

Artifact-free variational MPEG decomposition

Kristian Bredies and Martin Holler

Institute for Mathematics and Scientific Computing, University of Graz, Austria

Abstract. We propose a variational method for artifact-free video decomposition that is capable of processing *any* MPEG-2 encoded movie. The method extracts, from a given MPEG-2 file, a set of admissible image sequences and minimizes an artifact-penalizing spatio-temporal regularization functional over this set, giving an optimal decompressed image sequence. For regularization, we use the *infimal convolution* of spatio-temporal *Total Generalized Variation* functionals (ICTGV). Numerical experiments on MPEG encoded files show that our approach significantly increases image quality compared to standard decomposition.

Keywords: MPEG decomposition, image sequence regularization, total generalized variation, infimal convolution type functionals.

1 Introduction

The MPEG (Motion Picture Experts Groups) video compression standard is one of the most well-known and widely-used methods to compress and store digital video data. The underlying concepts of MPEG compression, namely motion compensation and Block Discrete Cosine Transform (BDCT) encoding, are at the heart of almost any modern video compression method. Currently, several versions of MPEG are available of which MPEG-2 and MPEG-4 are the most popular, with MPEG-2 being used for DVD videos and HDTV broadcasting. MPEG achieves high compression rates such that the storage and transmission of the huge amount of data usually required by videos becomes feasible. However, the high compression comes with the cost of loss of data due to quantization (rounding) of BDCT coefficients which is responsible for disturbing artifacts in the decompressed movie.

In this work, we propose a variational method for improved MPEG decomposition which is able to reduce compression-induced artifacts. By utilizing the information provided by the compressed MPEG file, our approach relies on the extraction of quantization intervals associated with the compressed data. These intervals must contain the “original data”, i.e., the data prior to compression, and are used to describe the convex set D of all possible reconstructions. We then variationally decompress the video by minimizing a spatio-temporal regularization functional \mathcal{R} subject to the data being contained in D . As the decoding process, i.e., the mapping which takes the stored coefficients to the spatio-temporal video data, constitutes a linear operator A , the variational problem can be phrased as

$$\min_{d \in D} \mathcal{R}(Ad). \quad (1)$$

This allows, to the best knowledge of the authors for the first time, to apply the very successful concept of variational regularization to genuine MPEG compressed data and, depending on the regularization functional \mathcal{R} , has the potential for a strong quality improvement for any MPEG compressed movie.

Our method conceptually works for all versions of MPEG compression, but is implemented for MPEG-2 compressed video data, as this version is still one of the standard encoding formats and can be seen as a realistic but still tractable blueprint for most video compression methods. For regularization, we use the *Infimal Convolution of Total Generalized Variation* (ICTGV) functionals as introduced in [8], but also compare to spatio-temporal *Total Generalized Variation* (TGV) regularization [5] as this turned out to be successful in artifact-reduced JPEG still image decompression which shares the concept of BDCT coding with MPEG, see [3, 2] and [15] for an early work on variational JPEG decompression.

For the numerical solution of (1) we use the primal dual algorithm of [6] and provide an implementation that covers all orders of spatio-temporal TGV regularization and infimal convolutions thereof. Our numerical experiments confirm that, using ICTGV regularization, we are able to significantly reduce compression artifacts in MPEG compressed videos and obtain a good reconstruction quality even at relatively high compression rates.

As already mentioned, we are not aware of any other work on variational MPEG decompression. In contrast to that, there are many filter based post-processing techniques for MPEG compressed video available and we refer to [9] for an overview. The field of image sequence regularization is also not yet as well investigated as its still image counterpart. For recent approaches in the direction of spatio-temporal regularization we refer to [12, 8]. In [12] the authors aim at reconstructing video which has been compressed using a particular video compressive sensing method and employ the total variation applied to the difference between adjacent frames as temporal regularization. In [8] the ICTGV functional, which is also used in the present work, is introduced in an analytical setting and employed for spatio-temporal regularization of MJPEG compressed grayscale video data, i.e., data where each frame undergoes simple JPEG encoding.

2 A Variational Model for MPEG decompression

The key observation for improving MPEG decompression within a variational framework is that the decoding of MPEG compressed video data can be seen as a linear operator mapping block cosine data to image space, and that bounds for the block cosine data can be derived from the compressed file. Hence the problem of artifact-free MPEG decompression can be cast in the form (1).

