# GEOMETRY-BASED DISTRIBUTED CODING OF MULTI-VIEW OMNIDIRECTIONAL IMAGES

*Ivana Tosic and Pascal Frossard*

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory (LTS4), CH-1015 Lausanne
{ivana.tosic, pascal.frossard}@epfl.ch

**ABSTRACT**

This paper presents a distributed and occlusion-robust coding scheme for multi-view omnidirectional images, which relies on the geometry of the 3D scene. The Wyner-Ziv coder uses a multi-view correlation model that relates 3D features in different images using local geometric transforms in order to perform coset code design and the coset decoding of each feature. The meaningful image features are extracted by a sparse decomposition over a dictionary of localized geometric atoms. However, in such a decomposition, occlusions or low-correlated features appear as independent elements in the encoded stream, which can lead to erroneous reconstruction at the decoder. To ameliorate this problem, we propose to leave a controlled redundancy by sending additional syndrome bits that are computed by channel coding across the atoms of the Wyner-Ziv image. This offers resiliency against occlusions, or against inaccuracy in the view correlation model. The experimental results demonstrate the coding performance of the proposed scheme at low bit rate, where it performs close to the joint encoding strategy.

***Index Terms***— 3D scene, sparse approximations, DSC

## 1. INTRODUCTION

Collecting visual information from the 3D environment is nowadays possible with networks of distributed cameras. Still, interpreting and compressing the acquired data represents a challenging task as the amount of data is typically huge. Moreover, in plenty of cases the communication among cameras is limited because of the bandwidth constraints or the introduced time delay. Development of distributed processing and compression techniques in the multi-camera setup thus becomes necessary in a variety of applications.

Distributed coding of multi-view images captured by camera networks recently attained great interest among researchers. Its information theoretical background was established already in the seventies by the Slepian-Wolf [1] and Wyner-Ziv [2] theorems, but its application to imaging problems has been delayed due to the difficulty of modeling correlations between real sources. The existing approaches are generally based on the disparity estimation between views under epipolar constraints. For example, in [3] cameras are divided into conventional cameras that perform independent image coding and Wyner-Ziv cameras that do distributed coding, while the joint decoder performs disparity estimation in order to decode Wyner-Ziv images. A geometrical approach for distributed coding is proposed in [4], where multi-view correlation is modeled by relating the locations of 1D linear boundaries in the 2D piecewise polynomial image representation. In previous work [5] we proposed a geometry-based correlation model between multi-view images that relates image projections of 3D scene features in different views, assuming that these features are correlated by local transforms, such as translation, rotation or scaling. We proposed to represent these features by sparse image expansion with geometric atoms taken from a redundant dictionary of functions. The correlation model is applied to the design of a distributed coding method with side information for multi-view omnidirectional images mapped to spherical images. Omnidirectional images are particulary convenient for scene representation due to their wide field of view. The Wyner-Ziv coder is designed by partitioning the redundant dictionary into cosets based on atom dissimilarity. The joint decoder uses the proposed correlation model to select the best candidate atom within the coset and to find corresponding features in two views. While the scheme has been shown to perform well at very low bit rate, it is penalized at higher rates due to occlusions and failures of decoding atoms that represent non-prominent scene features.

This paper presents a new shape correlation model, based on the distance between shape parameters of corresponding atoms. This model is able to identify similarity between highly anisotropic atoms after rotation, which permits improvements to the coset design and the performance of the joint decoder. We propose an occlusion-robust coding solution that is able to deal with decoding failures due to occlusions and non-prominent features. In sparse decompositions, atom occlusions or low-correlated features are represented as a subset of atoms, which are then erroneously decoded by the decoder. To correct these errors, we apply a channel code across all atoms, and send an additional syndrome to the decoder. The experimental results confirm that the new shape correlation model improves the coding performance, and that the occlusion-resilient coding corrects the saturation effect in the RD performance towards higher bit rate, where it becomes close to the joint encoding strategy.

