WYNER-ZIV CODING OF MULTI-VIEW OMNIDIRECTIONAL IMAGES WITH OVERCOMPLETE DECOMPOSITIONS

Ivana Tosic and Pascal Frossard

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Institute, CH-1015 Lausanne
{ivana.tosic, pascal.frossard}@epfl.ch

ABSTRACT

This paper addresses the problem of distributed coding of light fields in camera networks. A novel distributed coding scheme with side information is presented, based on spherical image expansion over an overcomplete dictionary of geometric atoms. We propose to model the correlation between views with local geometrical transformations of corresponding features in the sparse representations of different views. We design a Wyner-Ziv encoder by partitioning the dictionary into cosets of dissimilar atoms, with respect to their shape and position on the image. The joint decoder finds pairwise correspondences between atoms in the reference image and atoms in cosets of the Wyner-Ziv image. It selects the most likely correspondence among pairs of atoms that satisfy epipolar geometry constraints. This permits to estimate local transformations between correlated images that eventually help to refine the side information provided by the reference image. Experiments demonstrate that the proposed method is capable of estimating the geometric transformations between views, and hence to reconstruct the Wyner-Ziv image.

Index Terms— 3D scene, sparse approximations, DSC

1. INTRODUCTION

Vision sensor networks have recently gained a tremendous interest among researchers as they find popular applications in 3DTV, surveillance or robotics. Since information from cameras is most generally highly correlated, one can try to eliminate redundancy in order to reduce the bit rate in such sensor networks. A distributed approach to data compression in these networks then becomes very appealing, since it does not require communication among cameras. Interestingly, results from information theory show that the correlation among sources can be exploited at the decoder, even with independent encoders. Slepian and Wolf [1] proved in 1973 that the lower bound of the total rate for separate encoding of correlated sources is actually the joint entropy of the sources, just as it is the case in the joint encoding scheme. Few years later, Wyner and Ziv [2] extended the distributed framework to lossy compression with side information at the decoder. The first practical distributed source coding (DSC) schemes only appeared more than 2 decades later (e.g., [3, 4]), when the link of DSC with channel coding has been established. However, only a few works have addressed the problem of distributed coding in camera networks (e.g., [5, 6]), mainly due to the difficulty of modeling the statistical correlation among sources. Schemes based on channel codes assume the correlation on the level of pixel bit planes, modeled by the statistics of a virtual channel. However, as small camera movements can introduce large variations in the pixel values, the choice of an appropriate channel rate often becomes impossible without a feedback from the decoder.

The correlation between images in camera networks mostly lies in the motion of the objects in the scene, and motion estimation at the decoder certainly improves the performance of the distributed compression scheme. However, simple motion estimation is generally limited to translational motion of observed objects, and it cannot cope with local transforms such as scaling or rotation. In this work we propose a new correlation model for multi-view omnidirectional images based on local geometrical transformations of corresponding features in sparse image representations. Omnidirectional images are very suitable for 3D scene representation as they offer a 360 degrees view of the scene. Moreover, they can be easily mapped and processed on spherical manifolds. Sparse approximations, such as the one obtained with the Matching Pursuit algorithm [7], offer very good approximation performance at low bit rate as they are able to capture the most prominent signal components with a few elements selected in a redundant dictionary of atoms. Under the assumption that these components are present in correlated views, possibly under some local transform, we design a Wyner-Ziv coder by partitioning the redundant dictionary into cosets, based on atom dissimilarity. The joint decoder selects the best candidate atom within the coset with help of the side information image. The correspondences that are found during the decoding between atom expansions of both images, are further used to estimate local transformations and to build a transform field between correlated views. These transformations are used to refine the side information for decoding the following atoms. Experimental results show that the proposed method successfully finds the geometric transforms between sparse image components, and that the proposed scheme outperforms independent coding strategies at low bit rate.