A short overview on MPEG-2 compression. The MPEG standard builds on the JPEG still image standard for which a short overview is given first. For JPEG encoding, the image first undergoes a discrete blockwise cosine transformation (BDCT) on blocks of size 8×8 pixels. This yields a representation of each block as a linear combination in terms of frequencies. Based on the assumption

that higher frequencies are less important for the perceived image quality, the cosine coefficients are then quantized according to a predefined quantization table (which is stored in the JPEG file) and rounded to integer. For color images, the JPEG standard encodes each channel of the image in YCbCr color space separately. Additionally, the chroma components, i.e., the Cb and Cr channel, may be subsampled (by a factor of 2 either in horizontal or both directions) accounting for the fact that the human visual system is less sensitive to spatial color variations.

For movies, MPEG compression now essentially incorporates additional prediction of neighboring frames by motion estimation and possible additive correction by BDCT data. Like JPEG, it codes 3 channels in YCbCr color space, allowing for different types of subsampling. The MPEG standard realizes three different types of frames, the I (intra), P (forward prediction) and B (bidirectional prediction) frames, respectively. An I frame codes the whole spatial image data independently from other frames in terms of quantized block DCT coefficients, similarly to the JPEG standard. For these frames, the quantization intervals for the BDCT data can be obtained from the MPEG file. Unlike I frames, a P frame refers to the spatial data of the previously decoded I or P frame. Such a frame is reconstructed macroblockwise (i.e. in blocks of size 16×16) by either decoding 8×8 quantized BDCT data for this macroblock or copying data from the previous I or P frame (motion compensation). For the latter, the position of the source block is determined by *motion vectors* which are coded in the MPEG file. Additionally, the outcome is possibly corrected by adding some BDCT-coded residual data whose quantization intervals can also be obtained. Finally, B frames may appear between two I or P frames and allow, in addition to P frames, to choose, for each macroblock, whether data should be copied from the previous or next I or P frame, or whether the data from both surrounding I or P frames should be averaged. Again, the corresponding motion vectors are obtained from the file.

Finally, in the MPEG data stream the coded frames are arranged in *groups of pictures* (GOPs) which usually contain exactly one I frame and some subsequent P or B frames. A typical GOP spans 12–15 frames and its structure could look like IBBPBBP...BBP, but also BBIBBP...BBP is possible. Consequently, one can consider each GOP as essentially independent and for the variational MPEG decompression, it suffices to optimize over one or two GOPs and not over the whole video.

Modelling related aspects. An MPEG compressed file provides, for each macroblock of each frame, BDCT data and/or motion vectors. Each macroblock is decoded by performing some of the following basic operations:

- Copying/Averaging of motion compensated macroblock data from other frames.
- Copying/Addition of inverse BDCT transformed 8×8 data blocks.

This gives a linear decompression operator which maps BDCT coefficients to the decoded image sequence. Standard decompression can be interpreted as applying

this operator to the integer coefficients given by the file. In our model we allow the BDCT coefficients, as these result from rounding, being contained in the intervals

$$J_i = [d_i^0 - 1/2, d_i^0 + 1/2] = [d_i^l, d_i^r], \quad (2)$$

where $d_i^0 \in \mathbb{Z}$ is the i th quantized BDCT coefficient and d_i^l, d_i^r abbreviate the interval bounds. With that, the set of possible source coefficients is then given by

$$D = \{(d_i)_i \mid d_i \in J_i\}. \quad (3)$$

The above-mentioned decompression operator, denoted by A , maps any element $d \in D$ to an image sequence (with possibly subsampled color components). To compensate for the color subsampling, we also introduce an upsampling operator \hat{S} that performs pixel repetition. The missing color detail information is modelled by \hat{s} , which is required to be in the kernel of the color subsampling operator S . MPEG decompression then amounts to finding an image sequence \hat{u} satisfying

$$\hat{u} = \hat{s} + \hat{S}A\hat{d} \quad (4)$$

with $\hat{s} \in \ker(S)$ and $\hat{d} \in D$ unknown.

Before this is incorporated in an optimization problem, we discuss two additional aspects concerning the data intervals $(J_i)_i$. First, as the BDCT coefficients of a single block are potentially used to construct image data for multiple frames, optimization over D would not take into account block matching errors, e.g., the same block might be copied to several positions at different frames. To compensate for that, we decouple the blocks by introducing new BDCT coefficients with the same data bounds each time a set of coefficients influences different blocks. Second, as the residual BDCT data of a macroblock in which motion compensation is performed is typically highly quantized, the resulting error intervals for the dequantized residual data are large, even though often only the DC coefficient is coded. We compensate for that by setting all error intervals describing residual data to be point intervals containing the center point. As this is just done for residual data, each block coefficient has still at least one degree of freedom that results from reproducing the data of other blocks.

