}\) represent the new forward and inverse transform scaling factors, respectively.

<table>
<thead>
<tr>
<th>New Scale Factor</th>
<th>New Scale Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>New transform scaling</td>
<td>(2^{6N})</td>
</tr>
<tr>
<td>(S_{T1})</td>
<td>(2^{(16-28-N)})</td>
</tr>
<tr>
<td>(S_{T11})</td>
<td>(2^{(21-B)})</td>
</tr>
<tr>
<td>Forward transform total scaling</td>
<td>(2^{(15-B-N)})</td>
</tr>
<tr>
<td>Inverse transform total scaling</td>
<td>(2^{(15-B-N)})</td>
</tr>
</tbody>
</table>

4. Transforms Switch: As the RDOT definition depends on the residual blocks initially extracted to build the training set, it is possible that some combinations of Intra prediction modes and TU sizes were not used and, consequently, no mode-dependent transform may be defined for these combinations. The last change applied in the HEVC Reference Software was the insertion of a switch that decides between using a RDOT and using the HEVC default transforms (separable DCT or DST). The HEVC default transforms are used whenever no optimized non-separable transform is available for a specific Intra prediction mode of a specific TU size.

After applying all the necessary software changes, the last step towards the design of a codec including dependent transforms is completed. Now, the mode dependent transforms coding solution performance is ready to be assessed.

IV. MODE DEPENDENT TRANSFORM BASED CODING OF HOLOGRAPHIC DATA: PERFORMANCE ASSESSMENT

In this section, the mode dependent transform based coding solution defined in the previous section will be assessed in comparison with relevant coding alternatives.

A. Test Material and Coding Conditions

The holographic test material used for this experiment was courteously provided by Prof. Frédéric Dufaux from Paris Tech. The holographic content in this dataset is computer generated, based on three well-known 3D virtual models [5], notably Bunny, Luigi and Girl. The ParisTech dataset has been made available already converted into 4 different holographic representation formats: i) intensities of three interferograms (I1, I2 and I3); ii) Amplitude/Phase; iii) Real/Imaginary; and iv) Phase Shifted Distances (D1 and D2). Figure 4 and Figure 5 show the components of the various representation formats data for the Bunny object displayed as a luminance image.
performance) phases, notably the quantization parameters will take the values 12, 17, 22, 27, 32 and 37 to consider various levels of quality. In this section, the Main profile is used.

Figure 6: Natural test images: (a) Blowing Bubbles; (b) Basketball Pass and (c) Bus.

For each type of data, holographic and natural images, the optimized transforms for each test element will be created using as training data the residuals obtained for the remaining elements of the same type, e.g. for one image the optimized transforms will be obtained using the residuals extracted from the HEVC Intra coding of the remaining images. However, to assess the impact of transform adaptation on the coding solution RD performance, results will be also provided for the situation where the residuals used for the transform optimization are obtained from the test element (i.e. the test and the training element are the same).

B. Performance Assessment Methodology

In the proposed performance assessment methodology for the holographic data domain, the hologram representation format components before and after coding are compared using appropriated metrics. Note that in this domain the reconstruction resulting from holographic data is not considered, i.e. this domain does not take into account what is being displayed/viewed by the user.

As mentioned above, the holographic data in the ParisTech dataset correspond to 2D matrices with a floating-point representation. In this context, in order to be able to perform the holographic data domain performance assessment described above, the following steps are required:

- **Scaling and Quantization** - First, it is necessary to apply a suitable transformation over the floating point values to obtain an 8-bit representation for each data sample; the standard image coding solutions, such as the HEVC codec, only accept as input data in integer representation (typically with 8- or 10-bit depth samples).
- **HEVC Coding Solution** - At this stage, the HEVC Intra coding solution is applied to each hologram representation format component represented with 8-bit depth samples (obtained from the previous stage).

At the end of the process, both the original and decoded hologram representation format components are in 8-bit integer precision and can, therefore, be compared.

