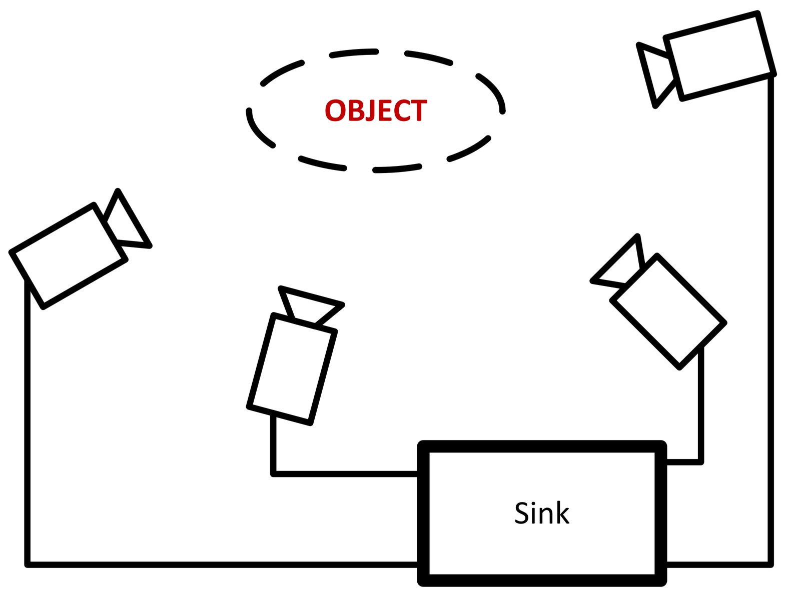
MULTI-VIEW DISTRIBUTED SOURCE CODING OF BINARY FEATURES FOR VISUAL SENSOR NETWORKS

Nuno Monteiro, Catarina Brites, Fernando Pereira, João Ascenso
Instituto de Telecomunicações — Instituto Superior Técnico

1. Context

⇒ **Visual Sensor Networks** (VSNs):
- → **Sensing Nodes**: Visual data is acquired and features are extracted.
- → **Sink Node**: Features are gathered and analyzed.
- → Strict constraints in computational power, energy and bandwidth at the sensing nodes.



2. Objectives and Solution

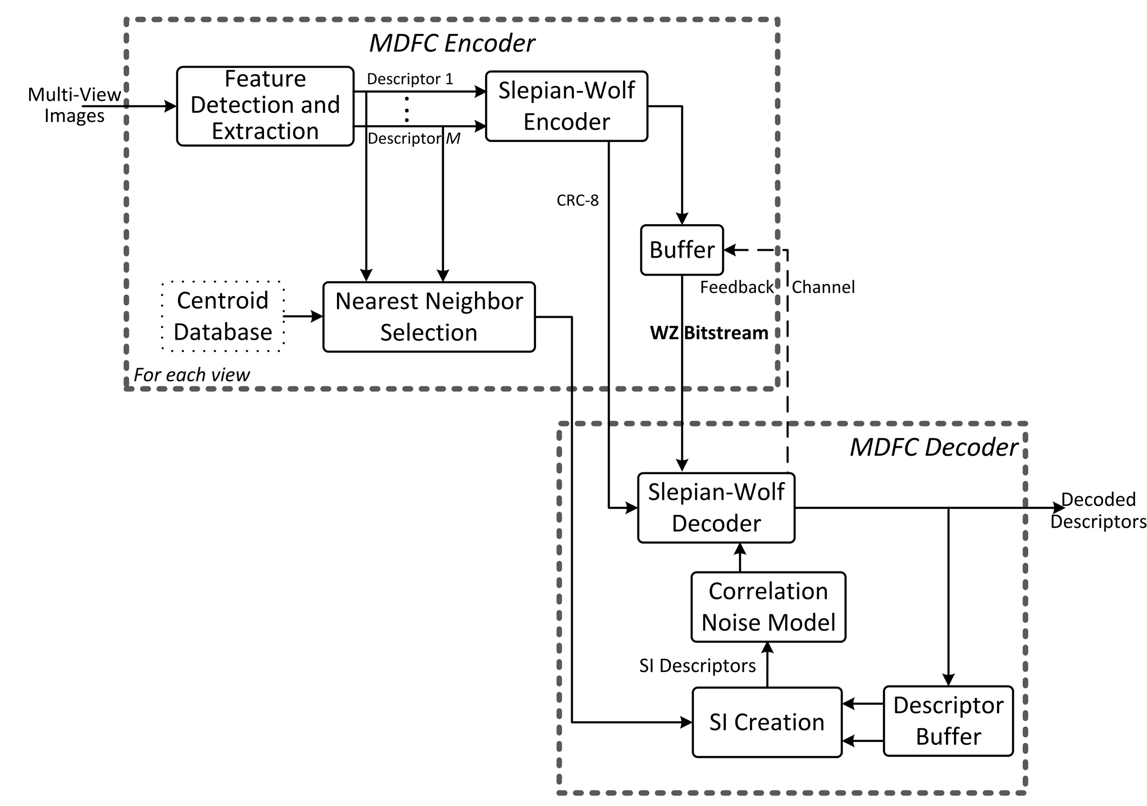
⇒ **Multi-view Distributed Feature Codec** (MDFC) **Objectives**:
- → Exploit the correlation between features extracted from overlapped views of the same scene.
- → Propose coding techniques with minimal routing overhead, that work under severe bandwidth restrictions and that are parsimonious in terms of computational resources.