The optimization problem. Now suppose that the space of discrete full resolution color movies is given by $U = \mathbb{R}^{n \times m \times T \times 3}$, where n, m are the horizontal and vertical number of pixels and T is the number of image frames. Recall that, as the I frames are encoded independent of the other frames, we can typically process one or two GOPs separately and hence T is not too large. Further define the space of coefficient data as $Z = \mathbb{R}^N$. The data set D is then given as

$$D = \{d \in Z \mid d_i \in J_i \text{ for all } i = 1, \dots, N\}, \quad (5)$$

with $(J_i)_i$ as described above. With \mathcal{R} a convex regularizer, artifact-free MPEG decompression can then be phrased as solving

$$\min_{s \in U, d \in Z} \mathcal{R}(s + \hat{S}Ad) + \mathcal{I}_D(d) + \mathcal{I}_{\ker(S)}(s) \quad (6)$$

where \mathcal{I}_L denotes the convex indicator function of a given set L . \mathcal{R} needs to be a suitable spatio-temporal regularization functional. Motivated by the results of [8], we use the *Infimal Convolution of Total Generalized Variation* type functionals as regularization. For comparison, we also employ spatio-temporal *Total Generalized Variation* (TGV) regularization. Before providing a definition of these functionals in the finite dimensional setting, we consider well posedness of the minimization problem (6) given in a general form.

In the infinite dimensional setting, all the above-mentioned functionals are lower semi-continuous semi-norms that are invariant on a finite dimensional space and coercive on its orthogonal complement (see [8, 4]). These results transfer to the finite dimensional setting, where any semi-norm is continuous, invariant on its (linear) kernel and coercive on the complement of its kernel. This allows the following general existence result, from which existence of a solution to (6) with \mathcal{R} being any semi-norm will immediately follow.

Proposition 1. *Let V be a finite dimensional vector space, $\mathcal{C} = \{v \in V \mid Mv \leq b\}$ be a convex polyhedron with $M : V \rightarrow \mathbb{R}^m$ linear, $m \in \mathbb{N}$, $b \in \mathbb{R}^m$, and \mathcal{H} a semi-norm on V . Then there exists a solution to*

$$\min_{v \in \mathcal{C}} \mathcal{H}(v). \quad (7)$$

Proof. Denote by $P : V \rightarrow \ker(\mathcal{H})^\perp$ the linear projection to the orthogonal complement of the kernel of \mathcal{H} . Then, since \mathcal{C} is a polyhedron, the set $P(\mathcal{C})$ is again a polyhedron, in particular closed. This can be deduced from linearity of P and the fact that, by the Minkowski-Weyl theorem (see [13, Corollary 7.1b]), any polyhedron can equivalently be written as the sum of the convex hull of finitely many vectors and a finitely generated convex cone.

As \mathcal{H} is coercive and continuous on the closed set $P(\mathcal{C})$, it admits a minimizer $v_{P(\mathcal{C})}$ over this set. Denoting $v^* \in \mathcal{C}$ such that $P(v^*) = v_{P(\mathcal{C})}$, it follows that v^* is a solution to (7). Indeed, for any $v \in \mathcal{C}$ we have

$$\mathcal{H}(v) = \mathcal{H}(P(v)) \geq \mathcal{H}(v_{P(\mathcal{C})}) = \mathcal{H}(v^*)$$

and hence the proof is complete.

Now as for the MPEG decomposition case all data intervals $(J_i)_i$ are closed, the set $\mathcal{C} := \ker(S) \times D \subset U \times Z$ can be written as convex polyhedron. Choosing \mathcal{R} to be any semi-norm, existence of a solution to (6) then follows from the above proposition with $\mathcal{H}(s, d) = \mathcal{R}(s + \hat{S}Ad)$. In particular (6) is well-posed when choosing either of the regularizers described on the following.

Spatio-temporal second order Total Generalized Variation. The Total Generalized Variation functional was introduced in [5] in the still image context as convex regularizer that is aware of higher order smoothness and hence avoids first order staircasing artifacts that result from TV regularization. As flickering artifacts that have been observed with spatio-temporal TV regularization [1] might be interpreted as temporal staircasing, there is hope that TGV regularization is superior also in the spatio-temporal context.