To assess the quality of each decoded hologram representation format component, the PSNR metric will be used, as it provides a reliable metric to assess the fidelity of the decoded data against the original data. The PSNR metric is calculated between the original hologram data and the decoded hologram in 8-bit integer representation. The performance assessment methodology for the natural images is similar to the one described above except for the scaling and quantization process, which is not needed; the test natural images are already in an (8-bit) integer representation. As for the holographic data, the PSNR, BD-PSNR and BD-Rate metrics will be used to compare the original and the decoded natural images.

C. Assessing the Improved HEVC Codec

This section will present the RD performance results obtained with the developed HEVC extension using the optimized transforms determined using the RDOT process detailed in the previous chapter and other relevant coding solutions. First, the optimized transforms based codec will be applied to natural images and after to holographic data. As before, due to space constraints, this section will show charts for some representative situations, notably first for the Blowing Bubbles image and after for the Real, Amplitude and Phase components of the Bunny holographic data element.

I) RD Performance: Natural Images

To obtain a better understanding of the behavior of the developed codec, the RD performance will be presented for various coding cases and configurations, notably:

1. **Standard HEVC (labeled HEVC):** Corresponds to standard HEVC coding using the (separable) DCT and DST available in the HEVC reference software.

2. **Codec using all optimized transforms obtained with the residuals from the same natural image being coded (labeled All_Opt_Own):** Corresponds to the case where the image is coded using optimized transforms determined using its own residuals; it allows to have an idea of the benefits when the transform is fully adapted 'ideal' in terms of residuals, i.e. an upper bound of the RD performance gains. This means that the training set above defined for Blowing Bubbles is not used, instead the residuals obtained from the previous HEVC coding of Blowing Bubbles itself are used to determine the optimized transforms.

3. **Codec using all optimized transforms obtained with the residuals from the remaining natural images (labeled All_Opt_Other):** This is the more natural way of using the proposed optimized transforms coding as it uses the residuals set from the remaining images to determine optimized transforms for all combinations of TU sizes and Intra prediction modes; these optimized transforms are applied for the image being evaluated. This means that the training and testing set of residuals are completely separated.

4. **Codec using the optimized transforms only for the angular Intra prediction modes obtained with the residuals from the remaining natural images (labeled Angular_Opt_Other):** This solution is similar to the previous one but now only the angular Intra prediction modes use optimized transforms; for the other modes, the standard HEVC DCT/DST is applied.

5. **Codec using the optimized transforms only for the DC and Planar Intra prediction modes obtained with the residuals from the remaining natural images (labelled as DC/Planar_Opt_Other):** This codec is similar to the previous one but now only the DC and Planar Intra prediction modes use optimized transforms; for the angular modes, the HEVC DCT/DST is applied.

6. **Codec using the non-separable DCT/DST transform obtained by applying the Kronecker product to the default HEVC DCT/DST transforms (labelled as Non-Separable_DCT/DST):** This codec uses the non-separable DCT/DST transforms resulting from using the default HEVC transforms (DST for 4x4 Intra and DCT for remaining TU sizes and modes) to calculate the Kronecker product.

Figure 7 shows the RD performance curves for the various tested coding solutions. Analyzing the RD performance results in Figure 7, some interesting conclusions may be taken:

- The standard HEVC codec has the worst performance, highlighting that the use of any optimized transforms is always beneficial for natural images as already reported in the literature. This behavior validates the RDOT process and the developed optimized transforms coding solution as it implies that the various
optimized transforms are adapted to the image residuals set characteristics.

- The non-separable DCT coding solution achieves a very good RD performance thus confirming what has been said in the previous chapter: the non-separable approach (non-separable_DCT) allows a better RD performance than the separable approach (HEVC) mainly because it exploits all the spatial correlation of the pixels inside a block.

- As expected, the optimized transforms codec using the residuals from the same natural image to determine the specific used optimized transforms reaches the best RD performance. As, in this case, the optimized transforms are adapted to the specific residuals set of the image being coded, this best RD performance allows to confirm that the developed solution is really creating transforms adapted to the residuals set.

