A l$_1$-NORM PRESERVING MOTION-COMPENSATED TRANSFORM FOR SPARSE APPROXIMATION OF IMAGE SEQUENCES

Markus Flierl

ACCESS Linnaeus Center, School of Electrical Engineering
KTH Royal Institute of Technology, Stockholm, Sweden
mflierl@kth.se

ABSTRACT

This paper discusses an adaptive non-linear transform for image sequences that aims to generate a l_1-norm preserving sparse approximation for efficient coding. Most sparse approximation problems employ a linear model where images are represented by a basis and a sparse set of coefficients. In this work, however, we consider image sequences where linear measurements are of limited use due to motion. We present a motion-adaptive non-linear transform for a group of pictures that outputs common and detail coefficients and that minimizes the l_1 norm of the detail coefficients while preserving the overall l_1 norm. We demonstrate that we can achieve a smaller l_1 norm of the detail coefficients when compared to that of motion-adaptive linear measurements. Further, the decay of normalized absolute coefficients is faster than that of motion-adaptive linear measurements.

Index Terms— Sparse approximation, l_1 norm, motion compensation, image sequence processing.

1. INTRODUCTION

We assume that video signals are sparse or compressible in the sense that they depend essentially only on a small number of degrees of freedom. Most sparse approximation problems employ a linear model where the signal can be written either exactly or accurately as a superposition of a small number of vectors in some fixed basis [1]. Recently, it has been shown that l_1 minimization is an efficient and correct method for sparse signal recovery. In particular, this is applicable to compressive sensing where a small number of linear measurements is used to recover sparse signals. The compressive sensing framework states that if a signal can be approximated using a sparse representation, it can also be accurately reconstructed from a small collection of linear measurements [2]. For example, exact signal reconstruction from highly incomplete frequency information is demonstrated in [3]. Further, an application of compressive sensing to images and video is presented in [4].

We are interested in sparse representations of video that can be coded efficiently. For example, [5] presents a spatio-temporal representation that uses a sparse decomposition algorithm along motion trajectories. Sparse approximations are obtained by utilizing the matching pursuit algorithm on redundant dictionaries. In the present work, however, we argue that linear measurements are of limited use due to motion. We present a motion-adaptive non-linear transform for a group of pictures that outputs common and detail coefficients and that minimizes the l_1 norm of the detail coefficients while preserving the overall l_1 norm. We compare our results to a motion-compensated orthogonal transform for image sequences that offers

a motion-adaptive linear representation while preserving the overall l_2 norm [6].

The paper is organized as follows: Section 2 introduces the l_1-norm preserving motion-compensated transform and the l_1 minimization of the detail coefficients. Section 3 presents experimental results, compares the l_1 norm of the detail coefficients, and discusses the decay of normalized absolute coefficients.

2. MOTION-COMPENSATED TRANSFORM

Let $\mathbf{x}_1$ and $\mathbf{x}_2$ be two positive vectors representing consecutive pictures of an image sequence. The non-linear transform T maps these vectors according to

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \end{bmatrix} = T\left(\begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right) \tag{1}$$

into two vectors $\mathbf{y}_1$ and $\mathbf{y}_2$ which represent common and detail coefficients, respectively. Let the transform be dependent on the motion vectors between the two input pictures. Given T, there exists an unique inverse transform $T^{-1}(\mathbf{y}) = \mathbf{x}$ that recovers the original signal from the coefficients. Both T and T^{-1} conserve the l_1 norm of the signal and the coefficients such that

$$\|\mathbf{y}\|_1 = \|\mathbf{x}\|_1. \tag{2}$$

The problem can be formulated as follows: Given the vector of input pictures $\mathbf{x}$, find a motion-dependent transform whose coefficient vector $\mathbf{y}$ recovers the vector of input pictures $\mathbf{x}$ while minimizing the l_1 norm of the detail coefficients $\mathbf{y}_2$.

$$\min \|\mathbf{y}_2\|_1 \quad \text{s.t.} \quad T^{-1}(\mathbf{y}) = \mathbf{x} \tag{3}$$

To simplify the problem of constructing an invertible transform, we write the non-linear transform T as a concatenation of k incremental transforms T_κ such that

$$T = T_k \circ T_{k-1} \circ \cdots \circ T_\kappa \circ \cdots \circ T_2 \circ T_1, \tag{4}$$

where each incremental transform has a unique inverse and conserves the l_1 norm of its input vector. This guarantees that T is invertible and l_1 conserving. It can be imagined that the pixels in image $\mathbf{x}_2$ are processed from top-left to bottom-right in k steps where each step κ is represented by the incremental transform T_κ.