For $u = (u_{i,j,t}) \in U$ we define the forward finite difference operator

$$(\delta_1^+ u)_{i,j,t} = \begin{cases} u_{i+1,j,t} - u_{i,j,t} & \text{if } i = 0, \dots, n-2 \\ 0 & \text{else} \end{cases} \quad (8)$$

and δ_2^+ , δ_3^+ similarly. Note that $(\delta_1^+ u)_{i,j,t} \in \mathbb{R}^3$ as u has three color components. For $\lambda > 0$, a weighted spatio-temporal gradient $\nabla_\lambda : U \rightarrow U^3$ is given as

$$\nabla_\lambda u = (\delta_1^+ u, \delta_2^+ u, \lambda \delta_3^+ u). \quad (9)$$

The weight λ is introduced to fix the ratio between the temporal and spatial step-size for the finite difference operands which is not given a-priori. We further define the backward finite difference operator

$$(\delta_1^- u)_{i,j,t} = \begin{cases} u_{i,j,t} - u_{i-1,j,t} & \text{if } i = 1, \dots, n-1 \\ 0 & \text{else} \end{cases} \quad (10)$$

and δ_2^- , δ_3^- similarly. The symmetrized gradient \mathcal{E}_λ with temporal weighting is then defined for a vector field $v = (v^1, v^2, v^3) \in U^3$ as

$$\mathcal{E}_\lambda v = \frac{1}{2}(\nabla_{2,\lambda} v + (\nabla_{2,\lambda} v)^T), \quad \nabla_{2,\lambda} v = \begin{pmatrix} \delta_1^- v_1 & \delta_2^- v_1 & \lambda \delta_3^- v_1 \\ \delta_1^- v_2 & \delta_2^- v_2 & \lambda \delta_3^- v_2 \\ \delta_1^- v_3 & \delta_2^- v_3 & \lambda \delta_3^- v_3 \end{pmatrix}. \quad (11)$$

The second order TGV functional (see [5, 4]) can now be defined as

$$\text{TGV}_{\lambda,\alpha}^2(u) = \min_v \alpha_1 \|\nabla_\lambda u - v\|_1 + \alpha_0 \|\mathcal{E}_\lambda v\|_1 \quad (12)$$

where, abusing notation, $\|\cdot\|_1$ denotes discrete L^1 norms using the Frobenius norm on the matrix and tensor components, respectively. Notice that in practice the symmetrized parts of $\mathcal{E}v$ are stored only once and the Frobenius norm is adapted accordingly.

Infimal Convolution of second order Total Generalized Variation. The weighting of the temporal derivative by λ is an additional degree of freedom that appears in spatio-temporal regularization and balances spatial versus temporal regularity. It has been observed in [8] that this degree of freedom can be exploited to further improve reconstruction quality by optimally balancing between two choices of λ . This balancing might for instance be realized by the infimal convolution of two second order total generalized variation functionals (ICTGV), which is given, for $\lambda_1, \lambda_2, \beta > 0$, as

$$\text{ICTGV}_{\lambda,\beta,\alpha}^2(u) = \min_v \text{TGV}_{\lambda_1,\alpha}^2(u - v) + \beta \text{TGV}_{\lambda_2,\alpha}^2(v). \quad (13)$$

The rationale behind this balancing is that one aims at an additive decomposition of the image sequence into two parts for which either strong or weak temporal regularization is favourable, in particular regularizing almost static background strongly and moving objects weakly in time.

3 Numerical solution

In order to solve the minimization problem (6) numerically, we use the primal-dual algorithm of [6]. As we are interested in testing different spatio-temporal regularization approaches, we formulate the resulting framework in a way that covers all orders of spatio-temporal TGV regularization and infimal convolutions thereof.

For motivational purposes, consider for instance a saddle-point reformulation for minimizing the second order TGV functional plus a data fidelity \mathcal{D} , which is given formally as

$$\min_{(u,v) \in X} \|K(u,v)\|_{1,\gamma} + \mathcal{D}(u) = \min_{(u,v) \in X} \sup_{\substack{w \in Y \\ \|w\|_{\infty,\gamma} \leq 1}} (K(u,v), w) + \mathcal{D}(u) \quad (14)$$

with

$$K = \begin{pmatrix} \nabla_\lambda & -I \\ 0 & \mathcal{E}_\lambda \end{pmatrix} : X \rightarrow Y \quad (15)$$

a linear operator and $\|\cdot\|_{1,\gamma}$ a weighted sum of ℓ^1 norms. Here, the primal and dual variables are given as $(u,v) \in X = U \times U^3$ and $w \in Y = U^3 \times U^9$, respectively, with u containing image data and v balancing between different orders of differentiation. Keeping this in mind, we now consider the numerical solution of (6) in a slightly more general form. For $L_X, L_Y \in \mathbb{N} \cup \{0\}$, $l_k^X, l_k^Y \in \mathbb{N}$, we define the spaces of primal and dual variables X and Y as