## 2. CORRELATION MODEL BY SPARSE APPROXIMATIONS

The correlation model between multi-view images introduced in [5] relates image components that approximate the same 3D object in different views, by local transforms that include translation, rotation and anisotropic scaling. Given a redundant dictionary of atoms

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

$\mathcal{D} = \{\phi_k\}, k = 1, ..., N$, in the Hilbert space $H$, we say that image $y$ has a *sparse* representation in $\mathcal{D}$ if it can be approximated by a linear combination of a small number of vectors from $\mathcal{D}$. Therefore, sparse approximations of two[1] multi-view images can be expressed as $y_1 = \mathbf{\Phi_{I_1}} c_1 + \eta_1$ and $y_2 = \mathbf{\Phi_{I_2}} c_2 + \eta_2$, where $I_{1,2}$ labels the set of atoms $\{\phi_k\}_{k \in I_{1,2}}$ participating in the sparse representation, $\mathbf{\Phi_{I_{1,2}}}$ is a matrix composed of atoms $\phi_k$ as columns, and $\eta_{1,2}$ represents the approximation error. Since $y_1$ and $y_2$ capture the same 3D scene, their sparse approximations over the sets of atoms $I_1$ and $I_2$ are correlated. The geometric correlation model makes two main assumptions in order to relate the atoms in $I_1$ and $I_2$:

1. The most prominent (energetic) features in a 3D scene are present in sparse approximations of both images, with high probability. The projections of these features in images $y_1$ and $y_2$ are represented as subsets of atoms indexed by $J_1 \in I_1$ and $J_2 \in I_2$ respectively.
2. These atoms are correlated, possibly under some local geometric transforms. We denote by $F(\phi)$ the transform of an atom $\phi$ between two image decompositions that results from a viewpoint change.

Under these assumptions the correlation between the images is 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, remaining atoms in $I_2$, and noise $\eta_2$:

$$y_2 = \sum_{i \in J_1} c_{2,i} F_i(\phi_i) + \sum_{k \in I_2 \setminus J_2} c_{2,k} \phi_k + \eta_2. \quad (1)$$

The model from Eq. (1) is applied in [5] to atoms from the sparse decompositions of omnidirectional multi-view images mapped onto the sphere. The approach is based on the use of a structured redundant dictionary of atoms that are derived from a single waveform subjected to rotation, translation and scaling. More formally, given a generating function $g$ defined in $H$ (in the case of spherical images $g$ is defined on the 2-sphere), the dictionary $\mathcal{D} = \{\phi_k\} = \{g_\gamma\}_{\gamma \in \Gamma}$ is constructed by changing the atom index $\gamma \in \Gamma$ that defines rotation, translation and scaling parameters applied to the generating function $g$. This is equivalent to applying a unitary operator $U(\gamma)$ to the generating function $g$, i.e.: $g_\gamma = U(\gamma)g$. The main property of the structured dictionary is that it is transform-invariant, i.e., the transformation of an atom by any of the combination of translation, rotation and anisotropic scaling transforms results in another atom in the same dictionary. Let $\{g_\gamma\}_{\gamma \in \Gamma}$ and $\{h_\gamma\}_{\gamma \in \Gamma}$ respectively denote the set of functions used for the expansions of images $y_1$ and $y_2$. When the transform-invariant dictionary is used for both images, the transform of the atom $g_{\gamma_i}$ in image $y_1$ to the atom $h_{\gamma_j}$ in image $y_2$ reduces to a transform of its parameters, i.e., $h_{\gamma_j} = F(g_{\gamma_i}) = U(\gamma')g_{\gamma_i} = U(\gamma' \circ \gamma_i)g$. Due to the geometric constraints that exist in multi-view images, only a subset of all local transforms between $\{g_\gamma\}$ and $\{h_\gamma\}$ are feasible. This subset can be defined by identifying two constraints between corresponding atoms, namely the *epipolar* constraint and the *shape similarity* constraint. Given the atom $\{g_\gamma\}_{\gamma \in \Gamma}$ in image $y_1$ these two constraints give the subset of possible parameters $\Gamma_i \subseteq \Gamma$ of the correlated atom $\{h_{\gamma_j}\}$. Pairs of atoms that correspond to the same 3D points have to satisfy the epipolar constraints that represent one of the fundamental relations in multi-view analysis. Two corresponding atoms are said to match when their epipolar atom distance $d_{EA}(g_{\gamma_i}, h_{\gamma_j})$ is smaller than a certain threshold $\kappa$ (for more details on this distance we refer the reader to [5]). The set of possible candidate atoms in $y_2$, that respect epipolar constraints with the atom $g_{\gamma_i}$ in $y_1$, called the *epipolar candidates set*, is then defined as the set of indexes $\Gamma_i^E \subset \Gamma$, with:

