IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

Towards Visualization and Searching: a Dual-Purpose Video Coding Framework

RENAM C. DA SILVA[1,†], (Member, IEEE), FERNANDO PEREIRA[2], (FELLOW, IEEE), AND EDUARDO A. B. DA SILVA[3], (Senior Member, IEEE)

[1]Universidade Federal do Rio de Janeiro, Brasil (e-mail: renam.silva@smt.ufrj.br)
[2]Instituto Superior Técnico – Universidade de Lisboa and Instituto de Telecomunicações, Portugal (e-mail: fp@lx.it.pt)
[3]PEE/COPPE/DEL, Universidade Federal do Rio de Janeiro, Rio de Janeiro 21941-972, Brasil (e-mail: eduardo@smt.ufrj.br)

†This work was done while the author was a graduate student at Universidade Federal do Rio de Janeiro

ABSTRACT To offer more powerful video-enabled applications, it is increasingly more critical not only to visualize the decoded video but also to provide efficient searching capabilities for similar content. Video surveillance and personal communication are critical application examples asking for these dual visualization and searching functionalities. However, the current video coding solutions are strongly biased towards visualization needs. In this context, this paper introduces a novel dual-purpose video coding framework targeting both visualization and searching needs by adopting a hybrid coding approach where the usual pixel-based coding approach is combined with an innovative feature-based coding approach. In a particular implementation of this dual-purpose video coding framework, some frames are coded using a set of keypoint matches, which not only allow decoding for visualization, but also provide the decoder valuable feature-related information, extracted at the encoder from the original frames, instrumental for efficient searching. Experimental results for video surveillance and personal communications scenarios show competitive performance regarding the state-of-the-art HEVC standard both in terms of visualization and searching performances.

INDEX TERMS descriptor, dual-purpose video coding, HEVC, keypoint, matching, searching, visualization

I. INTRODUCTION

DIGITAL video applications have been exploding in recent years both in variety and amount of captured, stored, transmitted, searched and visualized data. This explosion has required successive generations of video coding standards, creating a continuous saga towards more efficient video coding solutions. The most recent video coding standard, the High Efficiency Video Coding (HEVC) standard [1] [2], jointly developed by ISO/IEC MPEG and ITU-T SG16 VCEG, is the result of decades of research work and investment. Through its multiple profiles, it offers highly efficient video coding solutions for a wide variety of applications, from video surveillance and personal communications to UHD television and streaming. HEVC and all the previous video coding standards adopt a pure pixel-based video coding approach, which essentially targets visualization capabilities and thus visual quality. However, with the increasing amount and omnipresence of digital video, users are increasingly not just visualizing the decoded video but also using it for other purposes, notably searching for similar visual content. This is happening in many application domains, notably in video surveillance and personal communications, where the decoded video is often used for searching in very rich, available databases. Naturally, besides good visual quality, it is also critical to provide good searching performance. However, with the current video coding approach, the searching process has to rely on keypoints and descriptors extracted from the lossy decoded video, sometimes with rather low quality, what usually significantly penalizes the searching performance, especially for lower rates.

In this context, this paper proposes a novel dual-purpose video coding framework that is more adjusted to the current role of digital video in modern applications. In these applications, the decoded video should be friendly and efficient both in terms of visualization (visual quality performance) and searching (descriptor matching performance). This is achieved by proposing an architecture adopting a hybrid approach where pixel-based and feature-based coding are jointly used. First, periodic k-frames are coded using a standard pixel-based approach and used as reference frames to code the f-frames using a feature-based approach. For GOP sizes longer than 2, also the k-frames may become reference frames in a hierarchical GOP structure. Once the k-frames are decoded, frame rate-up conversion is performed to obtain a first coarse estimation of the f-frames. The basic idea to code the f-frames is to refine this coarse estimation by migrating appropriate image patches from the decoded reference frames. This is achieved by establishing correspondences between features/patches in the original f-frames and the already available decoded reference frames. In this way, the quality of the f-frames may be gradually improved by reusing appropriate image patches from the reference frames, relying on the fact that video sequences usually exhibit significant temporal redundancy. In addition, since keypoint positions extracted from original uncompressed video data are available for the f-frames at the encoder, the visual searching performance may be boosted compared with the performance associated to decoder-extracted keypoints based on lossy decoded video.

This is a conceptually refreshing coding approach which tries to conciliate some degree of backward compatibility with HEVC, the most recent video coding standard (through the k- frames) with a new video coding approach targeted at boosting the searching performance (through the f-frames). The proposed dual-purpose video coding framework is flexible enough to allow adjusting the balance between the visualization and searching capabilities, up to the extreme cases where one of them is dominating, depending on the specific application scenario requirements. This framework offers a synergetic video coding approach between two key user capabilities, which, to the authors' knowledge, is rather unusual in the literature. The dual-purpose conceptual framework, which the authors claim to be the main contribution of this research work, builds on novel and challenging concepts and tools to the video coding landscape.

The performance assessment of a specific design and implementation of the proposed video coding framework has been carried out using the very efficient HEVC coding standard as reference, thus setting a very challenging benchmark. Although such an assessment involves comparing video coding solutions with very different degrees of maturity, investment and optimization, the performance results have shown that the implemented instance of proposed video coding framework is competitive both in terms of visual quality and descriptor matching performances.

This paper is organized as follows: after this first section where the context, motivation and objectives of the proposed video coding framework are presented, Section II provides a review of the related background literature. Then, Section III presents the architecture and the processing pipeline of the implemented instance of proposed video coding framework while Section IV details the most novel and technically original coding modules. Section V presents the performance assessment using representative test conditions and meaningful benchmarks and metrics, taking video surveillance and personal communication as the proof-of-concept application scenarios. Finally, Section VI closes the paper with final comments and future work directions.

II. BACKGROUND REVIEW

Developments in computer vision have led to the emergence of new forms of visual information representation which are better suited for visual analysis tasks than just pixel-based representations. Local visual features are a powerful type of such representations, which have been playing a central role in modern digital image and video applications such as mobile visual search, object recognition, and scene classification. Such local features describe image characteristics that are distinctive, representative and informative, usually by first performing keypoint detection to identify salient image regions and then extracting a descriptor to capture the local characteristics. Th Scale Invariant Feature Transform (SIFT) [4] and Speeded-Up Robust Features (SURF) [8] are two major description tools in this context. Following this recent trend, distributed visual analysis systems, for instance, may aggregate a huge amount of data captured from multiple and distributed visual sensors and perform complex visual analysis, targeting to provide services such as augmented reality in sport events, behavior analysis in security systems and mobile visual search [3] [9] [10]. The latter is a rather mature and increasingly popular application that uses local visual features to retrieve, from a remote server, relevant information for a query image or video. In this context, three main approaches have been considered to meet different constraints when performing feature-based analysis in scenarios involving remote searching. These are the Compress-then-Analyze (CTA), the Analyze-then-Compress (ATC) and the Hybrid-Analyze-Then-Compress (HATC) approaches [6] [11] [12]. In the CTA approach, the remote analysis is carried out using visual features extracted from compressed, transmitted and decompressed video content, thus enabling also visualization. However, the compression usually has a detrimental effect in the extracted visual features, which in turn impairs the visual analysis performance. Some works have tried to modify existing standard image and video coding solutions to better preserve the features of interest [13] [14] [15].

On the other hand, in the ATC approach, the visual analysis performed at the remote server has to solely rely on the set of compressed visual features extracted and transmitted by the sender. Naturally, such approach has the drawback of not enabling visualization at the server side, which limits the range of applications [16]. For this approach, a significant amount of work has been done with several authors proposing coding schemes to efficiently compress state-of-the-art local visual features, such as SIFT and SURF, both for images and videos [6] [17] [18]. Also, new visual feature descriptors have been carefully designed, targeting lower bitrate representations [19] such as the so-called binary descriptors [20]. Still in the ATC domain, the recently issued MPEG-CDVS (Compact Descriptors for Visual Search) standard [9] [10] provides description tools to enable interoperability in the context of image searching.

Finally, the HATC approach aims at overcoming the limitations of the two previous paradigms by combining pixel-based and feature-based coding, notably offering simultaneously visualization capabilities (decoded video) and original extracted features (not extracted from decoded video). Considering that visualization and searching are becoming very popular together, the HATC approach has recently attracted attention. In [21] [22], an image coding solution based on SIFT descriptors is proposed already inspired by the technique reported in [23]. SIFT descriptors are extracted from the original image and differentially coded with respect to SIFT descriptors extracted from a poor quality, down-sampled and low rate version of the image, that is first conveyed to the decoder. This first image is used to guide the target quality reconstruction, since it should carry enough information about the edges, colors and objects. The decoded descriptors are used to retrieve highly correlated images available in the cloud, which shall provide image patches to enable a higher quality image reconstruction. In the context of scene classification and pedestrian detection, a two-part predictive coding architecture is proposed in [7], targeting both the signal (image) and feature fidelities. Related systems are proposed in the contexts of Visual Sensor Networks (VSN) [12] and augmented reality applications [24]. In [11], a video coding solution is proposed where keypoint information detected on the uncompressed video frames is coded in parallel with regularly coded video, thus not exploiting their mutual synergies. It was experimentally demonstrated that keypoints detected on uncompressed video are effective in reducing the detrimental effects of compression on feature matching performance, even if the descriptors themselves are extracted from lossy decoded video [11]; this highlights the critical importance of using keypoint information extracted from uncompressed data for efficient searching.

In these previous HATC works, pixel and feature-based representations are essentially designed and used independently from each other, meaning that the feature-level data, targeting searching is not exploited to aid the pixel-level coding, targeting visualization, and vice-versa. But this scenario is starting to change. In [16], a hybrid framework for jointly coding the feature descriptors and the visual content is proposed, exploiting their interaction. While the feature descriptors are efficiently represented by taking advantage of the structure and motion information in the compressed video stream, the already compressed descriptors can be used to further improve the video compression efficiency by applying feature matching based affine motion compensation. The novel video coding solution proposed in this paper also adopts the HATC approach; however, differently from [16], the proposed solution explicitly codes just the keypoint data detected at the encoder from original video as the descriptors themselves are extracted at the decoder using the keypoint based reconstructed f-frames. The proposed solution is based on a flexible, joint Lagrangian optimization framework where pixel-based and feature-based processing are combined to find the most appropriate trade-off between the visualization and searching performances. Moreover, the proposed solution provides quality scalability for the f-frames and some degree of backward compatibility with the latest video coding standard HEVC with the k-frames. A very preliminary version of this coding solution, not performing joint optimization of the visualization and searching performances, and not quality scalable, has been published in [25].