2.1. Incremental Transform

Let $\mathbf{x}_1^{(\kappa)}$ and $\mathbf{x}_2^{(\kappa)}$ be two vectors representing consecutive pictures of an image sequence if $\kappa = 1$, or two output vectors of the incremental transform $T_{\kappa-1}$ if $\kappa > 1$. The incremental transform T_κ

maps these vectors according to

$$\begin{bmatrix} \mathbf{x}_1^{(\kappa+1)} \\ \mathbf{x}_2^{(\kappa+1)} \end{bmatrix} = T_\kappa \left(\begin{bmatrix} \mathbf{x}_1^{(\kappa)} \\ \mathbf{x}_2^{(\kappa)} \end{bmatrix} \right) \tag{5}$$

into two vectors $\mathbf{x}_1^{(\kappa+1)}$ and $\mathbf{x}_2^{(\kappa+1)}$ which will be further transformed into common and detail coefficients, respectively.
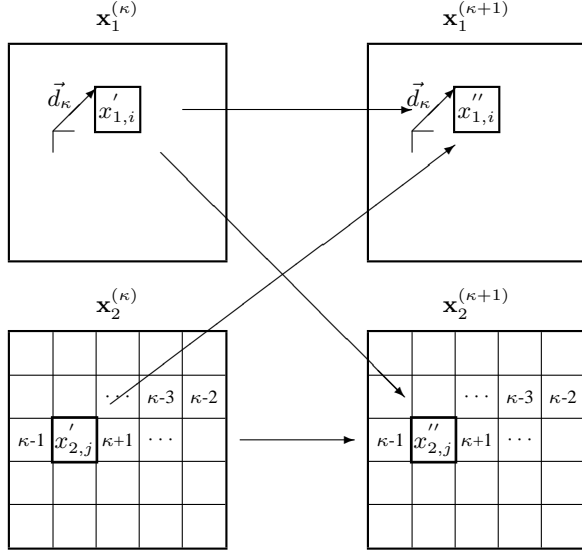

Fig. 1. The incremental transform T_κ for two frames $\mathbf{x}_1^{(\kappa)}$ and $\mathbf{x}_2^{(\kappa)}$ at step κ that uses the motion vector $\vec{d}_\kappa$.

Fig. 1 depicts the process accomplished by the incremental transform T_κ with its input and output images as defined above. The incremental transform removes the l_1 norm of the j-th pixel $x'_{2,j}$ in the image $\mathbf{x}_2^{(\kappa)}$ with the help of the i-th pixel $x'_{1,i}$ in the image $\mathbf{x}_1^{(\kappa)}$ which is linked by the motion vector $\vec{d}_\kappa$ (or of the j-th block with the help of the i-th block if all pixels of the block have the same motion vector $\vec{d}_\kappa$). The l_1-removed pixel value $x''_{2,j}$ is a function of the pixel values $x'_{1,i}$ and $x'_{2,j}$. The l_1-concentrated pixel value $x''_{1,i}$ is also a function of the pixel values $x'_{1,i}$ and $x'_{2,j}$. All other pixels are simply kept untouched.

As each incremental transform modifies only the pixel values $x_{1,i}$ and $x_{2,j}$, it can essentially be captured by the simple transform H_α as defined by

$$\begin{bmatrix} x''_{1,i} \\ x''_{2,j} \end{bmatrix} = H_\alpha \left(\begin{bmatrix} x'_{1,i} \\ x'_{2,j} \end{bmatrix} \right), \tag{6}$$

which rotates signal points in a plane by the angle α while conserving their l_1 norm $|x''_{1,i}| + |x''_{2,j}| = |x'_{1,i}| + |x'_{2,j}|$. For images, $x'_{1,i}$ and $x''_{1,i}$ are always positive. Hence, the l_1 conservation law yields

$$x''_{1,i} = \frac{1 + \left| \frac{x'_{2,j}}{x'_{1,i}} \right|}{1 + \left| \frac{x''_{2,j}}{x''_{1,i}} \right|} x'_{1,i}. \tag{7}$$