- The coding solution using the optimized transforms determined using the training set created with the remaining images and the non-separable DCT coding solution have very similar RD performances. This may be justified by the small training set as it only contains two other natural images. Anyway, the coding solution with all the optimized transforms allows some RD performance improvements regarding the standard HEVC.

- The codec using only the optimized transforms for the angular modes reaches a better RD performance than the codec using only the optimized transforms for the Planar and DC modes showing that the transforms computed for directional Intra prediction modes are responsible for significant performance gains.

Regarding the Real component, by analyzing the RD performance results in Figure 9, the following conclusions were obtained:

- As expected, and already occurring for previous cases, the codec using the residuals from the same Bunny’s component to determine the optimized transform reaches the best RD performance.

- The non-separable DCT codec achieves the second best RD performance, outperforming all the ‘realistic’ approaches that use disjoint training and testing sets.

- The optimized transforms determined using the training set created with the remaining holographic data elements obtains better RD performance, outperforming the standard HEVC codec. This means that the training set residuals allows a reasonable transform adaptation to the residuals set used for the Bunny’s Amplitude. Although this approach outperforms the standard HEVC codec, it is outperformed by the non-separable DCT, which seems to imply that the optimization process was not able to create adapted transforms to outperform the non-separable DCT codec.

- The codec using only the optimized transforms for the angular modes also allows reaching a better RD performance than the standard HEVC.

- The codec using only optimized transform for the DC and Planar modes also outperforms the standard HEVC. It may be concluded that this coding approach reaches a better RD performance than the codec using only the optimized transforms for the angular modes. This means that there is a better RDOT adaptation to the DC and Planar modes than to the angular modes.

2) **RD Performance: Holographic Data**

Figure 8, Figure 9 and Figure 10 show the RD performances for holographic data for the same codecs already considered for the natural images. The RD performance curves use the same codec labels as for the natural images.

Regarding the Real component, by analyzing the RD performance results of Figure 8 the following conclusions were obtained:

- As expected, and already occurring for natural images, the codec using the residuals set from the same Bunny’s component to determine the optimized transforms reaches the best RD performance.

- The non-separable DCT codec achieves a very good RD performance, even outperforming all the approaches using the training set residuals from the other holographic data elements to calculate the optimized transforms.

- The codec with the optimized transforms determined using the training set created with the remaining holographic data elements reaches the worst performance, and is outperformed by the standard HEVC RD performance. This fact means that exploiting all the optimized transforms does not allow to reach a better RD performance than the standard HEVC RD performance as expected and as occurred for natural images.

- The codec using only the optimized transforms for the angular modes reaches a RD performance very similar to the standard HEVC RD performance. It may be concluded that the number of angular modes residuals in the training set is not enough to allow the optimized transforms to adapt to the angular modes residuals characteristics.

- The codec using only optimized transform for the DC and Planar modes is outperformed by the standard HEVC coded, showing that the optimized transforms are not able to adapt to the DC and Planar modes characteristics. This may be justified by the fact that the DC mode is not a directional mode and so it may not have the characteristics to which to adapt for all the holographic data elements.
Regarding the Phase component, by analyzing the RD performance results in Figure 10 the following conclusions were obtained:

- Again, the codec using the residuals from the same Bunny’s component to determine the optimized transforms reaches the best RD performance. In this case, the difference between this RD performance and the other codecs’ RD performance is not as clear as for the other holographic components.

- The non-separable DCT codec achieves a good RD performance, outperforming the standard HEVC RD performance and the RD performance of the codec using the optimized transforms for the DC/Planar modes.

- The optimized transforms determined using the training set created with the remaining holographic data elements reaches a RD performance slightly above the non-separable DCT codec. This behavior shows that the optimized transforms are well adapted to the Intra prediction modes characteristics. On the other hand, this RD performance difference is small maybe due to the small training set used, only containing 4 other holographic data elements.