$$\Gamma_i^E = \{\gamma_j | h_{\gamma_j} = U(\gamma')g_{\gamma_i}, d_{EA}(g_{\gamma_i}, h_{\gamma_j}) < \kappa\}. \quad (2)$$

[1]Two images are taken for the sake of clarity, but the correlation model can be generalized to any number of images.

## 3. SHAPE PARAMETERS CORRELATION MODEL

The shape similarity constraint assumes that the change of viewpoint on a 3D object results in a limited difference between shapes of corresponding atoms since they represent the same object in the scene. From the set of atom parameters $\gamma$, the last three parameters $(\psi, \alpha, \beta)$ describe the atom shape (its rotation and scaling), and they are thus taken into account for the shape similarity constraint. We propose to model the shape correlation based on the distance between atom shape parameters $\vec{v} = [\psi \ \ \alpha \ \ \beta]^T$ of two corresponding atoms $g_{\gamma_i}$ in $y_2$ and $h_{\gamma_j}$ in $y_2$. Camera movements result in parameter changes of atom $g_{\gamma_i}$ in the image decomposition that form a manifold in the space of shape parameters $(\psi, \alpha, \beta)$. Therefore, changing the viewpoint from camera 1 to camera 2 will induce the change of shape parameters from point $\vec{v}_i$ to point $\vec{v}_j$, which both belong to the manifold. The distance between shapes can then be defined by the manifold distance $d_m(\vec{v}_i, \vec{v}_j)$ as:

$$d_m(\vec{v}_i, \vec{v}_j) = \inf_{L_{ij}} \int_{L_{ij}} ||d\vec{v}||, \quad (3)$$

where $L_{ij}$ denotes a path on the parameters manifold from $\vec{v}_i$ to $\vec{v}_j$. Given this similarity measure, we define the *shape candidates set* as the set of atoms $h_{\gamma_j}$ in $y_2$ that are possible transformed versions of the atom $g_{\gamma_i}$ with respect to the shape similarity constraint, i.e.:

$$\Gamma_i^S = \{\gamma_j | h_{\gamma_j} = U(\gamma')g_{\gamma_i}, d_m(i,j) < t\}, \quad (4)$$

where $t$ is a chosen threshold. An alternative shape similarity measure is the inner product $\mu(i,j) = |\langle g_{\gamma_i}, h_{\gamma_j} \rangle|$ between atoms centered at the same position $(\tau, \nu)$, as proposed in [5]. The shape candidates set then includes atoms $h_{\gamma_j}$ such that $\mu(i,j) > s$, where $s$ is a chosen threshold. Inner product represents a general measure between two vectors in the image space and it is independent of the dictionary construction. However, this model is not able to identify the correlation between rotated highly anisotropic atoms (like edges). An example of two atoms that differ by a rotation of $11.25^\circ$ is shown on the Fig. 1. These atoms have similar shapes, but their inner product is very small (0.072) and the shape correlation model based on the inner product ($s$ is usually much higher) will classify them as uncorrelated atoms.