The input signal point is located in the plane at angle ϕ (see Fig. 2). H_α rotates by α such that the output signal point is located in the plane at angle $\phi - \alpha$.

$$\frac{x'_{2,j}}{x'_{1,i}} = \tan(\phi) \tag{8}$$

$$\frac{x''_{2,j}}{x''_{1,i}} = \tan(\phi - \alpha) \tag{9}$$

Using trigonometric identities, we obtain the values of the output signal point as a function of α.

$$\frac{x''_{2,j}}{x''_{1,i}} = \frac{\frac{x'_{2,j}}{x'_{1,i}} - \tan(\alpha)}{1 + \frac{x'_{2,j}}{x'_{1,i}} \tan(\alpha)} \tag{10}$$

The rotation angle α is determined in the next subsection which discusses the l_1 minimization.

Note that to accomplish the transform T, each pixel in $\mathbf{x}_2$ is touched only once whereas the pixels in $\mathbf{x}_1$ may be touched multiple times or never. Further, the order in which the incremental transforms T_κ are applied does not affect the l_1 conservation of T. But the order may affect l_1 minimization.

2.2. l_1 Minimization

The rotation angle α for each pixel touched by the incremental transform has to be chosen such that the l_1 norm is minimized for the image $\mathbf{x}_2$. We discuss a method that reduces the l_1 norm of detail coefficients to zero if each motion vector connects pixels with identical intensity.
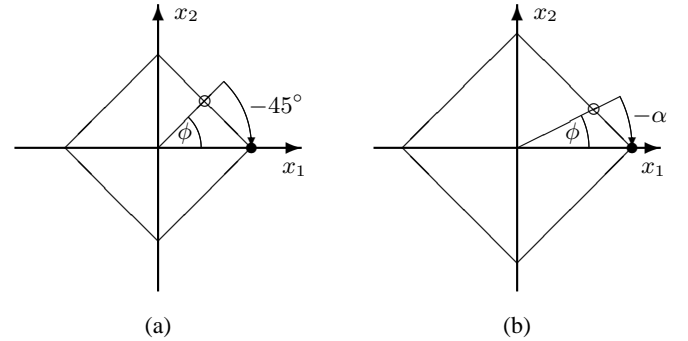

(a)　　　　　　　　　　(b)

Fig. 2. The incremental transform T_κ as a rotation of a signal point on the l_1 circle. Consider perfect translatory motion: (a) As not modified in previous incremental transforms, pixel x_1 and x_2 have the same intensity value. (b) Pixel x_1 has been modified by a scale factor in a previous incremental transform.

Consider the pixel pair $x_{1,i}$ and $x_{2,j}$ to be processed by the incremental transform T_κ. To determine the rotation angle α for the pixel $x_{2,j}$, we assume that the pixel $x_{2,j}$ is connected to the pixel $x_{1,i}$ such that $x_{2,j} = x_{1,i}$. Consequently, the resulting detail pixel $x''_{2,j}$ shall be zero. Note that the pixel $x_{1,i}$ may have been processed previously by T_τ, where $\tau < \kappa$. Therefore, let v_1 be the *scale factor* for the pixel $x_{1,i}$ such that $x'_{1,i} = v_1 x_{1,i}$. The pixel $x_{2,j}$ is used only once during the transform process T and no scale factor needs to be considered. But for sake of generality, let v_2 be the scale factor for the pixel $x_{2,j}$ such that $x'_{2,j} = v_2 x_{2,j}$. Let u_1 be the scale factor

for the pixel $x_{1,i}$ after it has been processed by T_κ. Now, the pixels $x'_{1,i}$ and $x'_{2,j}$ are processed by T_κ as follows:

$$\begin{bmatrix} u_1 x_{1,i} \\ 0 \end{bmatrix} = H_\alpha \left(\begin{bmatrix} v_1 x_{1,i} \\ v_2 x_{1,i} \end{bmatrix} \right) \tag{11}$$