- The codec using only the optimized transforms for the angular modes reaches the second best RD performance. This fact may be justified by the frequent use of angular modes in the training set, thus, the optimized transforms are able to adapt to the characteristics of the Phase angular modes.

- The codec using only the optimized transforms for the DC and Planar modes reaches the second worst RD performance, only outperforming the standard HEVC RD performance. The explanation in the previous bullet also fits here: the frequent use of the angular modes in the training set results in a poor optimized transforms adaptation for the DC and Planar modes and thus does not allow to reach a better RD performance.

D. Analyzing Mode-Dependent Transform based Coding of Holographic Data

As it was concluded in the previous section, using mode-dependent optimized transforms for coding natural images allows to obtain a better RD performance than the standard HEVC Intra coding. Unlike natural images, the coding performance improvement when using (mode-dependent) optimized transforms based coding of holographic data depends on the holographic component being coded. This means that the data characteristics influence the performance gains that the proposed type of transform may achieve. To better understand the shortcomings of the proposed transforms for holographic content, three metrics will be used, notably:

1. Residuals Variance: The objective of this metric is to assess the residuals values distribution around the mean value. Thus, for each natural image and for each holographic data component, three steps are performed:
   i) The residuals mean block is calculated;
   ii) After, it is calculated the squared difference between each pixel of the residuals mean block and the corresponding pixel of each residual block. For each residual block, the squared differences are summed and divided by the block number of pixels. The variance of a residual block is computed as:
   \[
   \sigma^2_b = \frac{1}{M^2} \sum_{i=1}^{M} \sum_{j=1}^{M} (x_{ij} - \bar{x}_{b})^2
   \]  
   (1.13)
   where \( M \) represents the block size, \( x_{ij} \) represents the residual value at the \((i, j)\) position within the \(b\)th block being assessed, \( \bar{x}_{b} \) represents the residual value at the \((i, j)\) position within the mean block and \( \sigma^2_b \) represents the variance of the \(b\)th block.
   iii) The last step calculates the variance of all residuals blocks by performing the average on the variances of all blocks.

2. Transformed Coefficients Compactness: This metric intends to assess how well the transform basis functions can represent the residuals set by evaluating the energy compactness of the transformed coefficients. This metric simply counts the average number of non-zero transformed coefficients when a transform is applied to a residuals block.

In the following, the results obtained for each one of these three metrics will be presented. For each metric, the results obtained for the Blowing Bubbles image will be used as reference, as the RD performance for this image behaves as expected. Due to space limitation, results will be only provided considering residuals resulting from HEVC Intra coding with a QP equal to 22.

Residuals Variance

This metric is applied separately to each TU size, mixing all the Intra prediction modes, and the results obtained are presented in Table 2. As this metric intends to represent the variance per TU size, each TU size will have one value representing the variance of the residuals set belonging to that TU size; in Table 2, ‘x’ corresponds to a TU size that was not used in the HEVC coding and so no residuals were evaluated regarding that specific TU size.

As explained above, the results shown in Table 2 allow to compare the residuals variance of the Bunny holographic data element with the Blowing Bubbles natural image for each TU size. The variance assumes higher values for a residuals set containing a higher variation around the mean residuals value, e.g., the higher the variance value, the higher is, on average, the distance between each residuals block and the residuals blocks’ mean. Assessing the residuals variance assumes, therefore, an important role because it helps to understand
how difficult is adapting the RDOTs for residuals sets whose distribution is spread around the mean value. In general, the analysis of the Table 2 results shows that the adaptation of optimized transforms is easier for the Blowing Bubbles image residuals than for the Bunny holographic components residuals. This may partly explain the difference in the RD performance shown in Sections C.1) and C.2) between natural and holographic content for the optimized transforms.

It can also be observed from Table 2 that, regarding the holographic data elements, the 16x16 and 32x32 TU’s sizes of the Bunny’s Phase component reach lower (residual) variance values than the corresponding ones in the Blowing Bubbles image. Although those TU sizes are not the most commonly used in the Bunny’s Phase component coding, they certainly contribute to the overall coding solution RD performance gain regarding the ‘pure’ HEVC coding solution.