III. PROPOSED DUAL-PURPOSE VIDEO CODING FRAMEWORK: ARCHITECTURE

The proposed dual-purpose video coding (DPVC) framework combines pixel-based and feature-based coding to provide a powerful and efficient coding framework towards both visualization and searching. While the pixel-based component provides backward compatibility with the most efficient visualization-targeted video coding standard, the feature-based component boosts the searching performance by providing precise keypoint locations, extracted from the original, uncompressed video content. By targeting simultaneously two key functionalities, the dual-purpose coding process has to consider both a visual quality distortion, D_V, and a descriptor matching distortion, D_M, which assess the visualization and searching performances, respectively. The proposed dual-purpose coding architecture is presented in Figure 1 and its processing pipeline is briefly described in the following.

The original video frames are split in two sets, namely the so-called *k-frames* and *f-frames*. While the k-frames are coded using a conventional video coding solution, the f-frames are coded using a feature-based approach, explicitly making the overall coding framework searching friendly. After frame splitting, the k-frames are Intra coded and decoded using a standard video codec; in this case, the state-of-the-art HEVC standard is used [1]. To restore the original frame rate and effectively estimate regions with smoother spatial and temporal evolutions, an initial estimate of the f-frames is obtained by interpolating them from the neighboring decoded reference frames. This is performed both at encoder and decoder using a block-based motion-compensated frame interpolation algorithm using as references the closest past and future available decoded reference frames [35]. To determine the best patches from the reference frames to improve the interpolated f-frames, the original f-frames and the decoded reference frames feed a keypoint detection module which identifies the most distinctive positions inside a frame in terms of feature-based characterization and thus searching performance. For each keypoint, a descriptor is extracted to capture the local image patch. The objective is that only a parsimoniously selected number of keypoint matches are conveyed to the decoder as there is an associated rate cost. Their aim is twofold. First, they indicate the areas in the interpolation estimated f-frames that may have their quality more improved for visualization using decoder available patches. Second, they indicate distinctive f-frame areas likely to be correctly matched to a visual content database available at the receiver side as this improves the searching performance,

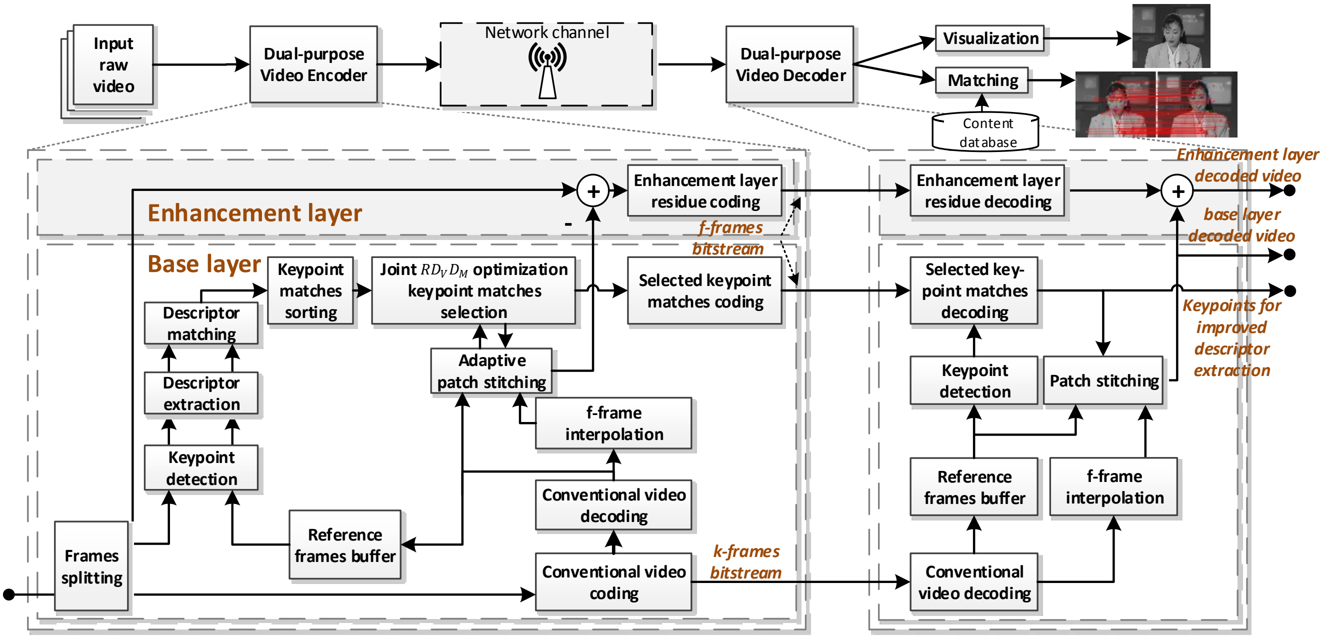

FIGURE 1. Architecture of the proposed dual-purpose video coding (DPVC) framework.

which in turn is underpinned by the ability to match descriptors properly [10]. In the proposed framework, at the decoder side, local descriptors are extracted at these reliable and parsimoniously selected keypoints in the f-frame. Figure 1 (upper right) shows an example of matching areas in two images (one decoded and another at the decoder database) which have been identified by matching their descriptors. To be able to improve the f-frames estimates with patches from the reference frames, the f-frames descriptors are matched to the reference frames descriptors using the Euclidean distance as matching metric. The intuition is that the matching descriptor pairs represent regions with similar visual content in the f-frames and reference frames.

IV. PROPOSED DUAL-PURPOSE VIDEO CODING SOLUTION: CODING TOOLS

This section targets the detailed presentation of the most novel and critical modules in the proposed dual-purpose video coding framework. In this context, before going any further, let us represent each visual feature by the pair $\{\mathbf{p}_{n,i}; \mathbf{d}_{n,i}\}$ where $\mathbf{p}_{n,i}$ denotes the vector with the keypoint position (x, y), scale σ and angle θ of the i-th feature in frame n and $\mathbf{d}_{n,i}$ the associated descriptor vector, *e.g.* SIFT coefficients.

This video coding solution is just the first implementation of the proposed dual-purpose video coding framework that will be used here to show its potential, notably using a very tough benchmark, the HEVC standard.

ADAPTIVE PATCH STITCHING

The patch stitching process targets improving the f-frames quality with appropriate patches extracted from the already available (decoded) reference frames. In the stitching process, the image patch $I_m^{(k)}|_{\Omega_{m,j}^{(k)}}$ defined over the region $\Omega_{m,j}^{(k)}$, centered at a selected keypoint location $(x_{m,j}, y_{m,j})^{(k)}$ from a reference frame $I_m^{(k)}$, is extracted and seamlessly stitched over the region $\Omega_{n,i}^{(f)}$, centered at the matching keypoint location $(\hat{x}_{n,i}, \hat{y}_{n,i})^{(f)}$ in the relevant f-frame $I_n^{(f)}$, thus generating the stitched f-frame $I'^{(f)}_n$. The superscripts $^{(k)}$ and $^{(f)}$ refer to the k-frames and f-frames, respectively. In the variables above, the subscripts m and n indicate the m-th and n-th frames, whereas j and i indicate the j-th and i-th keypoints; the hat $\hat{}$ over a variable indicates quantization. For simplicity, circularly-shaped patches are used in this work. The diameters of the image areas involved in the stitching process depend on the scale parameters $\sigma_{m,j}^{(k)}$ and $\hat{\sigma}_{n,i}^{(f)}$ of the matching keypoints. The reference frame patch diameter is $\mathrm{m}_s\sigma_{m,j}^{(k)}$ and the f-frame destination region diameter is $\mathrm{m}_s\hat{\sigma}_{n,i}^{(f)}$, where m_s is a scale parameter factor that is adaptively determined for each patch at the encoder, as explained in the sequel, and is coded in the bitstream sent to the decoder.

The stitching process aims to keep unchanged the pixel values both over and outside the boundary $\partial\Omega$ of $\Omega_{n,i}^{(f)}$, while blending inside the pixel values of the patch $I_m^{(k)}|_{\Omega_{m,j}^{(k)}}$ (from the reference frame) seamlessly with those from the f-frame $I_n^{(f)}$. A comprehensive formulation of this problem is given in [23] [26]. In this paper, the core patch stitching process is performed using the Poisson stitching technique proposed in [26]. The patch stitching process is carried out using the non-quantized keypoint parameters of the decoded reference frame and the quantized keypoint parameters of the original f-frame as the second will have to be quantized when coding.

Given the matching keypoints $\hat{\mathbf{p}}_{n,i}$ and $\mathbf{p}_{m,j}$, the first one in the current f-frame (as reconstructed up to this point) and the second one in the most similar frame found in the reference buffer together with the corresponding frames $I_n^{(f)}$ and $I_m^{(k)}$, the stitching process proceeds as follows:

1. ***Initialization***: Set the current visual quality minimum distortion $D_{V,curr}$ as the distortion between the current f-frame and the corresponding original f-frame. The mean squared error (MSE) is used here to assess the visual quality distortion.
2. ***Support size adaptation***: For each m_s value in the selected range do: Let $I_m^{(k)}|_{\Omega_{m,j}^{(k)}}$ be the image patch defined over a circularly-shaped domain $\Omega_{m,j}^{(k)}$ defined by its diameter $\mathsf{m}_s\sigma_{m,j}^{(k)}$ and centered at the keypoint location $(x_{m,j}, y_{m,j})^{(k)}$ and $\Omega_{n,i}^{(f)}$ the destination region in the f-frame, centered at the keypoint location $(\hat{x}_{n,i}, \hat{y}_{n,i})^{(f)}$ with diameter $\mathsf{m}_s\hat{\sigma}_{n,i}^{(f)}$.

   a. ***Geometric transform***:

      i. Rotate the reference patch by $\varphi = \hat{\theta}_{n,i}^{(f)} - \theta_{m,j}^{(k)}$ and scale it by $s = \frac{\hat{\sigma}_{n,i}^{(f)}}{\sigma_{m,j}^{(k)}}$ around the point $(x_{m,j}, y_{m,j})^{(k)}$ applying the transformation:

$$A = \begin{bmatrix} \alpha & \beta & (1-\alpha)x_{m,j}^{(k)} - \beta y_{m,j}^{(k)} \\ -\beta & \alpha & \beta x_{m,j}^{(k)} + (1-\alpha)y_{m,j}^{(k)} \end{bmatrix} \quad (1)$$

      where $\alpha = s \cdot \cos\varphi$ and $\beta = s \cdot \sin\varphi$

      ii. Translate the reference patch to the appropriate f-frame position by applying:

$$T = \begin{bmatrix} 1 & 0 & \hat{x}_{n,i}^{(f)} - x_{m,j}^{(k)} \\ 0 & 1 & \hat{y}_{n,i}^{(f)} - y_{m,j}^{(k)} \end{bmatrix} \quad (2)$$

   b. ***Poisson stitching***: Carry out the stitching process as described in [26].

   c. ***Visual quality assessment***: Compute the visual quality distortion between the resulting stitched f-frame and the original f-frame for each successive m_s value. If the visual quality distortion is reduced regarding $D_{V,curr}$, then $D_{V,curr}$ is updated with the new minimum distortion value and the new best scale parameter factor m_s is adopted.