Fig. 2 depicts the incremental transform T_κ as a rotation of a signal point on the l_1 circle. If the scale factor for both pixels is one, we need to rotate the signal point by $-45°$ to obtain a zero detail coefficient. For general scale factors, we need to compensate the angle ϕ. In any case, l_1 minimization of the detail coefficient implies $\tan(\phi - \alpha) = 0$, i.e. $\phi = \alpha$, such that

$$\tan(\alpha) = \frac{x'_{2,j}}{x'_{1,i}} = \frac{v_2 x_{1,i}}{v_1 x_{1,i}} \tag{12}$$

$$\tan(\alpha) = \frac{v_2}{v_1}. \tag{13}$$

Finally, rotation on the l_1 circle conserves the l_1 norm. It determines the scale factor u_1 of the common pixel after it has been processed by H_α.

$$|u_1 x_{1,i}| = |v_1 x_{1,i}| + |v_2 x_{1,i}| \tag{14}$$

$$u_1 |x_{1,i}| = v_1 |x_{1,i}| + v_2 |x_{1,i}| \tag{15}$$

$$u_1 = v_1 + v_2 \tag{16}$$

Now, let $n_1, n_2, m_1 \in \mathbb{N}_0$ be *scale counters* such that their corresponding scale factors satisfy

$$v_\mu = n_\mu + 1 \quad \text{for} \quad \mu = 1, 2 \quad \text{and} \tag{17}$$

$$u_1 = m_1 + 1. \tag{18}$$

n_1 simply counts how often the pixel $x_{1,i}$ is used as reference for motion compensation. In the beginning, i.e., before the transform is applied, the scale counter for each pixel $x_{1,i}$ is $n_1 = 0$ and its scale factor is $v_1 = 1$.

As noted previously, the transform T touches each pixel in $\mathbf{x}_2$ only once and, hence, no scale factor needs to be considered as long as only one transform T is applied. But if four of more images are to be transformed, a hierarchical procedure can be applied such that the common coefficients of the first level are the input to a transform on the second level. As the common coefficients come with scale factors, they have to be considered for both input images of the transform on the second level. Therefore, we use a scale counter n_2 for each pixel $x_{2,j}$ to count how often it has been used as reference in a transform on the first level. Obviously, the scale counters n_2 for the transforms on the first level are set to zero.

The conservation of the l_1 norm in (16) requires the scale factors to satisfy $u_1 = v_1 + v_2$. With the definition of scale counters in (17) and (18), we obtain a *scale counter update rule* for the l_1-norm preserving motion-compensated transform.

$$m_1 = n_1 + n_2 + 1 \tag{19}$$

That is, each pixel that is used as reference for motion compensation by the incremental transform receives a scale counter update by $n_2 + 1$. This rule applies within the transform T at any level of a hierarchical decomposition.

Finally, with the help of the definition of scale counters, the l_1 minimization of the detail coefficients is achieved by the transform H_α with the angle

$$\tan(\alpha) = \frac{n_2 + 1}{n_1 + 1}, \tag{20}$$

where the scale counters n_1 and n_2 are maintained according to (19).

2.3. Inverse Incremental Transform

The inverse incremental transform takes the common and detail coefficients and operates as the transform H_α, but rotates by the angle $-\alpha$. Again, the scale counters n_1 and n_2 are used to determine this angle.

$$\frac{x'_{2,j}}{x'_{1,i}} = \frac{\frac{x''_{2,j}}{x''_{1,i}} + \frac{n_2+1}{n_1+1}}{1 - \frac{x''_{2,j}}{x''_{1,i}} \frac{n_2+1}{n_1+1}} \tag{21}$$

The conservation of the l_1 norm recovers the positive pixel values in the images.

$$x'_{1,i} = \frac{1 + \left| \frac{x''_{2,j}}{x''_{1,i}} \right|}{1 + \left| \frac{x'_{2,j}}{x'_{1,i}} \right|} x''_{1,i} \tag{22}$$

After each inverse incremental transform, scale counters have to be reduced accordingly. Note that scale counter values are uniquely related to the applied motion field. If the sequence of incremental transforms as applied in the forward transform is known, no extra information is necessary to reduce the scale counters correctly.