Finally, we combine the epipolar and shape similarity constraints to define the set of possible parameters of the transformed atom in $y_2$ as $\Gamma_i = \Gamma_i^E \cap \Gamma_i^S$.

## 4. OCCLUSION-RESILIENT WYNER-ZIV CODER

### 4.1. Wyner-Ziv coding

Based on the above geometric correlation model, we can build a Wyner-Ziv coding scheme for multi-view omnidirectional images.

**(a)** $g_{\gamma_1}$ **(b)** $g_{\gamma_2}$

**Fig. 1**. Rotation on anisotropic atoms, $\langle g_{\gamma_1}, g_{\gamma_2} \rangle = 0.072$

The coding scheme illustrated in Fig. 2 is mostly based on the algorithm proposed in [5]. However, the shape cosets and the joint decoding process use the novel shape similarity metric described in Sec. 3. In addition, an occlusion-resilient coding block (shaded area) is included to improve the coding performance, as described in the next section. The scheme is based on coding with side information, where image $y_1$ is independently encoded, while the Wyner-Ziv image $y_2$ is encoded by coset coding of atom indices and quantization of their coefficients. The approach is based on the observation that when atom $h_{\gamma_j}$ in the Wyner-Ziv image $y_2$ has its corresponding atom $g_{\gamma_{i_j}}$ in the reference image $y_1$, then $\gamma_j$ belongs to the subset $\Gamma_i = \Gamma_i^E \cap \Gamma_i^S$. Since $\Gamma_i$ is usually much smaller than $\Gamma$, the Wyner-Ziv encoder does not need to send the whole $\gamma_j$, but can transmit only the information that is necessary to identify the correct atom in the transform candidate set given by $\Gamma_i$. This is achieved by coset coding, by partitioning $\Gamma$ into distinct cosets that contain dissimilar atoms with respect to their position $(\tau, \nu)$ and shape $(\psi, \alpha, \beta)$. Two types of cosets were constructed: Position cosets, and Shape cosets. The encoder eventually sends only the indexes of the corresponding cosets for each atom (i.e., $k_n$ and $l_n$ in Fig. 2). The Position cosets are designed as VQ cosets [5], which are constructed by 2-dimensional interleaved uniform quantization of atom positions $(\tau, \nu)$ on a rectangular lattice. The new shape correlation model in Eq.(4) relies on the shape parameters $(\psi, \alpha, \beta)$, so we propose to design the Shape parameter cosets by 3-dimensional interleaved uniform vector quantization of shape parameters on a cubic lattice.

The decoder matches corresponding atoms in the reference image and atoms within the cosets of the Wyner-Ziv image decomposition using the correlation model described earlier. The atom pairing is facilitated by the use of quantized coefficients of atoms, which are sent directly. Each identified atom pair contains the information about the local transform between the reference and Wyner-Ziv image, which is exploited by the decoder to update the transform field between them. The transformation of the reference image with respect to the transform field provides an approximation of the Wyner-Ziv image that is used as side information for decoding the atoms without a correspondence in the reference image. These atoms are decoded based on the minimal mean square error between the currently decoded image and the side information. Finally, the WZ image reconstruction $\hat{y}_2$ is obtained as a linear combination of the decoded image $y_d$, reconstructed by decoded atoms from $\mathbf{\Phi_{I_2}}$, and the projection of the transformed reference image $y_{tr}$ to the orthogonal complement of $\mathbf{\Phi_{I_2}}$ [5].