<table>
<thead>
<tr>
<th></th>
<th>Blowing Bubbles</th>
<th>Bunny’s Real Component</th>
<th>Bunny’s Amplitude Component</th>
<th>Bunny’s Phase Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>4x4</td>
<td>167.51</td>
<td>x</td>
<td>x</td>
<td>2729.60</td>
</tr>
<tr>
<td>8x8</td>
<td>82.71</td>
<td>x</td>
<td>x</td>
<td>293.09</td>
</tr>
<tr>
<td>16x16</td>
<td>102.38</td>
<td>x</td>
<td>x</td>
<td>76.38</td>
</tr>
<tr>
<td>32x32</td>
<td>X</td>
<td>222.81</td>
<td>223.33</td>
<td>63.11</td>
</tr>
</tbody>
</table>

Transformed Coefficients Compactness

This metric is applied separately to each combination of TU size and Intra prediction mode. Figure 11, Figure 12, Figure 13 and Figure 14 show the compactness metric results obtained respectively for the Blowing Bubbles image and the Bunny’s Real, Amplitude and Phase components. To assess the compactness metric, three different transform solutions used in Section C will also be used here: notably HEVC, Non-Separable_DCT and All_Opt_Other. Note that this metric computes the number of non-zero coefficients after performing a quantization equal to the HEVC quantization with a certain quantization parameter. Thus, the residuals used in this study are the residuals associated with the HEVC coding of a specific data element using the specified quantization parameter. The QP value of 22 was selected since it represents an average quality, i.e. it lies in the center of the RD curves presented in the previous sections.

Analyzing Figure 11, it may be concluded that the number of non-zero coefficients for the Blowing Bubbles image is similar for all the three transform solutions but the All_Opt_Other solution, which has, on average, a higher number of non-zero coefficients than the other two solutions. This similarity is clearer for the smallest TU size, 4x4, than for the other TUs sizes. These results are expected since according to the RD results on Figure 7, the RD performance associated with quantization parameter 22 (fourth RD point counting from the left) of the All_Opt_Other solution has a rate slightly higher than the HEVC solution. Note that the number of non-zero coefficients has a direct relation with the coding rate. The higher the number of non-zero coefficients within a block the higher will be the rate needed to code that specific block, since the zero coefficients do not need to be directly coded and transmitted.

Figure 11: Compactness metric for the Blowing Bubbles image.

Figure 12 shows the compactness for the Bunny’s Real component. It may be observed from Figure 12 that the number of non-zero coefficients is similar for non-separable DCT and standard HEVC solutions. Also, for the optimized transform solution, a higher number of non-zero coefficients is obtained when compared to the other two DCT based solutions. This result shows that the optimized transforms are not well adapted to the residuals set characteristics and, thus, the residuals blocks’ energy is not compacted in a smaller number of coefficients compared with the standard HEVC transforms and the non-separable DCT solution.

Figure 13 shows the compactness for the Bunny’s Amplitude component. It may be concluded from Figure 13 that the compactness metric results are similar to the Real component ones. In terms of RD performance, when applying the optimized transforms to the Bunny’s Amplitude component, in general, it outperforms the standard HEVC but again the RD point associated with the quantization parameter 22 achieves a rate a little bit higher than the ‘pure’ HEVC solution.

Figure 13: Compactness metric for the Bunny’s Amplitude component.
Analyzing the Figure 14., it is possible to concluded that for the 4x4 TU size of the Bunny’s Phase component, all the three transform solutions achieve similar number of non-zero coefficients. However, the All_Opt_Other solution achieves a slightly higher compactness, i.e. the number of non-zero coefficients is lower than for the other two non-optimized solutions. Unlike the smallest TU size, the 8x8, 16x16 and 32x32 TUs sizes present a higher number of non-zero coefficients for the optimized transforms with respect to the HEVC standard transforms. The difference between the number of non-zero coefficients of the optimized transforms and the HEVC standard transforms is smaller in the 8x8 TU size than in the 16x16 and 32x32 TUs sizes. As shown in the Figure 10, the RD point associated with the quantization parameter 22 in the All_Opt_Other curve has a rate higher than the HEVC solution, so these results were expected.