3. EXPERIMENTAL RESULTS

Experimental results assessing the l_1 minimization are obtained for the QCIF sequences *Mother & Daughter* and *Foreman*. We compare the output of the l_1-norm preserving motion-compensated transform to the orthogonal motion-compensated linear measurements obtained with the work in [6]. For both transforms, the same 8×8 block motion field is used. A scale counter n is maintained for every pixel of each picture. The scale counter values are an immediate result of the utilized motion vectors and are only required for processing.

Figs. 3 and 5 depict the average l_1-norm of detail coefficients normalized by the l_1-norm of all coefficients for l_2 and l_1-norm preserving motion-compensated transforms for *Mother & Daughter* and *Foreman*, respectively. The normalized average l_1 norm of detail coefficients is shown for groups of pictures of size 2, 4, 8, 16, and 32. This norm is decreasing with the size of the GOP as the l_1 norm is increasingly concentrated into only one common coefficient image. Moreover, the l_1 minimization results in a consistently lower normalized l_1 norm of detail coefficients when compared to that of orthogonal motion-adaptive linear measurements.

Figs. 4 and 6 visualize the decay of normalized absolute coefficients $|y[t]|/|y[1]|$ over the sorted coefficient index t for l_2 and l_1-norm preserving motion-compensated transforms for *Mother & Daughter* and *Foreman*, respectively. For this experiment, a GOP size of 8 has been chosen. Only the first 50 000 coefficients are shown. The remaining have a negligible small magnitude. Note that the decay of normalized absolute coefficients is faster than that of motion-adaptive linear measurements.

4. CONCLUSIONS

This paper presents a motion-adaptive transform for image sequences that outputs common and detail coefficients and that minimizes the l_1 norm of the detail coefficients while preserving the overall l_1 norm. It achieves a smaller l_1 norm of the detail coefficients when compared to that of motion-adaptive linear measurements. Moreover, the decay of normalized absolute coefficients is faster than that of motion-adaptive linear measurements.
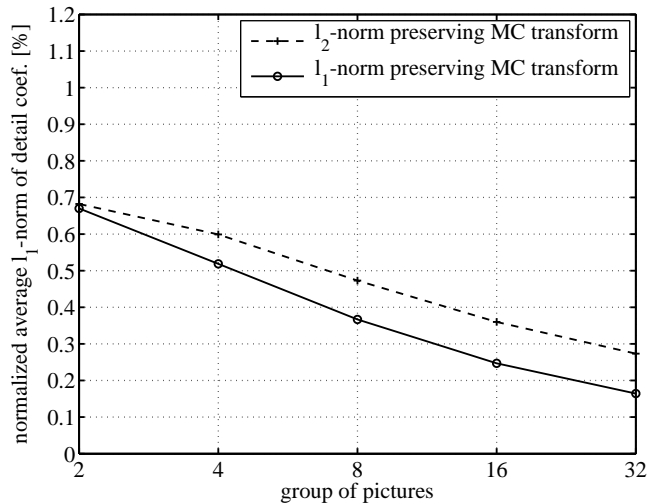
Fig. 3. Average l_1-norm of detail coefficients normalized by the l_1-norm of all coefficients over the size of the GOP for l_2 and l_1-norm preserving motion-compensated transforms for the QCIF sequence *Mother & Daughter*.
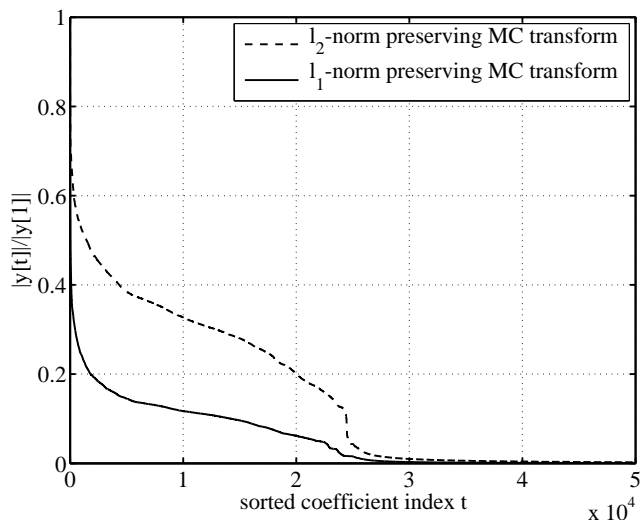

Fig. 5. Average l_1-norm of detail coefficients normalized by the l_1-norm of all coefficients over the size of the GOP for l_2 and l_1-norm preserving motion-compensated transforms for the QCIF sequence *Foreman*.