### 4.2. Occlusion-resilient coding

The described decoding procedure does not, however, give any guarantee that all atoms are correctly decoded. In a general multi-view system, occlusions are quite probable, and clearly impair the atom pairing process at the decoder. The atoms which approximate these occlusions cannot be decoded based on the proposed correlation model since they do not have a corresponding feature in the reference image. Moreover, there are also features which are not sufficiently prominent to appear in sparse approximations of both views. Hence, the encoder needs to send additional information about these atoms for correct decoding. For example, it can send the index of the atom's parameters within the Shape and Position cosets, denoted respectively as $a_n$ and $b_n$ for the $n$-th atom in the sparse decomposition. Along with the coset indices $k_n$ and $l_n$, indices $a_n$ and $b_n$ uniquely define $\gamma_n$ in the set $\Gamma$. However, the main problem in the distributed setting is that the encoder does not know which of the atoms in the decomposition are occlusions and non-prominent features since there is no communication between separate encoders. Therefore, the encoder cannot send $a_n$ and $b_n$ only for the problematic atoms. Still, the encoder can make the assumption that at least $M$ out of total $N$ atoms in the sparse decomposition represent occlusions or non-prominent features that cause decoding failures, without necessarily knowing their position in the sparse decomposition. The decoder correctly decodes $N - M$ atoms, which leads to an error probability of $M/N$. Following this observation, we propose to modify the described Wyner-Ziv encoder by performing channel coding on all $a_n$ and $b_n$, $n = 1, ..., N$, together. The encoder thus sends a unique syndrome for all atoms, which is then used by the decoder to correct the erroneously decoded atom parameters. The modification of the Wyner-Ziv coder with the occlusion resilient block (shaded part) is shown in Fig. 2.

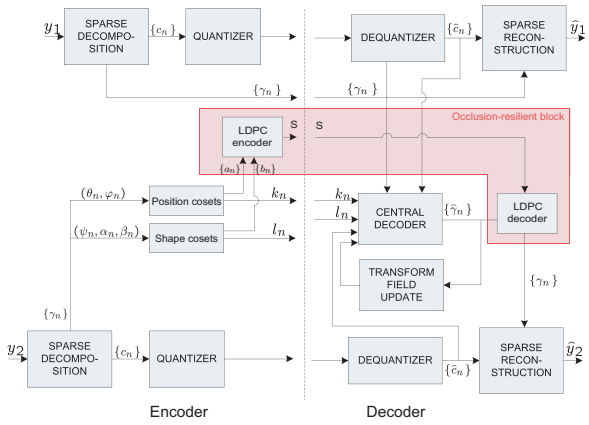


**Fig. 2**. Occlusion-resilient Wyner-Ziv coder

## 5. EXPERIMENTAL RESULTS

***Implementation details.*** The manifold distance given by Eq.(3) is usually hard to compute, since different parameters are not normalized. We therefore propose to calculate the Euclidean distance between parameters using their indices in the discrete dictionary as an approximation of the real manifold distance. When the parameters are sorted in a strictly ascending order, the Euclidean distance metric defined on the space of parameter indices will preserve the Euclidean distance relations between pairs of atom transforms in the parameter space itself.

***Numerical results.*** The performance of Wyner-Ziv coder with

the proposed shape correlation model and the performance of the occlusion-resilient Wyner-Ziv coder are compared to the method reported in [5]. The results are presented for the synthetic Room image set that consists of two $128 \times 128$ spherical images $y_1$ and $y_2$ (see Fig. 3). The sparse image decomposition is obtained using the Matching Pursuit (MP) algorithm on the sphere with the dictionary used in [5]. It is based on two generating functions: 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). The position parameters $\tau$ and $\nu$ can take 128 different values, while the rotation parameter uses 16 orientations. The scales are distributed logarithmically with 3 scales per octave. 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.