This process returns the stitched f-frame $I'^{(f)}_n$ with the support size $\mathsf{m}_s\hat{\sigma}_{n,i}^{(f)}$ providing the largest visual quality gain. At the decoder, the patch stitching process does not have to be adaptive as the appropriate m_s value is transmitted by the encoder as side information.

The adaptive patch stitching is the most computation demanding encoder module as it involves solving a system of linear equations for each candidate keypoint match. The adopted conjugate gradient method guarantees convergence in at most S steps [38], where S is the number of samples within the stitching region. The non-optimized DPVC implementation takes approximately 48s to code a CIF resolution frame, whereas the HEVC reference software (with QP = 25) takes around 1.47s; however, these numbers have a rather limited meaning since no serious efforts have yet been invested in optimizing the DPVC software. In the end, it may be true that the proposed DPVC solution is more complex but investing complexity is normal in video coding evolution since also the encoding platforms become more powerful in time. There are also application scenarios where encoding is performed offline and thus the encoding complexity is less critical and may be paid to buy other functionalities such as better searching at the decoder as proposed by the DPVC framework.

KEYPOINT MATCHES SORTING

The order by which the keypoint matches are considered in the joint RD_VD_M optimization process has a significant impact on the final performance, both in terms of visual quality as well as searching performance; therefore, it is essential to previously and appropriately sort the keypoint matches using some appropriate criterion as performing an exhaustive search over all possible keypoint matches arrangements is simply impractical due to the prohibitive computational cost. A reasonable solution is to evaluate each candidate keypoint match independently and sort them using a criterion which is able to express its effectiveness in contributing to reduce the visual quality distortion and the descriptor matching distortion (thus ultimately increasing the number of descriptor matches). Naturally, the quality of the descriptors extracted at the decoder to be used for the matching process strongly depends on the quality of the reconstructed frames. Thus, before carrying out the joint triple RD_VD_M optimization process, it is considered here that an appropriate criterion to perform the sorting is the MSE reduction, relative to the original f-frame, associated to the refinement of the interpolated f-frame using the image patch corresponding to a specific keypoint match.

To reduce the complexity associated to the Poisson stitching process integrated in the sorting process, the potential MSE reduction for each keypoint match is assessed by simply copying the image patch centered at the keypoint location in the reference frame to the matching keypoint location in the f-frame and computing the difference to the original f-frame. This is a low complexity stitching process which avoids solving the involved Poisson equation [26] [23] at the penalty of obtaining only an estimation of the MSE reduction; this is, however, enough for sorting purposes. Moreover, for complexity reasons, this process is performed for every keypoint match independently, implying that the cumulative effect of the keypoint matches is not considered. At the end, the list will include all the keypoint matches ordered by their MSE reduction potential.

JOINT RD_VD_M OPTIMIZATION KEYPOINT MATCHES SELECTION

The proposed dual-purpose video coding framework aims at delivering optimal visual quality for visualization and reliable keypoint information for searching. As previously outlined, to accomplish such objectives, the proposed coding framework combines pixel-based and feature-based approaches to represent the k-frames and f-frames arranged in a GOP structure. The periodic k-frames are coded using a standard video codec and also reused as source of image patches to improve the f-frames. In turn, each f-frame is coded using a feature-based approach on top of a first estimation obtained by motion-compensated frame interpolation using the available reference frames, only k-frames or also f-frames, depending on the GOP size.

More specifically, the f-frames are coded resorting to a set $\mathcal{M}_{kp}$ of keypoint matches $\hat{\mathbf{p}}_{n,i}^{(f)} \mapsto \mathbf{p}_{m,j}^{(k)}$, where $\hat{\mathbf{p}}_{n,i}^{(f)}$ (the hat over indicates quantization) belongs to the current f-frame and $\mathbf{p}_{m,j}^{(k)}$ to a reference frame in the buffer.

Such dual-purpose coding framework creates the challenge of allocating the bit budget to those keypoint matches which provide the best trade-off between reducing the visual quality distortion (visualization performance) and increasing the number of correct descriptor matches for the images in a given decoder content database (searching performance).

Measuring the visual quality distortion is straightforward as the availability of the original and decoded f-frames at the encoder facilitates the measurement of the distortion reduction associated to a specific keypoint match. However, the situation is very different for the searching capability, as the descriptor matching performance cannot be precisely measured at the encoder as only the decoder has access to the target content database. In the sequel, the joint optimization framework and joint keypoint matches selection process will be presented.

JOINT LAGRANGIAN OPTIMIZATION FRAMEWORK

The optimization goal is to select a set of keypoints matches, $\mathcal{M}_{kp}$, which minimize the following Lagrangian cost function:

$$\arg\min_{\mathcal{M}_{kp}^{*}} J = (D_V + \gamma D_M) + \lambda R(\mathcal{M}_{kp}) \tag{3}$$

where D_V is the visual quality distortion, D_M is descriptor matching distortion and $R(\mathcal{M}_{kp})$ is the total rate for coding the set of selected keypoint matches. The parameter γ weights the importance given to the searching performance regarding the visualization performance, while λ weights the overall rate regarding the combined distortion.

An iterative procedure to be presented in the sequel is adopted to determine the best set of keypoint matches M_{kp} which minimize the cost function as defined in Eq. (3). At each iteration, the benefit (cost function reduction) is evaluated in terms of rate, visual quality distortion and descriptor matching distortion.

- *Rate*

The rate for coding each candidate keypoint match $\hat{\mathbf{p}}_{n,i}^{(f)} \mapsto \mathbf{p}_{m,j}^{(k)}$ is computed as follows:

$$R = R(r_k) + R(r_m) + R(\mathsf{m}_s) + R(\mathbf{p}_{n,i}^{(f)} - \mathbf{p}_{m,j}^{(k)}) \tag{4}$$

where $R(r_k)$ is the rate to code the reference frame index which signals which of the reference frames in the buffer is used for patch stitching (out of two for GOP size 2); $R(r_m)$ is the rate to code the keypoint match index in the reference frame, following the known order provided by the extractor; $R(\mathsf{m}_s)$ is the rate to code the selected scale parameter factor; and $R(\mathbf{p}_{n,i}^{(f)} - \mathbf{p}_{m,j}^{(k)})$ is the rate to (lossy) encode the keypoint match parameters residues.

To reduce the computational complexity associated to the joint optimization step, the total rate is estimated by computing the self-information of each syntactic element according to the probability models as described in Section IV.

- *Visual quality distortion metric*

The MSE is adopted for visual quality distortion in this work. To decide whether or not to select a particular candidate keypoint match for coding, the encoder computes the MSE between the original f-frame and the resulting frame after performing the adaptive patch stitching.

- *Descriptor matching distortion estimation metric*

The descriptor matching distortion D_M should provide an objective way to assess the contribution of each candidate keypoint match to the searching performance. In this context, the encoder should ideally only spend rate on those keypoint matches likely to produce correct descriptor matches at the decoder side. As only the decoder has access to the target content database, the descriptor matching performance cannot be accurately measured at the encoder. Thus, it is proposed here to estimate this performance at the encoder by mimicking in the best possible way the descriptor matching steps that are performed at the decoder. Such descriptor matching performance estimation enables to formulate a joint Lagrangian optimization [27] [28] [29] framework as defined in Eq. (3) to trade-off the rate against the joint visual quality and descriptor matching distortion.

More precisely, it is proposed to estimate the searching performance based on the number of matches between the descriptors extracted from the reconstructed f-frames (at keypoint positions to be selected) and those extracted from the original f-frames (at keypoint positions detected at the original f-frame), somehow assuming that the database includes a frame rather similar to the original f-frame. For a reliable searching distortion estimation, each candidate descriptor match should satisfy both the ratio test [4] and the symmetric match criterion as it is reasonable to adopt at the encoder the same criterion usually adopted for performing the searching at the decoder. The ratio test intends to discard those matches for which the ratio between the distance for closest and the second-closest descriptor is greater than 0.8, while the symmetric match criterion increases matching consistency.