2. CORRELATION IN SPARSE DECOMPOSITIONS

The main challenge in the design of an efficient distributed coding scheme for correlated sources is primarily the modeling of the underlying correlation. In the case of camera sensor networks, the correlation between images comes from the 3D motion of the objects in the scene, which results in local changes of image components

This work has been supported by the Swiss National Science Foundation under grant 20001-107970/1.

that represent the moving objects. If we decompose each image into sparse components that capture the objects in the scene, we can assume with high probability that the most prominent components are present in all images, possibly with some local transformations. We therefore propose to model the correlation between views by local geometrical transformations, which are estimated by pairing components in sparse image decompositions.

Given a certain basis, or a redundant dictionary of atoms $\mathcal{D} = \{\phi_k\}, k = 1, ..., N$, in the Hilbert space H, every image y can be represented as:

$$y = \mathbf{\Phi} x = \sum_{k=1}^{N} x_k \phi_k, \tag{1}$$

where the matrix $\mathbf{\Phi}$ is composed of atoms ϕ_k as columns. When the dictionary is over-complete, x is not unique. In order to find a compact image representation one has to search for a sparse vector x. We say that y has a *sparse* representation in $\mathcal{D}$ if the number of non-zero components in x is much smaller than the dimension of x. Therefore, the sparse representation of y is:

$$y = \mathbf{\Phi_I} c = \sum_{k \in I} x_k \phi_k, \tag{2}$$

where c is the vector of non-zero elements of x, I labels the set of atoms $\{\phi_k\}_{k \in I}$ participating in the representation, and $\mathbf{\Phi_I}$ is a submatrix of $\mathbf{\Phi}$ with respect to I.

In the case of two correlated images $y_1 = \mathbf{\Phi_{I_1}} c_1$ and $y_2 = \mathbf{\Phi_{I_2}} c_2$, there exists a subset of atoms indexed respectively by $J_1 \in I_1$ and $J_2 \in I_2$ that represent image projections of the same 3D features in the scene. We assume that these atoms are correlated, possibly under some local geometric transformation. Denote $F(\phi)$ the transform of an atom in the image decomposition that results from the motion of an object in the 3D space, or equivalently, the transformation imposed to atom ϕ in different views due to camera displacement. Therefore, the correlation between the images can be modeled as a set of transforms F_i between corresponding atoms in sets indexed by J_1 and J_2. The approximation of the image y_2 can be rewritten as the sum of the contributions of transformed atoms, and remaining atoms in I_2 :

$$y_2 = \sum_{i \in J_1} c_{2,i} F_i(\phi_i) + \sum_{k \in I_2 \setminus J_2} c_{2,k} \phi_k. \tag{3}$$

3. TRANSFORMS IN OMNIDIRECTIONAL IMAGES

Motions of objects in the 3D space introduce various types of transformations in the image projective space. Most of these transforms can be represented by the 2-D similarity group elements, which include 2-D translation, rotation and isotropic scaling of the image features. We also consider anisotropic scaling to further expand the space of possible transforms among image features. In order to efficiently capture transforms between sparse image components, we propose to use a structured redundant dictionary of atoms for image representation. Atoms in the structured dictionary are derived from a single waveform that undergoes rotation, translation and scaling. Hence, the transformation of an atom by any of the 2-D similarity group elements or anisotropic scaling, results in another atom in the same dictionary: the dictionary is invariant with respect to any transform action.

In particular, as we address the problem of distributed coding of omnidirectional images, which can be precisely mapped on a sphere, we use a dictionary of atoms on the 2-D unit sphere [7]. Given a generating function g defined in the space of square-integrable functions on a unit two-sphere S^2, $g(\theta, \varphi) \in L^2(S^2)$, the dictionary $\mathcal{D} = \{\phi_k\} = \{g_\gamma\}_{\gamma \in \Gamma}$ is constructed by changing the atom indexes $\gamma = (\theta, \varphi, \psi, \alpha, \beta) \in \Gamma$, i.e., by applying a unitary operator $U(\gamma)$: $g_\gamma = U(\gamma) g$. The triplet (θ, φ, ψ) represents Euler angles that respectively describe the motion of the atom on the sphere by angles θ and φ, and the rotation of the atom around its axis with an angle ψ. The parameters α, β then represent anisotropic scaling factors. In such a dictionary, the transform of one atom to another reduces to a transform of its parameter set γ, i.e., $g_{\gamma_j} = F(g_{\gamma_i}) = U(\gamma') g_{\gamma_i} = U(\gamma' \circ \gamma_i) g$. Note that the size and redundancy of the dictionary is directly driven by the number of distinct atom transformations.