⇒ **MDFC Solution:**
- → Improve the coding efficiency of binary features by exploiting multiple Side Information (SI) hypotheses in the Iterative Slepian-Wolf decoding process (**Turbo** and **LDPC**):
  1. **Multiple Inter-view SI creation step:** Several SI hypotheses are constructed by exploiting spatial correlation between different views.
  2. **Intra-view SI creation method:** Works in parallel with the Inter-view mode to decode independent features that are not highly correlated with the other views of the same scene.

⇒ **Novel contribution:**
- → Understand how Inter-view correlation can be exploited to obtain SI with higher accuracy.

3. Proposed Architecture

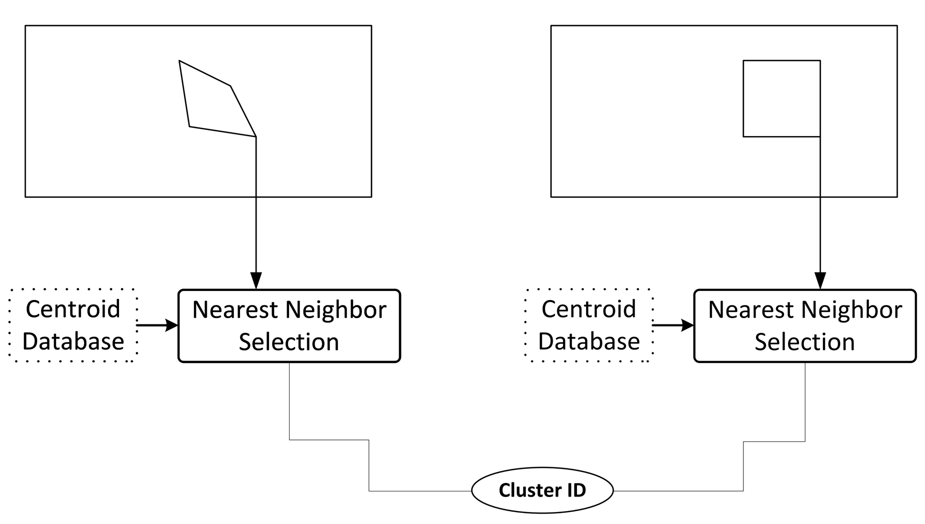
⇒ **Based on the DISCOVER codec used for (pixel-based) mono-view distributed video coding.**



4. Inter-View Side Information (SI) Creation

⇒ **To exploit the spatial redundancy between views it is necessary to decide which of the previously decoded descriptors is correlated with the descriptor being decoded.**

4.1. Centroid Based Strategy (CBS)
- → Descriptors are assigned to a cluster.
- → The same features, extracted from different views are expected to be represented by similar descriptors (same cluster ID).