Figure 14: Compactness metric for the Bunny’s Phase component.

The analysis performed in this Section attempts to characterize the characteristics of the holographic data and the efficiency of the optimized transforms when applied to holographic data elements. The first two proposed metrics (residuals block energy and variance) allow to conclude that the (Intra) prediction residuals of the holographic data have a higher variance when compared to the natural images. Also, the compactness metric shows that the optimized transforms are not able to achieve lower values of non-zero coefficients when compared with the HEVC standard transforms and the non-separable DCT transform. This fact is explained by the conclusions taken from the first metric. This means that the optimization process fails more often when adapting transforms to a residuals set exhibiting a high variability; a high variance value means that the residuals set may contain such diverse characteristics that makes harder the creation of an optimized transform capable of efficiently approximate – i.e. with a few transform coefficients’ number – that residuals set.

This section intends to prove that it is harder to obtain an optimized transform for the holographic data elements (than for natural images) due to the more diversified characteristics of their Intra prediction residuals. Note that not only the residuals distribution will influence the RD performance of the optimized transforms. The residuals Intra prediction modes distribution will also influence the optimized transforms creation. Although the holographic data residuals have a higher variance than the natural images ones, it is not impossible to reach improvements in the RD performance when using the optimized transforms. As it is shown in the Bunny’s Phase component RD performance (Figure 10), using the optimized transforms to code holographic data components containing explicit directionalities, and so using more often angular Intra prediction modes, allow a RD performance improvement comparing with the standard HEVC RD performance. Though, due to the holographic data characteristics this performance improvement is not as high as the performance improvement obtained in the natural images.

V. CONCLUSIONS AND FUTURE WORK

In summary, it can be concluded that the adopted solution allows improving the standard HEVC coding performance. This improvement is clearer in the natural image than in the holographic data elements; also, the RD performance of hologram depend on the holographic component that is coded. Compared with the HEVC standard DCT/DST transforms, the Phase and Amplitude components achieve a higher coding performance when using the optimized transforms, but the same is not verified for the Real component. The metrics evaluated 4 allow to explain the differences in the RD performance, when using the optimized transforms, between the holographic data elements and the natural image. These metrics show that the prediction residuals are more distributed (i.e. have a higher variance) when compared with the prediction residuals for the natural images and typically have higher energy. As the adopted solution intends to improve the coding performance by adapting the optimized transforms to the residuals characteristics, the adopted solution has a lower performance when the optimized transforms need to be adapted to residuals that exhibit higher variance and energy. Not only the residuals energy will impact the optimized transforms RD performance, the directionalities of the holographic components will also impact the optimized transforms RD performance. Despite the high energy and variance of the holographic residuals, the holographic components containing strong directionalities, and consequently using more often angular Intra prediction modes, are able to reach a better RD performance when using the optimized transforms comparing with the standard HEVC coding.

Since the coding of holography components is a relatively new topic and is nowadays emerging there are many interesting directions that can be followed to obtain a coding solution with higher performance for this type of data. Thus, some improvements are possible to achieve more consistent coding performance improvements, such as: performing clustering in the rate-distortion optimized transform creation to eliminate outlier residuals that may impair the optimized transforms creation; some pre-processing of the holographic data aiming to obtain holographic data with characteristics closer to natural images, naturally without losing any depth information, for instance a denoising filter; using Wavelet Transforms instead of using DCT and DST.

In conclusion, there is still a lot of work and research that needs to be performed in the holographic data coding field to allow the holograms to be efficiently represented and transmitted over bandwidth limited channels.

REFERENCES