$$X = \bigotimes_{k=0}^{L_X} U^{l_k^X} \quad \text{and} \quad Y = \bigotimes_{k=0}^{L_Y} U^{l_k^Y} \quad (16)$$

with $l_0^X = 1$ and the norms $\|\cdot\|_X, \|\cdot\|_Y, \|\cdot\|_Z$ on X, Y, Z being the standard inner product norms. Further we define the norm $\|\cdot\|_{1,\gamma}$ on Y as

$$\|w\|_{1,\gamma} = \sum_{k=0}^{L_Y} \gamma_k \|w_k\|_1 \quad (17)$$

where, abusing notation, the norms $\|\cdot\|_1$ on $U^{l_k^Y}$ are given as $\|r\|_1 = \sum_{i,j,t} |r_{i,j,t}|$, with $|r_{i,j,t}|$ the Frobenius norm of $r_{i,j,t} \in \mathbb{R}^{3 \times l_k^Y}$.

Letting $K : X \rightarrow Y$ be a linear operator, the minimization problem (6) with any order of TGV regularization and infimal convolutions thereof can be written as

$$\min_{\substack{(s,v) \in X, \\ d \in Z}} \|K(s + \hat{S}Ad, v)\|_\gamma + \mathcal{I}_D(d) + \mathcal{I}_{\ker(S)}(s). \quad (18)$$

A particular instance of this setting is for example obtained with

$$L_X = L_Y = 0, K = \nabla_\lambda, \gamma_0 = 1 \quad (19)$$

and corresponds to spatio-temporal TV regularization (which we do not consider in this paper). The case

$$L_X = L_Y = 1, K = \begin{pmatrix} \nabla_\lambda & -I \\ 0 & \mathcal{E}_\lambda \end{pmatrix}, \gamma_i = \alpha_{1-i} \quad (20)$$

corresponds to second order TGV regularization. ICTGV regularization can be obtained by setting

$$L_X = L_Y = 3, K = \begin{pmatrix} \nabla_{\lambda_1} & \nabla_{\lambda_1} & -I & 0 \\ 0 & 0 & \mathcal{E}_{\lambda_1} & 0 \\ 0 & \nabla_{\lambda_2} & 0 & -I \\ 0 & 0 & 0 & \mathcal{E}_{\lambda_2} \end{pmatrix}, \gamma_i = \alpha_{1-i}, \gamma_{i+2} = \beta\alpha_{1-i} \text{ for } i = 1, 2. \quad (21)$$

In order to find an optimal solution of (18) we apply the primal dual algorithm of [6] to the equivalent saddle point problem

$$\min_{\substack{(s,v) \in X, w \in Y \\ d \in Z}} \max_{(s,v) \in X, w \in Y} (K(s + \hat{S}Ad, v), w) - \mathcal{I}_{\{\|r\|_{\infty, \gamma} \leq 1\}}(w) + \mathcal{I}_D(d) + \mathcal{I}_{\ker(S)}(s) \quad (22)$$

where $\{\|r\|_{\infty, \gamma} \leq 1\} = \{r \in Y \mid \max_{i,j,t} |(r_k)_{i,j,t}| \leq \gamma_k \text{ for all } k\}$. Note that, as the regularization functional is continuous, it follows by standard arguments from convex analysis [7, Theorem III.4.1] that the saddle-point problem is indeed equivalent, in particular possesses a solution and the primal solution $((s, v), d)$ of the saddle-point problem is a solution to the original problem.

A first version of the primal-dual algorithm for MPEG decompression can now be written as Algorithm 1. There, the operator $\text{prox}_\gamma(w)$ is a projection on the set $\{\|r\|_{\infty, \gamma} \leq 1\}$ which can be calculated explicitly and pointwise. The operator $\text{proj}_D(d)$ can also be calculated explicitly and pointwise as the projection on the set $\{d \in Z \mid d_i \in J_i, i = 1, \dots, N\}$.

In order to ensure convergence of the algorithm we apply an adaptive step-size choice that has been proposed in [8]: With $\eta > 0$ used to fix a ratio between the primal and dual step-size σ and τ , we update the step-size parameters σ, τ at the end of each iteration as

$$\sigma = \mathcal{S}(\sigma\tau, \mathbf{n})\eta, \quad \tau = \mathcal{S}(\sigma\tau, \mathbf{n})/\eta \quad (23)$$

with $\mathbf{n} = \|(s_+ - s, v_+ - v, d_+ - d)\|_{X \times Z}^2 / \|K(s_+ - s + \hat{S}A(d_+ - d), v_+ - v)\|_Y^2$ (see the algorithm description). The function $\mathcal{S}(\cdot, \cdot)$ is chosen such that a reduction of the step-sizes σ and τ occurs finitely many – say k_0 – times, and ensures that after these reductions $\sigma\tau L_{k_0}^2 < 1$ with $L_{k_0} = \sup_{k > k_0} \{\sqrt{1/\mathbf{n}_k}\}$ and \mathbf{n}_k being the above quotient for the k th iterates. As can be seen by checking the convergence proof of [6], this is sufficient to ensure convergence of the proposed algorithm.