The proposed estimator for the descriptor matching distortion is simply defined in terms of the difference between the number of extracted descriptors (256 being the maximum in our framework as this has been considered enough) and the number of correctly matched descriptors. Figure 2 presents the procedure associated to the encoder estimation of the descriptor matching distortion. In detail, the descriptor matching distortion estimation proceeds as follows:

1. *Initial descriptor matching estimation*: At the beginning of the joint optimization process, the reconstructed f-frame, $I_R^{(f)}$, is equal to the interpolated f-frame, $I_I^{(f)}$, and thus an initial descriptor matching distortion estimation may be performed using only the set of descriptors extracted from the interpolated f-frame, here labeled as $\mathcal{DS}_I$ (in this case $\mathcal{DS}_R = \mathcal{DS}_I$ as the reconstructed f-frame is the interpolated f-frame at the beginning). As there are still no keypoint matches selected at this stage, the initial descriptor matching distortion estimation proceeds as follows (extreme left and right branches in Figure 2):

a. *Original f-frame keypoint detection and descriptor extraction*: Let $\mathbf{d}_{n,i}^{(f)} = \Psi(\mathbf{p}_{n,i}^{(f)}|I_O^{(f)})$ be a descriptor extracted at keypoint $\mathbf{p}_{n,i}^{(f)}$ detected in the original f-frame $I_O^{(f)}$ and $\mathcal{DS}_O$ the set of such descriptors.

b. *Interpolated f-frame keypoint detection and descriptor extraction*: Let $\bar{\mathbf{d}}_{k,i}^{(f)} = \Psi(\bar{\mathbf{p}}_{k,i}^{(f)}|I_I^{(f)})$ be descriptor extracted at the keypoint $\bar{\mathbf{p}}_{k,i}^{(f)}$ detected in the interpolated f-frame $I_I^{(f)}$ and $\mathcal{DS}_I$ the set of such descriptors.

c. *Descriptor matching for original versus interpolated f-frames*: Perform descriptor matching between the original f-frame descriptor set, $\mathcal{DS}_O$, and the interpolated f-frame descriptor set, as $\mathcal{DS}_R = \mathcal{DS}_I$.

d. *Descriptor matching distortion estimation*: Estimate the descriptor matching distortion between the original and interpolated f-frames descriptor sets according to:

$$D_M(\mathcal{DS}_R, \mathcal{DS}_O) = 1 - \frac{\sum_{i=1}^{|\mathcal{DS}_R|} \sum_{j=1}^{|\mathcal{DS}_O|} M_{\mathcal{DS}_R \to \mathcal{DS}_O}(\hat{\mathbf{d}}_i, \mathbf{d}_j) M_{\mathcal{DS}_O \to \mathcal{DS}_R}(\mathbf{d}_j, \hat{\mathbf{d}}_i)}{|\mathcal{DS}_R|} \tag{5}$$

M is defined as:

$$M_{\mathcal{X} \to \mathcal{Y}}(\hat{\mathbf{d}}_i, \mathbf{d}_j) = \begin{cases} 1, \text{if } \forall k \neq j \neq j' : \|\hat{\mathbf{d}}_i - \mathbf{d}_j\|_2 < \|\hat{\mathbf{d}}_i - \mathbf{d}_{j'}\|_2 < \\ \|\hat{\mathbf{d}}_i - \mathbf{d}_k\|_2 \text{ and } \dfrac{\|\hat{\mathbf{d}}_i - \mathbf{d}_j\|_2}{\|\hat{\mathbf{d}}_i - \mathbf{d}_{j'}\|_2} < 0.8 \\ 0, \text{ otherwise} \end{cases}$$

where $|\cdot|$ means cardinality and $\|\cdot\|_2$ is the Euclidean distance. Eq. (5) measures the fraction of descriptors extracted from the interpolated f-frame not finding a proper descriptor match at the original f-frame descriptor set; it is therefore a descriptor matching distortion. This fraction counts the proportion of descriptors not meeting the ratio test and symmetric matching criteria as expressed by the product $M_{\mathcal{X} \to \mathcal{Y}}(\hat{\mathbf{d}}_i, \mathbf{d}_j) M_{\mathcal{Y} \to \mathcal{X}}(\mathbf{d}_j, \hat{\mathbf{d}}_i)$.

2. *Iterative descriptor matching estimation within the joint Lagrangian optimization*: As the joint optimization process iterates over the sorted keypoint matches, each candidate keypoint match is evaluated regarding the descriptor matching distortion D_M. Naturally, at this stage, the reconstructed f-frame is no longer the interpolated f-frame but rather its improved version with the successively selected patches associated to the successively selected keypoint matches. The iterative descriptor matching distortion estimation proceeds as follows (central and right branches in Figure 2):

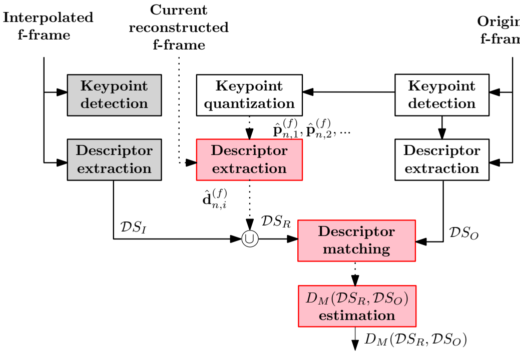

FIGURE 2. Encoder estimation of the descriptor matching distortion. The dashed arrows indicate the iterative steps.

a. *Keypoint parameters quantization*: Let $\mathbf{p}_{n,i}^{(f)} \mapsto \mathbf{p}_{m,j}^{(k)}$ be a given candidate keypoint match under consideration in the Lagrangian optimization. The residues between parameters of the original keypoint $\mathbf{p}_{n,i}^{(f)}$ and the parameters of its matching keypoint $\mathbf{p}_{m,j}^{(k)}$ are scalar quantized as detailed in Section IV, giving rising to the quantized keypoint $\hat{\mathbf{p}}_{n,i}^{(f)}$ used for descriptor extraction.

b. *Descriptor extraction at the quantized keypoint*: Extract descriptor $\hat{\mathbf{d}}_{n,i}^{(f)} = \Psi(\hat{\mathbf{p}}_{n,i}^{(f)}|I_R^{(f)})$ from current reconstructed f-frame (already improved with all the previously selected keypoint matches) at the quantized keypoint position.

c. *Descriptor matching for original versus current reconstructed f-frames*: Add the just extracted descriptor $\hat{\mathbf{d}}_{n,i}^{(f)}$ to the set $\mathcal{DS}_R$ and match $\mathcal{DS}_R$ to the descriptor set $\mathcal{DS}_O$ already extracted from the original f-frame.

d. *Descriptor matching distortion estimation*: Estimate the descriptor matching distortion between $\mathcal{DS}_R$ and $\mathcal{DS}_O$ as defined in Eq. (5). Here the descriptors to be matched comprise $\mathcal{DS}_I$ together with those in $\mathcal{DS}_R$, corresponding to the keypoints selected for coding so far. This descriptor matching distortion measures the fraction of used descriptors which did not result into a positive match.

The descriptor matching distortion estimation is performed for each candidate keypoint match, and feeds the joint Lagrangian optimization process that selects the keypoint match based also on the visual quality distortion and the rate, as described below.

JOINT LAGRANGIAN OPTIMIZATION PROCESS

At this point, with the joint optimization framework and metrics properly defined, it is time to design the joint RD_VD_M optimization process to select the optimal keypoint matches.

To determine the final set of keypoint matches to be coded, $\mathcal{M}_{kp}$, an iterative procedure considering all the available keypoint matches is adopted as follows:

1. *Initialization*: Given the selected values for γ and λ, set $\mathcal{M}_{kp} = \{\emptyset\}$ and initialize the minimum Lagrangian cost function as $J_{min} = (D_{V,ini} + \gamma D_{M,ini}) + \lambda R(\mathcal{M}_{kp})$. Here $D_{V,ini}$, the initial visual quality distortion, is defined as the MSE between the initially interpolated f-frame and the original f-frame as, at the beginning, the reconstructed f-frame is equal to the interpolated f-frame. Also here, $D_{M,ini}$, the initial descriptor matching distortion, is defined as the descriptor matching distortion between the descriptors extracted from the interpolated f-frame, $\mathcal{DS}_I$, and from the original f-frame, $\mathcal{DS}_O$, since at this point $\mathcal{DS}_R = \mathcal{DS}_I$. Lastly, since $\mathcal{M}_{kp} = \{\emptyset\}$, the rate is naturally zero.

2. *Iterative joint Lagrangian cost reduction*: For each candidate keypoint match in the available sorted list, temporarily add it to the set of selected keypoint matches, $\mathcal{M}_{kp}$, and evaluate its effectiveness in reducing the Lagrangian cost computed up to the current point. To do so, the following steps must be performed:

a. *Rate computation*: Compute the accumulated rate for the current set of keypoint matches in the set $\mathcal{M}_{kp}$ as detailed in Eq. (4).

b. *Visual quality performance impact assessment*: To check the visual quality benefit of additionally selecting the current keypoint match, and thus its associated patch, perform the adaptive patch stitching as described in Section IV over the current reconstructed f-frame. Then compute the visual quality distortion D_V between the resulting stitched f-frame and the original f-frame.

c. *Descriptor matching performance impact assessment*: To check the descriptor matching benefit of additionally selecting the current keypoint match, temporarily add the corresponding candidate descriptor to the reconstructed f-frame selected descriptor set $\mathcal{DS}_R$, and estimate the descriptor matching distortion D_M, between the set of descriptors for the current reconstructed frame and the original f-frame using Eq. (5).

d. *Lagrangian cost computation*: Using the rate, visual quality and descriptor matching distortions computed in steps a, b and c above, compute the Lagrangian cost $J = (D_V + \gamma D_M) + \lambda R(\mathcal{M}_{kp})$. If the Lagrangian cost is reduced relative to the current minimum Lagrangian cost, keep this candidate keypoint match in $\mathcal{M}_{kp}$ (and thus also its stitched patch in the updated reconstructed f-frame), keep its descriptor in $\mathcal{DS}_R$, and update the Lagrangian cost function minimum with this new minimum cost value. Otherwise, discard the keypoint match and its associated descriptor and process the next keypoint in the sorted list.

The above described keypoint matches selection procedure is able to consistently and jointly optimize the visualization and searching performances. The trade-off between visualization and searching distortions minimization depends on the specific application scenario, and can be set by appropriately tuning the Lagrangian cost parameters, λ and γ.

LAGRANGIAN COST PARAMETERS SELECTION

Naturally, the optimization control parameters γ and λ play a central role in the definition of the optimal configurations using the proposed video coding solution as different trade-offs between the optimization goals can be reached by adjusting them. In addition to γ and λ, another key control parameter is the quantization parameter (QP) value used to code the k-frames. In order to properly select the trio of parameters $(\mathrm{QP}, \gamma, \lambda)$ corresponding to various optimal operational points, extensive experiments have been performed as described next. In brief, the joint Lagrangian optimization process presented above was performed for multiple combinations of the parameters $(\mathrm{QP}, \gamma, \lambda)$, thus obtaining a dense cloud of RD_VD_M functional points. The $(\mathrm{QP}, \gamma, \lambda)$ parameter sets corresponding to RD_VD_M points lying on the convex hull of this dense cloud are selected as providing the best parameter choices. More specifically, this parameter selection process proceeds as follows:

1. RD_VD_M *space filling*: Run the coding solution for multiple combinations of the parameter set $(\mathrm{QP}, \gamma, \lambda)$ in some adopted dynamic range for each parameter. Let $config_w = (R, D_V, D_M, \mathrm{QP}, \gamma, \lambda)_w$ be each individual configuration vector including the resulting rate, visual quality distortion and descriptor matching distortion for a particular choice of the control parameter set $(\mathrm{QP}, \gamma, \lambda)_w$ and the parameter set itself. Let $CONFIG$ be the full set of such $config_w$ configuration vectors.