4.2. Geometry Based Strategy (GBS)
⇒ Exploits the geometric position of the extracted descriptors:
1. **Centroid Matching:** Centroid ID is used to identify a set of similar descriptors for each of the new descriptors being decoded.
2. **Affine Model Estimation:** Search for an affine model between the view being decoded and each reference view.
3. **Warping:** The homographic transformation is used to choose which descriptors shall be used as SI when trying to decode the new view descriptors.

5. Correlation Noise Model

⇒ **Motivation**:
- → A reliable model, that characterizes the correlation noise between the original descriptor and the SI descriptors, is needed.
- → Descriptors are binary memoryless sources where symbols ('0' and '1') have the same probability of occurrence.
- → SI descriptors corresponds to the set of already decoded descriptors that are highly correlated with the source.

⇒ **Binary Symmetric Channel** (BSC):
- → If a centroid's descriptor symbol is set to '0': The probability of the encoded symbol being '0' (P_-) is equal to the number of times, that same symbol, is set to '0' (N) divided by the number of descriptors (M) in the set of SI descriptors set. Equivalent for '1' (P_+).

$$P_- = p(B_n = 0 \mid Y_n^0, \dots, Y_n^M) = \frac{N}{M}$$
$$P_+ = p(B_n = 1 \mid Y_n^0, \dots, Y_n^M) = 1 - p(B_n = 0 \mid Y_n^0, \dots, Y_n^M)$$

6. Performance Evaluation

6.1. Test Conditions:
⇒ **Clusters**: 4096 centroids of 512 bits each.

⇒ **Unsupervised Learning**: 12456 images from Paris, Stanford landmarks and Oxford datasets.

⇒ **Keypoint Detector**: SURF

⇒ **Feature Extractor**: BRISK

⇒ **Reference Dataset**: Berkley Multiview Wireless (BMW)
- → 16 perspectives with 5 images per perspective.
- → Perspectives **0**, **3**, **6**, **9** and **12** are used as **queries.**
- → All the other images, from other prespectives, are used for the **database**.

⇒ **Predictive Features Codec** (PFC):
- → **Arithmetic** encoding of the residue between extracted descriptor and nearest centroid.

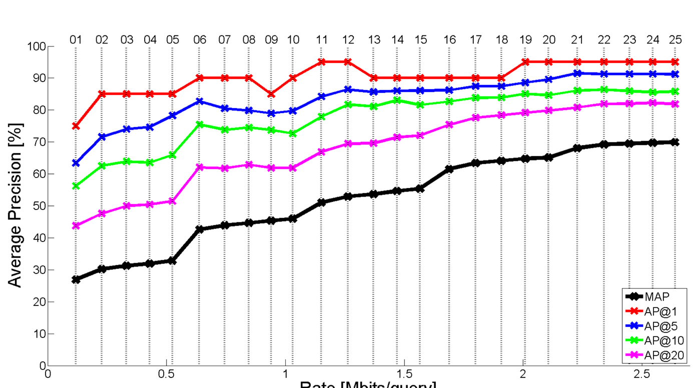
6.2. Experimental Results:
⇒ **Independent Encoding bitrates**: The bitrate compression acheived with 'PFC' slightly outperforms 'MDFC—LDPC'. Both outperform 'MDFC—Turbo' by 7 percentual points.

⇒ **Bitrate Reduction by using reference views**: 'MDFC—LDPC' outperforms 'PFC' when using 1 reference view. 'MDFC—Turbo' needs more views.

⇒ **Rate-accuracy:** Average Precision (AP) metrics show an improvement when using more view-points of the same object.

PFC and MDFC average Bitrate Reduction [%]

	Intra	CBS			GBS
Ref. Views	0	1	4	79	79
PFC	23.97				
MDFC - Turbo	16.41	22.47	27.57	28.44	33.23
MDFC - LDPC	23.04	28.05	32.45	33.23	37.36



'MDFC—LDPC' rate-accuracy performance

7. Conclusions and Future Work

⇒ Significant **bitrate savings** were obtained by exploiting Inter-view redundancy at decoder side.

⇒ **Accuracy** of the object recognition task **improves** by using more cameras (MAP goes from 30% to 70%).

⇒ A future improvement can be the design of a **selection coding scheme**, which prevents redundant features from being transmitted.

GreenEyes
Networked energy-aware visual analysis