At last, we define a suitable stopping criterion for Algorithm 1 that ensures optimality. For that purpose, we use a modified primal dual gap. By Fenchel-Rockafellar duality, a solution $(\hat{s}, \hat{v}, \hat{d}, \hat{w})$ of the saddle-point problem (22) must satisfy

$$0 = \|K(\hat{s} + \hat{S}A\hat{d}, \hat{v})\|_\gamma + \mathcal{I}_D(\hat{d}) + \mathcal{I}_{\ker(S)}(\hat{s}) + \mathcal{I}_{\{\|w\|_{\infty, \gamma} \leq 1\}}(\hat{w}) + G^*(\hat{r}_1, \hat{r}_2, A^* \hat{S}^* \hat{r}_1) \quad (24)$$

Algorithm 1 Regularized MPEG decomposition

```

1: function MPEG-REC(FILE)
2:    $D \leftarrow$  Decoding of FILE
3:    $d, d_+ \leftarrow$  Center point( $D$ ),  $s, s_+ \leftarrow 0$ ,  $v, v_+ \leftarrow 0$ ,  $w \leftarrow 0$ ,  $\sigma, \tau > 0$ 
4:   repeat
5:      $w \leftarrow \text{prox}_\gamma(w + \sigma K(s_+ + \hat{S}Ad_+, v_+))$ 
6:      $(s_+, v_+) \leftarrow K^*(w)$ 
7:      $d_+ \leftarrow d - \tau A^* \hat{S}^*(s_+)$ 
8:      $(s_+, v_+) \leftarrow (s, v) - \tau(s_+, v_+)$ 
9:      $s_+ \leftarrow s_+ - \hat{S}S(s)$ 
10:     $d_+ \leftarrow \text{proj}_D(d_+)$ 
11:     $((s, v, d), (s_+, v_+, d_+)) \leftarrow ((s_+, v_+, d_+), (2s_+ - s, 2v_+ - v, 2d_+ - d))$ 
12:  until Stopping criterion fulfilled
13:  return  $(s + \hat{S}Ad)$ 
14: end function

```

with $(\hat{r}_1, \hat{r}_2) = -K^*\hat{w}$ and G^* the convex-conjugate of $(s, v, d) \mapsto \mathcal{I}_D(d) + \mathcal{I}_{\ker(S)}(s)$. An easy calculation shows that G^* can be given explicitly as

$$G^*(s, v, d) = \left(\sum_{i=1}^N \frac{d_i^l + d_i^r}{2} d_i + \frac{d_i^r - d_i^l}{2} |d_i| \right) + \mathcal{I}_{\ker(I - \hat{S}S)}(s) + \mathcal{I}_{\{0\}}(v). \quad (25)$$

Replacing the equality constraints by an ℓ^1 -type penalty results in the following normalized approximate primal-dual gap as stopping criterion

$$\mathcal{G}(s, v, d, w) = \frac{1}{nmT} \left| \left\| K(s + \hat{S}Ad, v) \right\|_\gamma + \tilde{G}^*(r_1, r_2, A^* \hat{S}^* r_1) \right| \quad (26)$$

where again $(r_1, r_2) = -K^*w$ and

$$\tilde{G}^*(s, v, d) = \left(\sum_{i=1}^N \frac{d_i^l + d_i^r}{2} d_i + \frac{d_i^r - d_i^l}{2} |d_i| \right) + \left(\sum_{i,j,t} |(s - \hat{S}Ss)_{i,j,t}| + |v_{i,j,t}| \right). \quad (27)$$

Straightforward evaluation of the step-size rule and the approximate primal-dual gap would result in two additional evaluations of K and A and one additional evaluation of K^* and A^* in each iteration step. As these operations are computationally expensive, we use an equivalent formulation of the primal-dual algorithm, see Algorithm 2. It includes the step-size adaptation and approximate primal-dual gap computation without additional evaluations of K , K^* , A , A^* at the cost of slightly more memory requirement.