2. RD_VD_M *convex hull creation*: To find the set of RD_VD_M points from $CONFIG$ lying on the convex hull, the widely used convex hull algorithm Quickhull [30] has been used. It gives as output the facets of the convex envelope, that is, the smallest convex set of RD_VD_M points involving the input set of points; further details may be found in [30]. As the objective is here to find the parameter choices $(\mathrm{QP}, \gamma, \lambda)$ providing the optimal RD_VD_M tradeoffs, only the lowest facets are kept, this means, those facets which do not have any point below them.

Note that once the RD_VD_M points lying on the convex hull are obtained, the best set of parameter values $(\mathrm{QP}, \gamma, \lambda)$ are also obtained as every RD_VD_M tuple is associated to a particular $(\mathrm{QP}, \gamma, \lambda)$ tuple. An analytical relation for the parameters $(\mathrm{QP}, \gamma, \lambda)$ has been searched for by fitting a curve to the experimental results. However, the outcome was a function which was still very much content-dependent. For this reason, it was decided to obtain the $(\mathrm{QP}, \gamma, \lambda)$ values for optimal performance by using the convex hull obtained after exhaustive experiments. Figure 3 shows an example where the full cloud of RD_VD_M points is in red and the Delaunay triangulation for the RD_VD_M points lying on the convex hull are in blue for the video sequence Paris. In summary, this process where the RD_VD_M points on the convex hull are defined, allows to identify the $(\mathrm{QP}, \gamma, \lambda)$ combinations providing the optimal visualization-searching performances

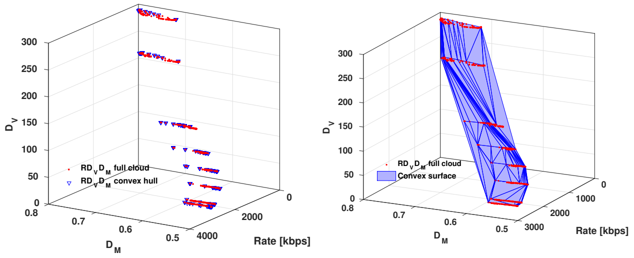

FIGURE 3. Left) Example of the $RD_V D_M$ points full cloud (red), the convex hull $RD_V D_M$ points are highlighted in blue; right) the corresponding convex surface for the sequence *Paris*.

trade-offs. In fact, a whole convex surface of optimal trade-offs can be found. Figure 3 shows an example of this convex surface obtained by performing Delaunay triangulation of the convex hull points.

SELECTED KEYPOINT MATCHES CODING

In the implemented instance of the proposed dual-purpose video coding framework, a set of keypoint matches is selected to code each f-frame at the encoder side. In terms of visualization, these selected keypoint matches indicate texture patches from reference frames which are worthwhile to be reused to improve regions in the interpolated f-frame most needing quality improvement. In terms of searching, the selected keypoint matches indicate image positions within the f-frame worth extracting feature descriptors as they are highly expressive is terms of searching.

In this context, to replicate the encoder patch stitching process and to indicate where to extract the descriptors at the decoder side, for each selected keypoint match, the following syntactic elements are coded: a) index of the reference frame in the reference frames buffer, r_k, e.g. previous or next for GOP size 2; b) index r_m of the matching keypoint in the reference frame available at the decoder considering the order given by the keypoint detector itself; c) encoder selected scale multiplicative factor, m_s; d) quantized keypoint parameters residues, notably the residues for the keypoint parameters position, angle and scale. Note that these parameters are residually coded using as reference the corresponding elements in the matching reference frame keypoint in order to exploit their inter-frame redundancy. In addition, to further reduce the rate, these residues are scalar quantized. This is detailed in the sequel.

1) Position residue and angle residue quantization

Both the position residue and the angle residue are quantized applying the same quantization scheme. For instance, the angle parameter residue for each matching keypoints pair, $c_\theta = \theta_{n,i}^{(f)} - \theta_{m,j}^{(k)}$, is quantized as:

$$\text{round}\left(\frac{\theta_{n,i}^{(f)} - \theta_{m,j}^{(k)}}{QS}\right) QS \tag{6}$$

where QS is the quantization step. Thus, the f-frame decoded keypoint angle parameter is given by:

$$\hat{\theta}_{n,i}^{(f)} = \theta_{m,j}^{(k)} + \text{round}\left(\frac{\theta_{n,i}^{(f)} - \theta_{m,j}^{(k)}}{QS}\right) QS \tag{7}$$

where $\theta_{m,j}^{(k)}$ for the reference frame is not quantized as it is extracted at the decoder.

The same procedure is carried out for the position residue quantization. In this work, the quantization is carried out using a common (fixed) quantization step, $QS = 0.25$, which has been validated with exhaustive experimentation.

2) Scale residue quantization

As for the scale quantization, there is one more step as the scale parameter depends on the integer parameters octave (o) and layer (l) according to:

$$\sigma = \sigma_0 \cdot 2^{(o + \frac{l}{3} + \Delta\sigma)} \tag{8}$$

where $\sigma_0 = 1.6$ and $\Delta\sigma$ is the scale offset, which resulted from the SIFT scale refinement [4]. Such parameters are also required for proper descriptor extraction at the decoder side. A differential scheme is used to code the octave and layer with respect to the octave and layer of the matching keypoint in the reference frame. No quantization is applied to the octave and layer residues in order to enable a proper descriptor extraction at the decoder side. Then, the scale residue between the "true" scale value and its approximation computed using the octave and layer values is quantized as follows:

$$\text{round}\left(\frac{\sigma_0\left(2^{(o+\frac{l}{3}+\Delta\sigma)} - 2^{(o+\frac{l}{3})}\right)}{QS}\right) QS \tag{9}$$

The same $QS = 0.25$ as above is used for scale residue quantization.

ENTROPY CODING

For better compression efficiency, the syntactic elements r_k, r_m and m_s are coded using arithmetic coding [31] with adaptive probability models, initialized with uniform probabilities. On the other hand, the keypoint parameter residues are coded using adaptive arithmetic coding with an initial statistical model set up for each parameter. The initial statistical models are obtained from a set of training sequences different from the set of test sequences. One can roughly estimate the rate associated to each syntactic element by considering a coding set up using 2 reference frames, a maximum of 256 keypoints per frame, 16 multiplicative scale factor values, CIF resolution and a maximum scale residue value of 80. In the worst case, using a uniform probability model for each syntactic element, the coding of each keypoint match would require 1 bit for r_k, 8 bits for r_m, 4 bits for m_s, 23 bits for the position residue, 12 bits for the angle parameter residue, 9 bits for the scale parameter residue, and 5 bits for the octave and layer, in a total of 62 bits per keypoint match. As it is proposed to use entropy coding with adaptive probability models, this rate can be reduced 1.6 times approximately.

ENHANCEMENT LAYER RESIDUE CODING

To improve the quality of the base layer (BL) reconstructed f-frames with its own novelty (and not only that migrating from the reference frames), a residue is computed between the original f-frame and the corresponding reconstructed BL f-frame. This residue is coded with an HEVC-like coding solution where the reconstructed BL f-frame plays the role of the HEVC prediction and the HEVC transform, quantization and entropy coding tools are used to code the enhancement layer (EL) residue. The adopted HEVC-like solution was built upon the HEVC reference software (HM ver. 16.3) [1] [2] [32].

Conceptually, this HEVC-like residue coding process consists in substituting the HEVC prediction module with the proposed BL decoder, which creates its prediction by using the keypoint matches (which behave like motion estimation) and patch stitching on top of an initially interpolated f-frame. The EL residue coding process includes the following steps:

1. *Enhancement layer prediction*: In the HEVC encoder, the residue for each coding unit (CU) is obtained after defining one or more prediction units (PUs) and subtracting the (Intra or Inter) predicted blocks from the original block. Similarly, in the proposed HEVC-like EL residue coding, the residue for each CU is obtained by subtracting the reconstructed BL output block from the original block.

2. *Base layer rate allocation map creation*: The rate associated to the prediction creation is here the rate used to code all the syntactic elements associated to the coding of the f-frame BL using the matching keypoint pairs. To perform the HEVC-like residue coding at CU level, it is necessary to compute its corresponding rate, in this case its rate share of the f-frame BL. To compute this rate, the BL produces a rate allocation map with an estimation of the BL rate expenditure for each area of the f-frame, notably depending on how the stitching process is distributed within the f-frame. More precisely, the number of bits spent for coding each keypoint match is divided by the number of pixels in the corresponding stitched area. Figure 4 shows an example of such rate allocation map where the whiter the area, the higher the rate estimation.

3. *Coding unit prediction rate estimation*: The rate allocation map is used by the HEVC-like residue coding module to estimate the rate already spent in the BL prediction. This estimation is needed in order each CU considers in its EL RD optimization the rate already spent by the BL in the corresponding image area. Otherwise, the EL residue coding process would be performed without taking into account the BL rate and thus its impact in reaching the final quality. In this case, the initial rate for each CU is estimated as the rate previously spent for the corresponding area in the BL rate allocation map as shown in Figure 4. In practice, for each CU, the proposed HEVC-like coding takes as prediction creation rate the rate corresponding to the stitching process parameters for the BL, this is done by summing the values within the rate allocation map overlapping the EL CU area (see Figure 4 right).

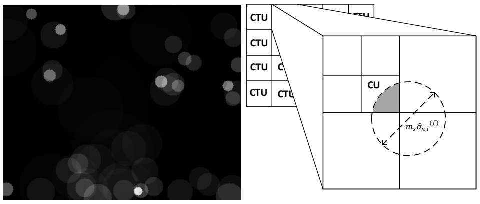

FIGURE 4. Left) Rate allocation map for the f-frame BL (frame 73 from the *Foreman* test sequence); right) Overlapping example between a coding unit and a stitched BL region (grey area).

4. *Residue coding*: After the prediction and rate estimation steps, the residual block coding occurs as in the HEVC encoder, notably involving the transform, quantization and entropy coding steps.