We are interested in finding correspondences between atoms that respectively represent the images y_1 and y_2, generated by two spherical cameras that capture the same scene. Consider an atom $g_{\gamma_i}, \gamma_i \in J_1$ from the sparse decomposition of image y_1. The subset of transforms $V_i^0 = \{\gamma' | g_{\gamma_j} = F(g_{\gamma_i}) = U(\gamma') g_{\gamma_i}\}$ allows to relate g_{γ_i} to the atoms g_{γ_j} in the expansion of y_2. However, not all these transforms are feasible under epipolar constraints. These constraints represent one of the fundamental relations in multi-view analysis, and define the relation between 3D point projections ($\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{R}^3$) on two cameras, as:

$$\mathbf{z}_2^T \hat{T} R \mathbf{z}_1 = 0, \tag{4}$$

where R and T are the rotation and translation matrices of one camera frame with respect to the other, and $\hat{T}$ is obtained by representing the cross product of T with $R\mathbf{z}_1$ as matrix multiplication, i.e., $\hat{T} R \mathbf{z}_1 = T \times R \mathbf{z}_1$. The set of possible transforms is therefore reduced to the transforms that respect epipolar constraints between the atom g_{γ_i} in y_1 and the candidates atoms g_{γ_j} in y_2. We evaluate the constraints given in Eq. (4) on atom centers denoted $m_l = [sin\theta_l cos\varphi_l \;\; sin\theta_l sin\varphi_l \;\; cos\theta_l]^T$ with $l \in \{i, j\}$, and define the set $V_i^E \subseteq V_i^0$ of possible transforms of atom g_{γ_i} as:

$$V_i^E = \{\gamma' | g_{\gamma_j} = U(\gamma') g_{\gamma_i}, m_j^T \hat{T} R m_i = 0\}. \tag{5}$$

Equivalently, the set of atoms g_{γ_j} in y_2 that are possible transformed versions of the atom g_{γ_i} is denoted as the *epipolar candidate set*. It is defined by the set of atom indexes $\Gamma_i^E \subset \Gamma$, with

$$\Gamma_i^E = \{\gamma_j | g_{\gamma_j} = U(\gamma') g_{\gamma_i}, \gamma' \in V_i^E\}. \tag{6}$$

A graphical interpretation of the epipolar constraint for spherical images is shown on the Figure 1, where we denote as S_1 and S_2

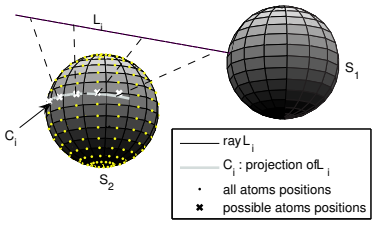


Fig. 1. Illustration of epipolar constraints in the selection of atom positions.

the two unit spheres corresponding to camera projection surfaces. The center m_j of the atom g_{γ_j}, lies on the part of a great circle $\mathcal{C}_i$ obtained by projecting the ray L_i on the sphere S_2. This ray originates from the center of camera 1 and passes through the center of atom g_{γ_i} on the sphere S_1.

Recall that corresponding atoms represent the same object in the 3D scene. Hence, we assume that the 3D motion of an object results in a limited difference between shapes of corresponding atoms, and we further restrict the set of possible transforms by constraints on the similarity of candidate atoms. We measure the similarity or coherence of atoms by the inner product $\mu(i,j) = |\langle g_{\gamma_i}, g_{\gamma_j} \rangle|$, and we impose a minimal coherence between candidate atoms, i.e., $\mu(i,j) > s$.