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

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

In the first experiment, we compare the performance of the Wyner-Ziv scheme for two shape correlation models, and thus two shape coset constructions. We will refer to the shape correlation model based on the inner product [5] as *model 1*, and to the model based on the parameters distance of Eq.(4) as *model 2*. Fig. 4 shows the rate-distortion curves for the Wyner-Ziv coding of image $y_2$, where the bit rate is swept by increasing the number of atoms. The dotted line corresponds to the model 1, and dash-dotted line to the model 2. In both cases we have used VQ Position cosets of size 1024, same as in [5]. For the shape cosets with the model 1, we have used 128 cosets, which corresponds to the values of shape parameters $s_G = 0.85$ (for Gaussian atoms) and $s_A = 0.75$ (for anisotropic atoms). In order to compare two shape models, we have selected the parameter $t$ such that $\Gamma_i^S$ includes the shape similarity set of the model 1. With 3D VQ coset design for the model 2 and $t = \sqrt{8}$, we have obtained 64 cosets. We can see that the model 2 outperforms the model 1, while requiring a smaller number of cosets and thus a smaller bitrate. DSC with either of the models outperforms the independent coding with MP (dashed line) and for low rates performs close to the joint encoding, where the encoder has access to the side information (red solid line). However, both curves saturate at higher rates and the effect of the new shape modeling becomes less important.

The second experiment examines the performance of the occlusion resilient Wyner-Ziv scheme, as explained in section 4.2, where LDPC codes are used for syndrome coding [6]. The rate of the LDPC code is chosen to be below the BSC channel capacity with the crossover probability $M/N$. We have used the same number of Position cosets and Shape cosets with model 2 as given above. The RD curve for the occlusion-resilient Wyner-Ziv scheme is given on the Fig. 4 with the blue solid line. We can clearly see that the occlusion resilient coding corrects the saturation behavior of the previous

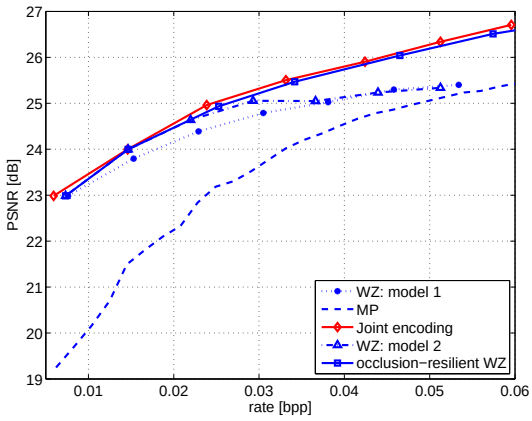


**Fig. 4**. Rate distortion performance for the image $y_2$

scheme, and performs close to the joint encoding scheme. The proposed method was also compared to independent coding of image $y_2$ using JPEG2000. Since JPEG2000 cannot perform at such low rates we have evaluated the rate at which it reaches the PSNR performance of the proposed DSC coder. To achieve the PSNR of 26.4 dB JPEG2000 uses 0.13 bpp, which is $\approx 2.4$ times greater than the rate the DSC coder needs to obtain the same image quality.

## 6. CONCLUSIONS

In this paper we have presented a distributed coding method for multi-view omnidirectional images, which is able to deal with occlusions or Wyner-Ziv decoding failures. Building upon a geometric correlation model and the distributed coding strategy in [5], we introduce a new shape correlation model between features in different views that improves the RD performance. Finally, we introduce an occlusion-resilient coding block that is able to correct the decoding failures due to occlusions or non-prominent features. This permits us to reach coding performances that are close to that of a joint encoding scheme for at low bit rates.

## 7. REFERENCES


[1] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inform. Theory*, vol. 19, no. 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, no. 1, pp. 1–10, January 1976.

[3] X. Zhu, A. Aaron and B. Girod, "Distributed compression for large camera arrays," in *Proceedings of IEEE SSP*, September 2003.

[4] N. Gehrig, P. L. Dragotti, "Distributed Compression of Multi-View Images using a Geometrical Coding Approach," in *Proceedings of IEEE ICIP*, September 2007, vol. 6, pp. VI –421–VI –424.

[5] I. Tosic and P. Frossard, "Geometry-based distributed scene representation with omnidirectional vision sensors," *IEEE Trans. on Image Processing*, vol. 17, no. 7, pp. 1033–1046, July 2008.

[6] "Methods for constructing ldpc codes: Available in url http://www.cs.utoronto.ca/pub/radford/ldpc-2001-05-04/pchk.html.," .