V. PERFORMANCE ASSESSMENT

This section presents extensive experimental results[1] for the assessment of the proposed DPVC solution under meaningful test conditions. The main objective here is to provide solid evidence that the proposed dual-purpose video coding framework shows already a promising performance with the proposed tools configuration regarding the very challenging HEVC standard. Naturally, more efficient coding solutions may be developed in the future within this same proposed framework. The results show the flexibility of the proposed coding solution to achieve different optimization trade-offs, notably when allocating the bitrate budget while jointly targeting visualization and searching capabilities. The state-of-the-art HEVC standard will be used as the natural benchmark to compare the obtained performance, considering not only joint optimization objectives but also special cases where the optimization is biased towards visualization or searching.

A. TEST MATERIAL AND CONDITIONS

To appropriately assess the proposed DPVC solution in terms of visualization and searching performances, the following materials and test conditions have been adopted:

- Six surveillance and personal communications video sequences have been selected, notably Hall, Container, Paris, Akiyo, FourPeople and KristenAndSara. The first four sequences are in CIF spatial resolution at 30Hz and 10 seconds long (300 frames); the last two sequences are in HD spatial resolution at 60Hz and 10 seconds long (600 frames). While it would be practically possible to provide results for datasets specifically targeted at search/retrieval such as MAR and MPEG-CDVS datasets, they would be highly questionable as those contents are either already compressed or consist of single frames from video sequences; unfortunately, this does

[1] Additional material may be found at [37].

not make a strong case for using those datasets for reliable coding experiments.

- To measure both the visualization and searching performances in a reliable way, each test sequence has been divided in two halves. To assess the visualization performance, the first half was used for coding. To assess the searching performance, the original version of the last frame of the second half (thus overall minimizing the correlation with the first half coded frames) was used to build the target content database at the decoder side; these frames play the role of target content for the queries based on decoded video frames.
- The selected QP values for k-frames coding were 45, 40, 37, 34, 30 and 25. The QP for f-frame residue coding is equal to the QP for k-frames incremented by 1 to implement some amount of quantization cascading. The γ values were set in the range 0 and 1000 and the λ values in the range 0 and 1 to accommodate the different distortion scales.
- A maximum number of 256 SIFT descriptors [4] was extracted per frame; for each keypoint, the residue of the parameters position and scale are quantized with precision of one quarter of pixel and the angle is quantized with precision of one quarter of degree.
- GOP sizes 2 and 4 are considered to assess the impact of the GOP size; low GOP sizes are used as the addressed application scenarios critically require low-delay. The reference frames buffer always includes two reference frames (one past and one future); for GOP size 2, these reference frames are the past and future k-frames for each f-frame. A hierarchical GOP scheme, where the f-frames also get patches from other f-frames organized in a dyadic way, is adopted for GOP sizes larger than 2.

B. BENCHMARKS AND METRICS

The natural benchmark for the proposed coding solution is the state-of-the-art HEVC standard, notably its reference software HM, version 16.3 [33]. The Main profile has been selected while using two prediction structures: All Intra and IBI (alternating Intra-predicted and Bi-predicted frames). It is important to stress that this is a very tough benchmark as it represents the best result of the video coding technology evolution designed by the related research community over the recent few decades. Comparing a new, by definition more immature, coding solution with such mature benchmark is by itself a challenge, especially because the new coding solution has a dual target, notably visualization and searching. The related HATC work [16] unfortunately does not provide a bit allocation strategy flexible enough to achieve arbitrary compromises between the visualization and searching performances, making the two solutions not amenable to a fair comparison. To perform a solid, wide and meaningful evaluation, the following performance metrics have been adopted:

- **Keypoint extraction performance** – The repeatability score [6] between the keypoints detected in the original f-frames and those extracted from the decoded f-frames is used to evaluate the impact of compression on the quality of the keypoints positioning. This is so because the compression artifacts impact the extracted descriptors. The repeatability score is defined as the ratio between the number of keypoint correspondences (those with overlap error below 0.4 [5]) in the two images and the smallest number of detected keypoints in the two images and is averaged over all f-frames.
- **Visual quality distortion** – The MSE and the SSIM are adopted as the visual quality distortion metrics to evaluate the performance regarding visualization. Furthermore, the Bjontegaard-Delta metrics [34], notably the BD-Rate is used to compare alternative coding solutions in terms of RD performance, that is, rate reduction for equivalent quality. The visual quality distortion assessment considers all coded frames, both f-frames and k-frames, as these frames types are not independently coded from each other.
- **Descriptor matching distortion** – The searching performance is evaluated by the average of the fraction of descriptors extracted from the decoded video (query descriptors) which positively match the descriptors extracted for each image in the target content database. These positive descriptor matches must satisfy both the ratio test and the symmetric matching criteria to be declared proper, positive matches. The descriptor matching distortion, as defined in Eq. (5), is the complementary fraction of the descriptor matching performance. Notice that here the ‘true’ descriptor matching performance is computed (and not an estimation), which may only be assessed at the decoder side with access to the (original) content database. The descriptor matching distortion assessment considers only the f-frames (against the HEVC B frames) as there are no keypoints coded for the k-frames.

C. KEYPOINT REPEATABILITY PERFORMANCE

Repeatability is a fundamental property for visual features. Matching performance based on visual features relies on the property of detecting the same distinguishing locations on images depicting the same scene content, although acquired or processed differently. Notably, image and video compression has a detrimental effect on keypoint detection, mainly at lower bitrates where a large quantization step and blocking artifacts may create spurious keypoint responses and erase valuable ones. This in turn would imply extracting descriptors at image locations unlikely to be correctly matched with descriptors extracted from original images. The first main advantage of the proposed DPVC solution is the availability at the decoder of originally extracted keypoint locations what is not possible for the alternative HEVC solution. Figure 5 shows the repeatability score averaged over all f-frames for DPVC and over all B-frames for a HEVC IBI configuration. The proposed DPVC consistently achieves a repeatability score of 100%, meaning that the keypoint locations are essentially the same as obtained from original frames (despite the light quantization applied to the keypoint parameter residues). This is essentially different from the HEVC repeatability behavior as the compression process has

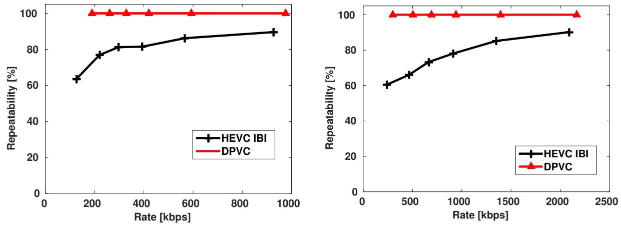

FIGURE 5. Repeatability score averaged over all f-frames/B-frames for the sequences: *Hall* and *Paris*.

a significant detrimental effect on the keypoint locations, especially at the lower bitrates where the repeatability score drops significantly.

D. TRADING-OFF VISUALIZATION AND SEARCHING PERFORMANCES

To show the flexibility of the proposed DPVC solution in trading-off visualization (visual quality distortion) and searching performances (descriptor matching distortion) while offering comparable performances regarding HEVC, this section presents and compares RD_V curves for a fixed descriptor matching distortion (D_M). These (level) curves are obtained from the convex surface fitted to the convex hull points as described in Section IV. While HEVC offers a fixed descriptor matching distortion D_M for a specific RD_V pair, the DPVC solution offers the same D_M performance for all the RD_V pairs lying along a curve implying that it is possible to trade-off rate with visual quality without 'touching' the descriptor matching performance. This is a powerful capability that only the proposed coding solution can offer, which results from the proposed joint optimization strategy.

In Figure 6, the RD_V performance for the proposed DPVC solution is presented for the six video sequences at a specific D_M value (set by HEVC IBI and IBBBI to allow a fair comparison). As shown, from an RD_V performance perspective, the proposed DPVC solution performs rather similarly to HEVC (both for the GOP 2 (IBI) and GOP 4 (IBBBI) cases) at the fixed descriptor matching distortion values while offering at the same time many other RD_V combinations for the same matching distortion. The key issue here is that the DPVC solutions offers a large set of RD_V operational points for each matching distortion, what is impossible with HEVC. For example, DPVC is able to offer a reasonable increase or reduction in the visual quality distortion by reducing or increasing the bitrate expenditure while keeping fixed the descriptor matching distortion. This behavior evidences that the jointly selected and coded keypoints are effective in holding the descriptor matching distortion at a certain level while trading-off the visual quality distortion. As consequence of the proposed flexible $RD_V D_M$ joint optimization, it is possible to appropriately set the pair (λ, γ), which are the parameters weighting the video distortion versus the descriptor matching distortion, to control the number of coded keypoints, in this way controlling the descriptor matching distortion.

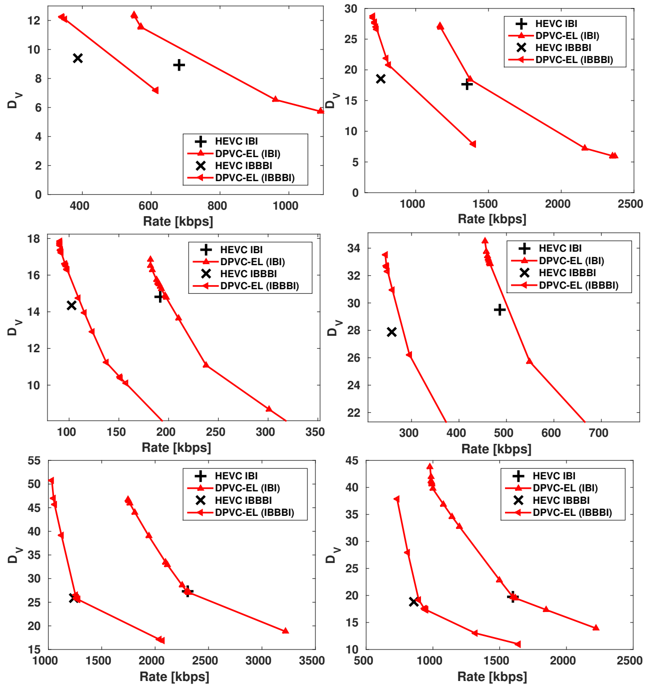

FIGURE 6. RD_V performance for a fixed descriptor matching performance (GOP size 2 and 4) for sequences: top) *Hall* and *Paris* ; middle) *Akiyo* and *Container*; and bottom) *FourPeople* and *KristenAndSara*. The D_M values for each sequence are given below for both GOP sizes.