This defines a set of possible transforms $V_i^\mu \subseteq V_i^0$ with respect to atom shape, as:

$$V_i^\mu = \{\gamma' | g_{\gamma_j} = U(\gamma') g_{\gamma_i}, \mu(i,j) > s\}, \tag{7}$$

and a set of candidate atoms in y_2, denoted the *shape candidate set*, whose indexes belong to $\Gamma_i^\mu \subset \Gamma$, with:

$$\Gamma_i^\mu = \{\gamma_j | g_{\gamma_j} = U(\gamma') g_{\gamma_i}, \gamma' \in V_i^\mu\}. \tag{8}$$

Finally, we combine the epipolar and shape similarity constraints to define the set of possible transforms for atom g_{γ_i}, as $V_i = V_i^E \cap V_i^\mu$. Similarly, we denote the set of possible parameters of the transformed atom in y_2 as $\Gamma_i = \Gamma_i^E \cap \Gamma_i^\mu$.

4. DISTRIBUTED CODING SCHEME

4.1. Coding with side information

Based on the correlation model defined by the local transformations of atoms, we propose an asymmetric scheme for coding with side information. The sparse decomposition of the reference image y_1 is independently encoded, while the decomposition of the Wyner-Ziv image y_2 is encoded by coset coding of atom indexes and quantization of their respective coefficients, as shown on the left side of the Fig. 2. We propose to partition the set of atom indexes Γ into distinct cosets, such that atoms that belong to the same transform candidate set Γ_i are placed in different cosets. Under the assumption that an atom g_{γ_j} in the image decomposition has its corresponding atom g_{γ_i} in the side information expansion, the encoder does not need to send the entire γ_j, but only the information that is necessary to identify the correct atom in the transform candidate set given by Γ_i. The side information and the coset index are therefore sufficient to recover the atom g_{γ_j} in the Wyner-Ziv image. The achievable bit rate for encoding the atom index γ_j is reduced therefore from $R \geq H(\gamma_j | \gamma_j \in \Gamma)$ to $R \geq H(\gamma_j | \gamma_j \in \Gamma_i)$.

Due to the independency of epipolar and shape constraints, the cosets are constructed separately for atom positions (θ, φ) and atom shape parameters (ψ, α, β). We therefore construct two types of cosets, respectively

1. EPI cosets: $K_k^E = \{(\theta_{k_n}, \varphi_{k_n})\}, k = 1, ..., N_1$
2. Shape cosets: $K_l^\mu = \{(\psi_{l_n}, \alpha_{l_n}, \beta_{l_n})\}, l = 1, ..., N_2$.

The EPI cosets are designed based on the knowledge that the centers of two corresponding atoms satisfy the epipolar constraint. In practice however, epipolar constraints are rarely satisfied exactly due to discrete atom positions. We therefore extend the epipolar candidates set Γ_i^E to atoms whose centers satisfy the epipolar constraint within a certain precision (i.e., the atoms whose center m_j lies within a distance δ from the great circle $\mathcal{C}_i$), and we form the set:

$$\tilde{\Gamma}_i^E = \{\gamma_j | g_{\gamma_j} = U(\gamma') g_{\gamma_i}, d[\mathcal{C}_i, m_j] \leq \delta\}, \tag{9}$$

where $d[\cdot, \cdot]$ stands for the point-quadratic distance.

We finally construct EPI cosets by separating in different cosets all atoms whose parameters belong to $\tilde{\Gamma}_i^E$. Similarly, we design shape cosets by distributing all atoms whose parameters belong to Γ_i^μ into different cosets. The encoder eventually sends for each atom only the indexes of the corresponding cosets (i.e., k_n and l_n in Figure 2).