Fig. 4. Decay of normalized absolute coefficients $|y[t]|/|y[1]|$ over the sorted coefficient index t for l_2 and l_1-norm preserving motion-compensated transforms with GOP size 8 for the QCIF sequence *Mother & Daughter*.
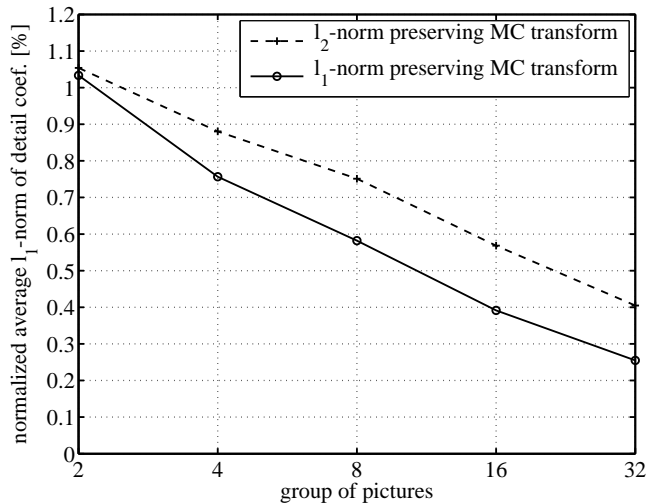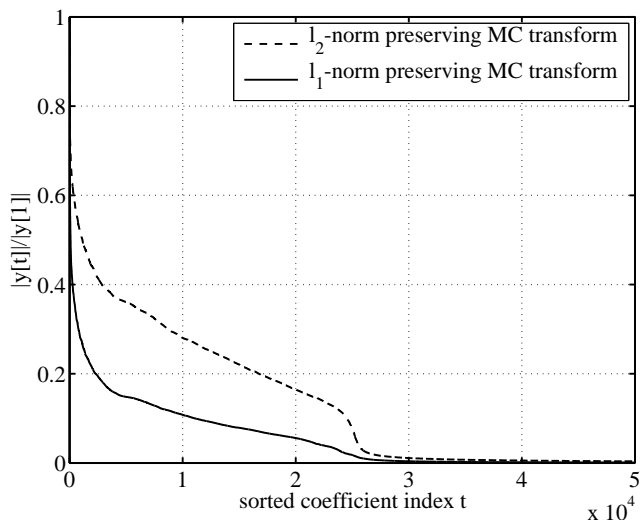

Fig. 6. Decay of normalized absolute coefficients $|y[t]|/|y[1]|$ over the sorted coefficient index t for l_2 and l_1-norm preserving motion-compensated transforms with GOP size 8 for the QCIF sequence *Foreman*.

5. REFERENCES

[1] E.J. Candes, "Compressive sampling," in *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, Aug. 2006.

[2] M.F. Duarte, M.A. Davenport, D. Takhar, J.N. Laska, T. Sun, K.F. Kelly, and R.G. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, pp. 83–91, Mar. 2008.

[3] E.J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] M.B. Wakin, J.N. Laska, M.F. Duarte, D. Baron, S. Sarvotham, D. Takhar, K.F. Kelly, and R.G. Baraniuk, "An architecture for compressive imaging," in *Proceedings of the IEEE International Conference on Image Processing*, Atlanta, GA, Oct. 2006, pp. 1273–1276.

[5] A. Rahmoune, P. Vandergheynst, and P. Frossard, "Flexible motion-adaptive video coding with redundant expansions," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 16, no. 2, pp. 178–190, Feb. 2006.

[6] M. Flierl and B. Girod, "A motion-compensated orthogonal transform with energy-concentration constraint," in *Proceedings of the IEEE Workshop on Multimedia Signal Processing*, Victoria, BC, Oct. 2006.