	Hall	*Paris*	*Akiyo*	*Container*	*Four People*	*Kristen AndSara*
IBI	0.57	0.63	0.67	0.67	0.65	0.61
IBBBI	0.56	0.57	0.63	0.66	0.64	0.59

As the GOP size gets longer, the number of f-frames increases over the number of k-frames. This could hurt the overall compression efficiency as it would limit the power of the reference frames to provide reusable patches, in turn, leading the encoder to struggle to balance visualization and searching performance. The unified optimization framework combined with the use of a hierarchical GOP structure effectively enable the encoder to balance visualization and searching performance for GOPs longer than 2 while being less bitrate-hungry than the GOP 2 setup.

E. BEST SEARCHING PERFORMANCE

Among the optimization trade-offs achievable with the proposed DPVC solution is the special case where the operational points are selected to provide the best searching performance. In this jointly optimized video coding framework, this situation is associated to the convex hull points which yield the best trade-off between searching performance and rate. Figure 7 shows RD_M curves expressing the best descriptor matching distortion from the $RD_V D_M$ points on the convex hull. For both the IBI and IBBBI cases, the DPVC solution consistently outperforms or performs equivalently to the

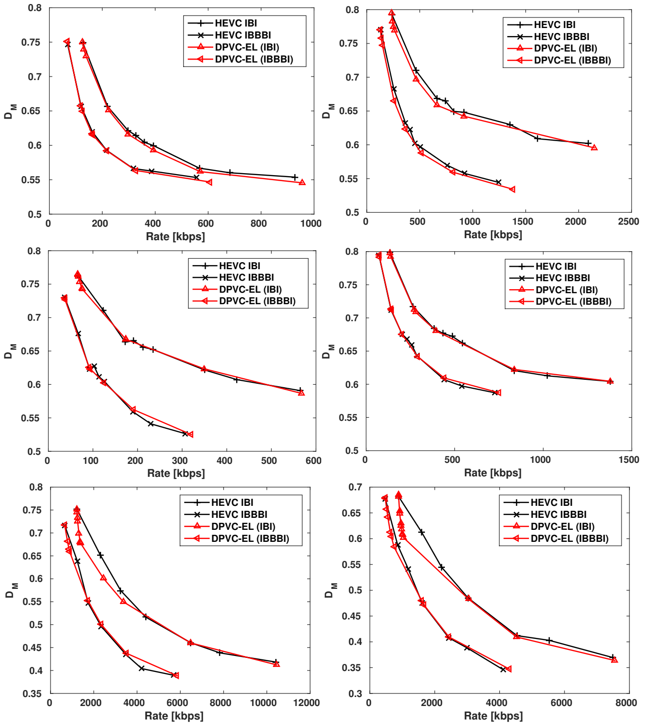

FIGURE 7. Best operational points in terms of RD_M performance (GOP size 2 and 4) obtained from the convex hull points for sequences: top) *Hall* and *Paris*; middle) *Akiyo* and *Container*; and bottom) *FourPeople* and *KristenAndSara*.

HEVC RD_M performance as it achieves a lower or close descriptor matching distortion than HEVC for the same bitrate. This again validates the importance of providing reliable keypoint location and consequently descriptor information for improved searching performance.

F. BEST VISUALIZATION PERFORMANCE

The other case of special relevance is when the optimization goal has to achieve the best visualization performance; this allows to assess to what extent the proposed coding solution is competitive with the best standard video coding solution available in terms of the most commonly used RD_V performance. To this end, the best operational points regarding visualization performance are selected from the convex hull points for the proposed DPVC.

Tables 1 and 2 present BD-Rate values for the proposed DPVC solution regarding HEVC, using the PSNR as visual quality metric instead of D_V. Table 1 provides results for GOP size 2 regarding the HEVC IBI and HEVC All Intra benchmarks for DPVC-BL and DPVC-EL, which correspond to the DPVC base and enhancement layers, and also for the so-called Motion-Compensated Frame Interpolation (MCFI) solution where f-frames only result from frame interpolation at no rate cost. This set of results allows concluding that DPVC-EL performs very close to HEVC IBI and easily outperforms HEVC All Intra, thus showing that the use of keypoint matches, which behave like motion vectors, in combination with patch stitching and residue coding do not introduce significant compression performance losses regarding the video coding state-of-the-art as represented by HEVC for GOP size 2. A slightly larger performance loss is observed for GOP size 4 as shown in Table 2. This is explained by the increase of the relative distances between f-frames and their reference frames, thus implying less temporal correlation. This effect tends to impair the BL compression performance which largely relies on the texture that can be reused from the reference frames. The use of a hierarchical GOP structure mitigates this BL performance loss as closer reference frames (in this case also f-frames) are used. As a compensation, the proposed solution offers, in a unified fashion, an explicit and flexible coding framework where visualization and searching can be jointly optimized, while still offering good performance for the cases where one optimization target dominates the other. It is interesting to notice that the EL (residue coding by a HEVC-like coding solution) is much more efficient for the HD sequences, this is because its coding tools were designed focusing on high spatial resolution video signals.

It is important to stress that the obtained BD-Rate loss is typically below 2.5% for GOP size 2 and 10% for GOP size 4 while offering some amount of quality scalability. When scalability is offered, it is common to accept a BD-Rate penalty up to 10% regarding the non-scalable solution [36]. It can observed that the penalty is much lower here. Notice that HEVC plays here the role of the non-scalable solution as it does not offer any quality scalability. These losses refer anyway to the extreme case where only the visualization performance is optimized while the proposed coding solution genuinely targets the case where visualization versus searching tradeoffs are needed. Moreover, the SSIM RD performance is shown in Figure 8 both for DPVC and HEVC. The results for this perceptual-inspired quality index corroborate the PSNR-based RD performance conclusion both for GOP sizes 2 and 4.

VI. FINAL REMARKS

In modern video applications, the role of the decoded video is much more than filling a screen for visualization. Among the emerging required user capabilities, searching plays a key role. In this context, this paper proposes a novel dual-purpose video coding framework that targets not only the usual visualization capabilities but it also potentiates simpler and better searching capabilities by combining the pixel-based and feature-based coding approaches. To this end, a flexible and unified Lagrangian optimization framework has been designed, which explicitly takes into account the rate and the visual quality and descriptor matching distortions. The experimental results show that the proposed framework allows to reach multiple trade-off points in terms of visualization and searching performances with no or only a negligible RD performance penalty. The results for the proposed implementation of the dual-purpose video framework provide solid evidence that it can be competitive with the

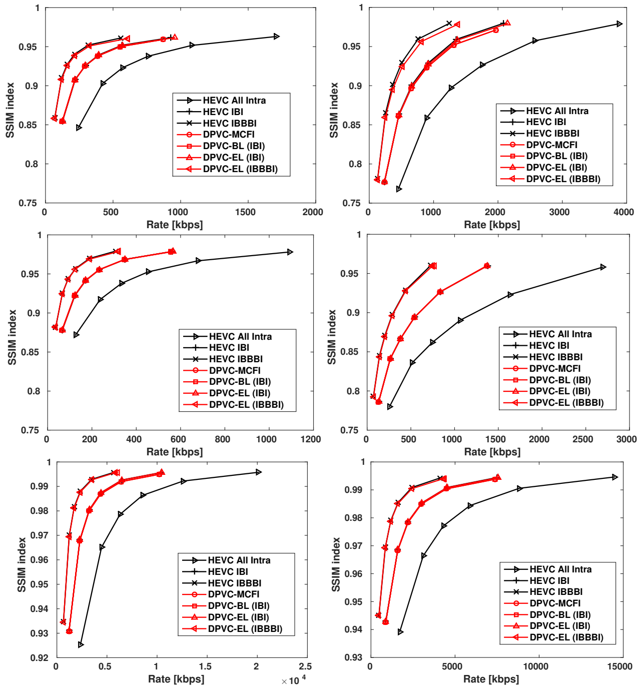

FIGURE 8. SSIM performance (GOP sizes 2 and 4) for sequences: top) *Hall* and *Paris*; middle) *Akiyo* and *Container*; and bottom) *FourPeople* and *KristenAndSara*.

TABLE 1. PSNR BD-Rate for DPVC regarding HEVC for GOP size 2

		HEVC All Intra	HEVC IBI
Hall	**DPVC-MCFI**	-42.406%	8.871%
	DPVC-BL	-42.796%	8.188%
	DPVC-EL	-46.980%	**1.560%**
Paris	**DPVC-MCFI**	-40.843%	15.630%
	DPVC-BL	-41.031%	15.243%
	DPVC-EL	-47.599%	**2.416%**
Akiyo	**DPVC-MCFI**	-50.971%	0.948%
	DPVC-BL	-51.063%	0.738%
	DPVC-EL	-50.955%	**0.817%**
Container	**DPVC-MCFI**	-51.769%	1.160%
	DPVC-BL	-51.829%	1.040%
	DPVC-EL	-51.916%	**0.789%**
Aver. BD-Rate		**-49.362%**	**1.395%**
Fourpeople	**DPVC-MCFI**	-49.077%	3.782%
	DPVC-BL	-49.119%	3.696%
	DPVC-EL	-50.951%	**0.022%**
KristenAndSara	**DPVC-MCFI**	-48.017%	5.666%
	DPVC-BL	-48.149%	5.411%
	DPVC-EL	-50.605%	**0.512%**
Aver. BD-Rate		**-50.778%**	**0.267%**

TABLE 2. PSNR BD-Rate for DPVC regarding HEVC for GOP size 4

		HEVC All Intra	HEVC IBBBI
Hall	**DPVC-EL**	-68.812%	6.408%
Paris	**DPVC-EL**	-68.353%	9.768%
Akiyo	**DPVC-EL**	-73.268%	4.264%
Container	**DPVC-EL**	-74.300%	5.819%
Aver. BD-Rate		**-71.183%**	**6.564%**
FourPeople	**DPVC-EL**	-73.637%	2.717%
KristenAndSara	**DPVC-EL**	-73.241%	3.819%
Aver. BD-Rate		**-73.439%**	**3.269%**

state-of-the-art HEVC solution if required, while offering increased operational flexibility regarding the compromise between visualization and searching capabilities. Naturally, more efficient coding solutions may be developed in the future within this same proposed framework. The proposed framework opens new avenues for the development of more efficient coding solutions in the future that are also good for searching applications. Future work will consider the design of a video coding framework where the f-frames are efficiently coded using the descriptors themselves and not only the keypoint matches. This should allow performing searching not only using original data extracted keypoints but also using original data extracted descriptors.