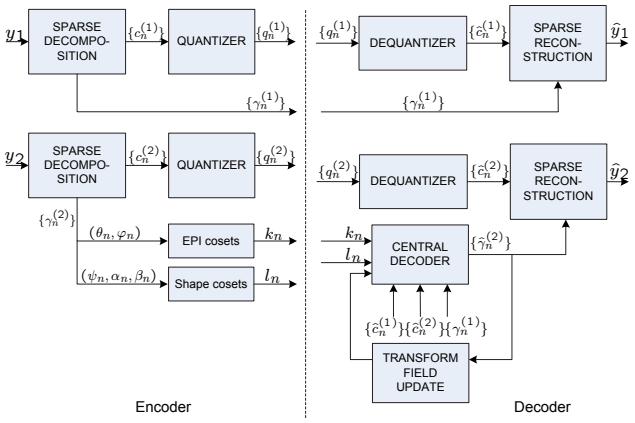


Fig. 2. Wyner-Ziv codec.

4.2. Decoder

The central decoder (CD) illustrated in the right side of Figure 2, uses the correlation model based on local atom transformations, in order to establish correspondences between atoms in the reference image and atoms within the cosets of the Wyner-Ziv image decomposition. It also uses the information provided by the quantized coefficients of atoms, along with the help of epipolar distance computed on the full atom support, in order to improve the atom matching process. Once a correspondence is identified, the decoder updates the transform field, which describes estimates of local transformation at each pixel in the image. The transformation of the reference image with respect to the transform field provides an approximation of the Wyner-Ziv image that is used as a side information. After checking all atoms in the Wyner-Ziv expansion for possible correspondences, the ones that do not have a correspondence in the reference image are simply decoded based on the maximal projection on the residual image, which is evaluated as a difference between the side information and previously decoded atoms. Finally, the WZ image reconstruction $\hat{y}_2$ is obtained as a linear combination of the decoded image y_d, formed of recovered atoms from $\mathbf{\Phi_{I_2}}$, and the transformed reference image y_{tr}, i.e.,:

$$\hat{y}_2 = y_d + \lambda \Psi_d \, y_{tr}, \tag{10}$$

where Ψ_d denotes the orthogonal complement to the basis formed by the decoded atoms in $\mathbf{\Phi_{I_2}}$, and λ is an optimization parameter. The

reconstructed Wyner-Ziv image benefits from both the decoded information and the transformed features that are not present in the decoded data. We estimate the value of λ from the energy conservation principle. Namely, under the assumption that $\|\Psi_d\, y_{tr}\| \approx \|\Psi_d\, y_2\|$, we get λ from Eq. 10 as $\lambda \approx \sqrt{1 - \|y_d\|^2/\|y_2\|^2}$, where the energy of the original image $\|y_2\|^2$ is send to the decoder as side information.

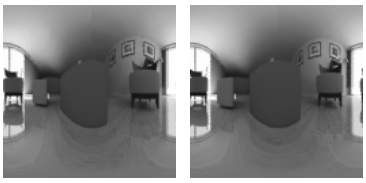
5. EXPERIMENTAL RESULTS

This section presents the performance of the proposed distributed scheme for a set of two 128×128 spherical images y_1 and y_2 that represent a synthetic room (see Figure 3), where the relative pose of one camera with respect to the other is given with $R = I$ and $T = [0 \;\; 0.3 \;\; 0]^T$. The sparse image decomposition is obtained using the Matching Pursuit algorithm on the sphere with a dictionary based on two generating functions that respectively consist in a 2D Gaussian function, and a 2D function built on a Gaussian and the second derivative of a 2D Gaussian in the orthogonal direction (i.e., edge-like atoms) [7]. The position parameters θ and φ can take 128 different values ($N_t = N_p = 128$), while the rotation parameter uses 16 orientations, between 0 and π. The scales are distributed in a logarithmic scale from 1 to $N_t/8$ for the Gaussian atoms and from 2 to $N_p/2$ for edge-like atoms, with 3 scales per octave. The choice of the dictionary is mainly driven by its good approximation properties demonstrated in [7].