4 Numerical Experiments

We carry out numerical experiments on two test videos that have been compressed using a standard MPEG-2 compression software. For regularization, we

Algorithm 2 Regularized MPEG decomposition (+ step-size and error control)

```

1: function MPEG-REC(FILE)
2:    $D \leftarrow$  Decoding of FILE
3:    $d_+, d \leftarrow$  Center point( $D$ ),  $s_+, s \leftarrow 0$ ,  $v_+, v \leftarrow 0$ ,  $w \leftarrow 0$ 
4:    $x_+, x \leftarrow K(s + \hat{S}Ad, v)$ ,  $\epsilon, \eta > 0$ ,  $\sigma \leftarrow \eta$ ,  $\tau \leftarrow 1/\eta$ 
5:   repeat
6:      $w \leftarrow \text{prox}_\gamma(w + \sigma(2x_+ - x))$ 
7:      $(s, v) \leftarrow -K^*(w)$ 
8:      $d \leftarrow A^* \hat{S}^* s$ 
9:      $\mathcal{G} \leftarrow \tilde{G}^*(s, v, d)$ 
10:     $(s, v, d) \leftarrow (s_+, v_+, d_+) + \tau(s, v, d)$ 
11:     $(s, d) \leftarrow (s - \hat{S}Ss, \text{proj}_D(d))$ 
12:     $n_1 \leftarrow \|(s, v, d) - (s_+, v_+, d_+)\|_{X \times Z}^2$ 
13:     $x \leftarrow K(s + SAd, v)$ 
14:     $n_2 \leftarrow \|x - x_+\|_Y^2$ 
15:     $(\sigma, \tau) \leftarrow (\mathcal{S}_\eta(\sigma\tau, n_1/n_2)\eta, \mathcal{S}_\eta(\sigma\tau, n_1/n_2)/\eta)$ 
16:     $\mathcal{G} \leftarrow (\mathcal{G} + \|x\|_\gamma)/(nmT)$ 
17:     $((s_+, v_+, d_+, x_+), x) \leftarrow ((s, v, d, x), x_+)$ 
18:  until  $\mathcal{G} < \epsilon$ 
19:  return  $(s_+ + \hat{S}Ad_+)$ 
20: end function

```

use spatio-temporal TGV and ICTGV regularization as in (12) and (13), respectively. The TGV_α^2 functional was defined using a time-weight of $\lambda = 0.5$ and parameters $\alpha_0 = 3$, $\alpha_1 = 1$. The parameters for the ICTGV functional were fixed empirically as $\lambda_1 = \beta = \lambda$, $\lambda_2 = 1/\lambda$, $\lambda = 0.2$ and α as before, which implies a weighting of the spatial and temporal derivative by 1 and 0.2 for the first term and by 0.2 and 1.0 for the second one, respectively. To ensure optimality, we require the approximate primal-dual gap as defined in (26) to be below 0.2.

For the first experiment, we consider the *Juggler* image sequence obtained from [11]. The compressed MPEG file consists of one GOP of 8 frames and requires 0.38 bits per pixel (bpp), which means a compression ratio of 62. The GOP consists of one initial I frame and seven P frames. Figure 1 shows the standard decompressed version of this image sequence together with a second order TGV and ICTGV regularized version. As one can observe, both regularizations are able to significantly reduce compression artifacts. The ICTGV based reconstruction shows more details for example in the slowly moving face of the juggler, which can be recovered due to the strong temporal regularization of the second component of the functional. We also observe that, in particular in the later frames, some compression artifacts are still visible. This is due to the fact that the data intervals for the residual BDCT data of motion-compensated macroblocks have been reduced to point intervals. Improving on this will be subject of future research. To allow a more detailed evaluation, we have made results for different compression rates available online at [10]. For a second test, we