REFERENCES

[1] G. J. Sullivan, J.-R. Ohm and W.-L. Han, "Overview of the High Efficiency Video Coding (HEVC) Standard," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 11, no. 12, pp. 1649-1666, December, 2012.

[2] F. Bossen, B. Bross and K. Suhring, "HEVC Complexity and Implementation Analysis," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1685-1696, December, 2012.

[3] B. Girod, C. Vijay, D. Chen, N.-M. Cheung, R. Grzeszczuk, Y. Resnik, G. Takacs, S. Tsai and R. Vedantham, "Mobile Visual Search," in IEEE Signal Processing Magazine, vol. 28, no. 4, pp. 61-76, June, 2011.

[4] D. Lowe, "Distinctive Image Features from Scale-invariant Keypoints," in International Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, November, 2004.

[5] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir and L. Van Gool, "A Comparison of Affine Region Detectors," in International Journal of Computer Vision, vol. 65, no. 1-2, pp. 43-72, October, 2006.

[6] L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi and S. Tubaro, "Coding Visual Features Extracted from Video Sequences," in IEEE Transactions on Image Processing, vol. 23, no. 5, pp. 2262-2276, March, 2014.

[7] S. D. Chen and P. Moulin, "A Two-part Predictive Coder for Multitask Signal Compression," in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Florence, Italy, 2014.

[8] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, "Speeded-Up Robust Features (SURF)," in Computer Vision and Image Understanding, vol. 110, no. 3, pp. 346-359, June, 2008.

[9] I. JTC, "Information Technology - Multimedia Content Description Interface - Part 13: Compact Descriptors for Visual Search," 2013.

[10] L.-Y. Duan, V. Chandrasekhar, J. Chen, J. Lin, Z. Wang, T. Huang, B. Girod and W. Gao, "Overview of the MPEG-CDVS Standard," in IEEE Transactions on Image Processing, vol. 25, no. 1, pp. 179-194, January, 2016.

[11] J. Chao and E. Steinbach, "Keypoint Encoding for Improved Feature Extraction from Compressed Video at Low Bitrates," in IEEE Transactions on Image Processing, vol. 18, no. 1, pp. 25-39, January, 2016.

[12] L. Baroffio, M. Cesana, A. Redondi, M. Tagliasacchi and S. Tubaro, "Hybrid Coding of Visual Content and Local Image Features," in IEEE International Conference on Image Processing (ICIP), Quebec City, Canada, 2015.

[13] J. Chao and E. Steinbach, "Preserving SIFT Features in JPEG-encoded Images," in IEEE International Conference on Image Processing (ICIP), Brussels, Belgium, 2011.

[14] M. Makar, H. Lakshman and V. Chandrasekhar, “Gradient Preserving Quantization,” in IEEE International Conference on Image Processing (ICIP), Orlando, USA, 2012.

[15] J. Chao and E. Steinbach, “SIFT Feature-preserving Bit Allocation for H.264/AVC Video Compression,” in IEEE International Conference on Image Processing (ICIP), Orlando, USA, 2012.

[16] X. Zhang and S. Ma and S. Wang and X. Zhang and H. Sun and W. Gao, “A Joint Compression Scheme of Video Feature Descriptors and Visual Content,” in IEEE Transactions on Image Processing, vol. 26, no. 2, pp. 633-647, February, 2017.

[17] A. Redondi, M. Cesana and M. Tagliasacchi, “Low Bitrate Coding Schemes for Local Image Descriptors,” in IEEE International Workshop on Multimedia Signal Processing (MMSP), Banff, Canada, 2012.

[18] V. Chandrasekhar, G. Takacs, D. Chen, S. S. Tsai, J. Singh and B. Girod, “Transform Coding of Image Feature Descriptors,” in IEEE Visual Communications and Image Processing, San Jose, USA, 2009.

[19] V. Chandrasekhar, G. Takacs, D. M. Chen, S. Tsai, Y. Reznik, R. Grzeszezuk and B. Girod, “Compressed Histogram of Gradients: A Low-Bitrate Descriptor,” in International Journal of Computer Vision, vol. 96, no. 3, pp. 384-399, February, 2012.

[20] J. Ascenso and F. Pereira, “Lossless Compression of Binary Image Descriptors for Visual Sensor Networks,” in International Conference on Digital Signal Processing (DSP), Fira, Greece, 2013.

[21] H. Yue and X. Sun, F. Wu and J. Yang, “SIFT-based Image Compression,” in IEEE International Conference on Multimedia and Expo (ICME), Melbourne, Australia, 2012.

[22] H. Yue, X. Sun, J. Yang and F. Wu, “Cloud-based Image Coding for Mobile Devices - Toward Thousands to One Compression,” in IEEE Transactions on Multimedia, vol. 15, no. 4, pp. 845-857, January, 2013.

[23] P. Weinzaepfel, H. Jégou and P. Pérez, “Reconstructing an Image from its Local Descriptors,” in IEEE CVPR, Colorado Springs, USA, 2011.

[24] S. Milani, G. Agresti and G. Calvagno, “A Rate Control Algorithm for Video Coding in Augmented Reality Applications,” in PCS, Nuremberg, Germany, 2016.

[25] R. C. da Silva, F. Pereira and E. A. B. da Silva, “Feature-based Video Coding: Designing a RD Efficient and Search Friendly Framework,” in PCS, Nuremberg, Germany, 2016.

[26] P. Pérez, M. Gangnet and A. Blake, “Poisson Image Editing,” in ACM T-G vol. 22, no. 3, pp. 313-318, July, 2003.

[27] A. Ortega and K. Ramchandran, “Rate-Distortion Methods for Image and Video Compression,” in IEEE Signal Processing Magazine vol. 15, no. 6, pp. 23-50, November, 1998.

[28] G. J. Sullivan and T. Wiegand, “Rate-Distortion Optimization for Video Compression,” in IEEE Signal Processing Magazine vol. 15, no. 6, pp. 74-90, November, 1998.

[29] H. Everett, “Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources,” in Operations Research vol. 11, no. 3, pp. 399-417, May, 1963.

[30] C. B. Barber, “The Quickhull Algorithm for Convex Hulls,” in ACM Transactions on Mathematical Software, vol. 22, no. 4, pp. 469-483, December, 1996.

[31] T. C. Bell, J. G. Cleary and I. H. Witten, “Text Compression,” Englewood Cliffs, New Jersey: Prentice-Hall, 1990.

[32] I.-K. Kim, J. Min, T. Lee, W.-J. Han and J. Park, “Block Partitioning Structure in the HEVC Standard,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1697-1706, December, 2012.

[33] JCTVC, “HEVC Test Model (HM),” [On-line]. Available: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/. Accessed on: October 18, 2017.

[34] G. Bjontegaard, “Calculation of Average PSNR differences between RD-curves,” in ITU-Telecommunications Standardization Sector, March, 2001.

[35] J. Ascenso and C. Brites and F. Pereira, “Improving Frame Interpolation with Spatial Motion Smoothing for Pixel Domain Distributed Video Coding,” in EURASIP CSIPMCS, Smolenice, Slovak Republic, 2005.

[36] H. Schwarz and D. Marpe and T. Wiegand, “Overview of the Scalable Video Coding Extension of the H.264/AVC Standard,” in IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 9, pp. 1103-1120, September, 2007.

[37] Renam Silva, “Additional Material,” [Online]. Available: http://www02.smt.ufrj.br/ renam.silva/tmm_additional_material.html. Accessed on: April 1, 2019.

[38] T. C. Meyer, “Matrix Analysis and Applied Linear Algebra,” Philadelphia, USA: Society for Industrial and Applied Mathematics, 2000.

RENAM C. DA SILVA (SM’2015–M’2018) received the Mechatronics Engineering degree from Universidade do Estado do Amazonas, Brazil, in 2010, M.Sc. and D.Sc. degrees in Electrical Engineering from Universidade Federal do Rio de Janeiro in 2013 and 2018, respectively. He is currently with the Samsung R&D Institute Brazil. His research interests include multimedia signal processing, image and video compression, and deep learning. He has served as a reviewer for IEEE Transactions on Multimedia.

FERNANDO PEREIRA (S’88–M’90–SM’99–F’08) received the B.S., M.Sc., and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Universidade de Lisboa, Lisboa, Portugal, in 1985, 1988, and 1991, respectively. He is currently with the Electrical and Computer Engineering Department, IST, and Instituto de Telecomunicações, where he is responsible for the participation of IST in many national and international research projects. He has authored or coauthored more than 250 papers. His areas of research interest include visual information analysis, processing, coding and description, and interactive multimedia services. He is or has been an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON MULTIMEDIA, and the IEEE Signal Processing Magazine, and the Editor-in-Chief of the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING. He is or has been a member of the IEEE Signal Processing Society Image Technical Committees on Video and Multidimensional Signal Processing and Multimedia Signal Processing Technical Committees, and of the IEEE Circuits and Systems Society Technical Committees on Visual Signal Processing and Communications and Multimedia Systems and Applications. He was an IEEE Distinguished Lecturer in 2005. He is an Area Editor of Signal Processing: Image Communication. He has been a member of the Scientific and Program Committees of many international conferences. He has been participating in the work of ISO/IEC MPEG and ISO/IEC JPEG for many years, notably as the Head of the Portuguese delegation, the Chairman of the MPEG and JPEG Requirements Subgroups.

EDUARDO A. B. DA SILVA (M’95–SM’05) was born in Rio de Janeiro, Brazil. He received the Electronics Engineering degree from Instituto Militar de Engenharia (IME), Brazil, in 1984, the M.Sc. degree in Electrical Engineering from Universidade Federal do Rio de Janeiro (COPPE/UFRJ) in 1990, and the Ph.D. degree in Electronics from the University of Essex, England, in 1995. He is a professor of Universidade Federal do Rio de Janeiro since 1989. He is co-author of the book "Digital Signal Processing - System Analysis and Design", published by Cambridge University Press, in 2002, that has also been translated to the Portuguese and Chinese languages, whose second edition has been published in 2010. He published more than 70 papers in international journals. His research interests lie in the fields of signal and image processing, signal compression, digital TV, 3D videos, computer vision, light fields and machine learning, together with its applications to telecommunications and the oil and gas industry. He is co-editor of the future standard ISO/IEC 21794-2, JPEG Pleno Plenoptic image coding system. Prof. Da Silva is a Senior Member of the Brazilian Telecommunications Society (SBrT) and of the IEEE.