The image y_1 is encoded independently at 0.23bpp with a PSNR of 30.95dB. The atom parameters for the expansion of image y_2 are coded with the proposed scheme. The coefficients are obtained by projecting the image y_2 on the atoms selected by MP, in order to improve the atom matching process, and they are quantized uniformly. We have used EPI cosets of size 1024 and shape cosets of size 128. The number of cosets depends on the correlation parameters δ (for the epipolar correlation) and s (for the shape correlation). In our simulations we have used $\delta = \pi/5$, $s_G = 0.85$ (for Gaussian atoms) and $s_A = 0.75$ (for anisotropic atoms), such that the atoms in the same coset are sufficiently different. In the cases where the center of an atom is close to the epipoles (i.e., degenerative case of epipolar constraints), its parameters are encoded independently. The rate-distortion (RD) performance of the proposed scheme for the Wyner-Ziv image is shown on the Figure 4. The dashed line represents the RD performance of independent coding with Matching Pursuit, while the solid line represents our distributed coding scheme, given by the RD curve of the reconstructed image $\hat{y}_2$. The proposed scheme clearly outperforms the independent coding, especially at low rates. The dash-dotted line represents the RD curve of the side information image, obtained by the application of the transform field on the reference image (dotted line with circles), showing that the transform field significantly improves the side information. Moreover, it can be noted that the combination of y_d (dotted line with triangles) and y_{tr} results in a better overall PSNR of the $\hat{y}_2$. We can thus conclude that these results are very encouraging, since the coset construction and the decoder are not fully optimized.

6. CONCLUSIONS

We have presented an algorithm for coding with side information networks of omnidirectional cameras. It relies on a novel correlation model between omnidirectional views that is based on local geometrical transforms of signal components given in sparse image decompositions. The distributed coding strategy exploits this correlation through the design of cosets of dissimilar atoms with respect to shape and epipolar matching. Correspondences between images allow to build a transform vector field, which is further used to construct the side information at decoder. Encouraging results show good approximation performance for the Wyner-Ziv image at low bitrate.



(a) y_1 **(b)** y_2

Fig. 3. Original Room images (128x128).

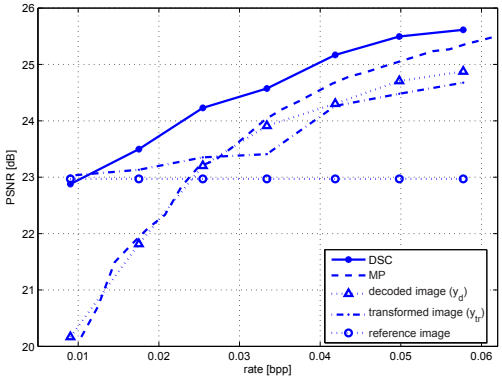


Fig. 4. RD performance of the proposed Wyner-Ziv coding scheme.

7. REFERENCES

[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources", IEEE Trans. on Inform. Theory, vol. 19(4), pp. 471-480, July 1973.

[2] A. D. Wyner and J. Ziv, "The rate-distortion function for source coding with side-information at the decoder", IEEE Trans. on Inform. Theory, vol. 22(1), pp. 1-10, January 1976.

[3] B. Girod, A. Aaron, S. Rane and D. Rebollo-Monedero, "Distributed video coding", Proc. of the IEEE, vol. 93(1), pp. 71 - 83, January 2005.

[4] R. Puri and K. Ramchandran, "PRISM: A "reversed" multimedia coding paradigm", Proc. of ICIP 2003, vol. 1, pp. 617-620, September 2003.

[5] J. Kusuma, L. Doherty and K. Ramachandran, "Distributed Compression for Sensor networks", Proc. of ICIP 2001, vol. 1, pp. 82-85, October 2001.

[6] M. Flierl and P. Vandergheynst, "Distributed Coding of Highly Correlated Image Sequences with Motion-Compensated Temporal Wavelets", EURASIP Journal on Applied Signal Processing, vol. 2006, Article ID 46747, 10 pages, 2006.

[7] I.Tosic, P. Frossard and P. Vandergheynst,"Progressive Coding of 3-D Objects Based on Overcomplete Decompositions," IEEE Trans. on Cir. and Sys. for Video Tech., vol. 16(11), pp. 1338–1349, November 2006.