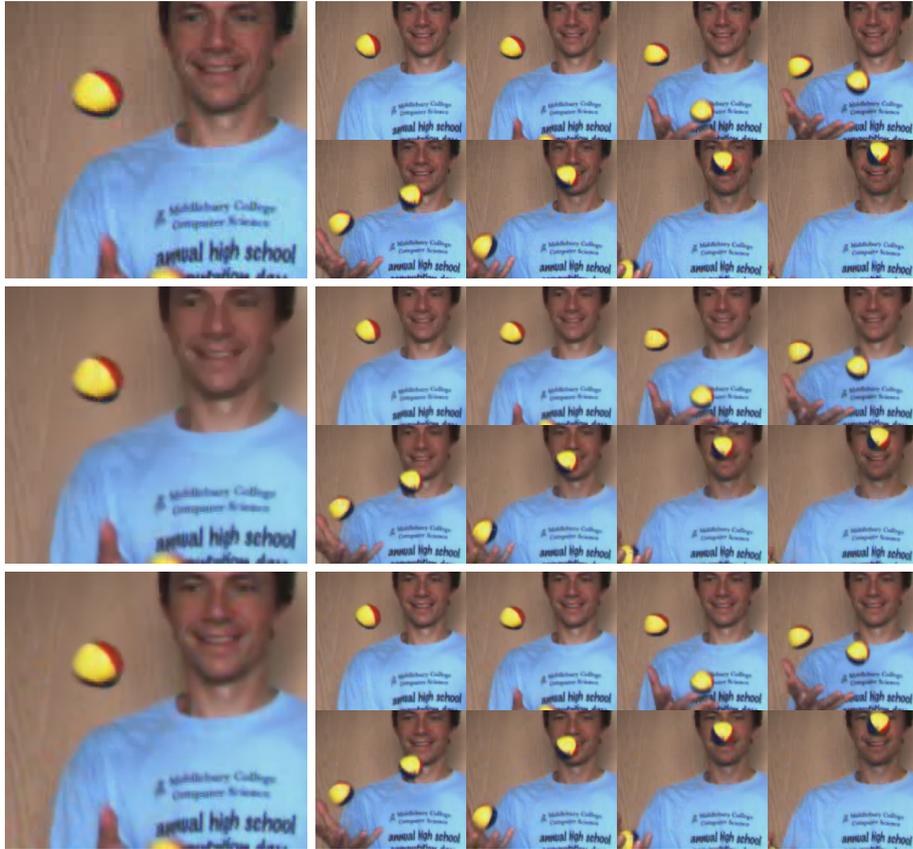


Fig. 1. From top to bottom: Standard, TGV-based and ICTGV-based decompression of the *Juggler* video. On the left, the second frame (P-frame) is shown in detail while on the right all 8 frames are depicted.

use the *Tempete* image sequence obtained from [14]. The compressed MPEG file consists of 32 frames and requires 0.98 bpp, which implies a compression ratio of 24. The GOP size was set to 9 and the frame structure of each GOP is given as IBPBPBPBP. The image sequence shows a slow zoom-out of a landscape with quickly moving leaves in the foreground, in particular there is, in contrast to the juggler sequence, no constant background region. Figure 2 show the result of standard decompression and ICTGV based decompression for the fourth frame, which is a B-frame. We can observe in the top images that the ICTGV based reconstruction still shows a high level of detail while, as can be seen in the close ups, compression artifacts are significantly reduced.

References

1. G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*. Springer, 2006.



Fig. 2. Left: Standard decomposition of frame four (B-frame) of the *Tempete* video with a close up. Right: The same frame of the ICTGV based decomposition.

2. K. Bredies and M. Holler. A total variation-based JPEG decomposition model. *SIAM J. Imag. Sci.*, 5(1):366–393, 2012.
3. K. Bredies and M. Holler. Artifact-free decomposition and zooming of JPEG compressed images with total generalized variation. In *CCIS*, volume 359, pages 242–258. Springer, 2013.
4. K. Bredies and M. Holler. Regularization of linear inverse problems with total generalized variation. *J. Inverse Ill-Posed Probl.*, 2014. doi:10.1515/jip-2013-0068.
5. K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM J. Imag. Sci.*, 3(3):492–526, 2010.
6. A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40:120–145, 2011.
7. I. Ekeland and R. Témam. *Convex Analysis and Variational Problems*. SIAM, 1999.
8. M. Holler and K. Kunisch. On infimal convolution of TV-type functionals and applications to video and image reconstruction. *SIAM J. Imag. Sci.*, 7(4):2258–2300, 2014.
9. B. Kocovski, T. Kartalov, Z. Ivanovski, and L. Panovski. An adaptive deblocking algorithm for low bit-rate video. In *ISCCSP 2008. 3rd International Symposium on*, pages 888–893, March 2008.
10. Webpage of Martin Holler. <http://www.uni-graz.at/~hollerm/research.html>.
11. Middlebury optical flow dataset. <http://vision.middlebury.edu/flow/>.
12. H. Schaeffer, Y. Yang, and S. Osher. Space-time regularization for video decomposition. 2014. Preprint.
13. A. Schrijver. *Theory of Linear and Integer Programming*. John Wiley & Sons, 1986.
14. Arizona State University. Video trace library. <http://trace.eas.asu.edu/>.
15. S. Zhong. Image coding with optimal reconstruction. In *International Conference on Image Processing*, volume 1, pages 161–164, 1997.