Exploiting Sparse Representations for Robust Analysis of Noisy Complex Video Scenes

Gloria Zen[1], Elisa Ricci[2], Nicu Sebe[1]

[1]{zen,sebe}@disi.unitn.it [2]elisa.ricci@diei.unipg.it

UNIVERSITA' DEGLI STUDI DI PERUGIA

UNIVERSITÀ DEGLI STUDI DI TRENTO

Contributions

- We formalize the task of **extracting activity patterns** as a **non-negative matrix factorization (NMF)** problem, considering as reconstruction function the robust **Earth Mover's Distance (EMD)** and imposing **sparsity constraints**.
- We derive an **alternating optimization approach** to solve the proposed problem efficiently and we show that it is reduced to a sequence of linear programs (LP).
- With respect to previous work on EMD clustering [1, 6], our method is **more scalable** and produces **more interpretable results** due to the sparsity constraints.

Efficient Earth Mover's Distance (EMD-L_1)

The $\mathcal{D}_{EMD}(\boldsymbol{h},\boldsymbol{p})$ is obtained as the solution of the transportation problem:

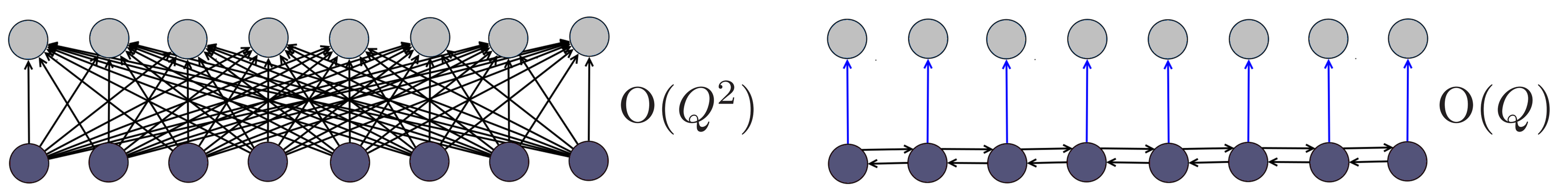
$$\min_{f_{qt}\geq 0} \sum_{t,q=1}^{Q} d_{qt} f_{qt} \quad \text{s.t.} \quad \sum_{q=1}^{Q} f_{qt} = h^t, \quad \sum_{t=1}^{Q} f_{qt} = p^q$$

By using **EMD with** L_1 as **ground distance** [2], this simplifies as:

$$\min_{f_{qt}\geq 0} \sum_q \sum_{t\in\mathcal{N}(q)} f_{q,t}$$

$$\text{s.t.} \quad \sum_{t\in\mathcal{N}(q)} f_{q,t} - \sum_{t\in\mathcal{N}(q)} f_{t,q} = h^q - p^q \;\; \forall q$$

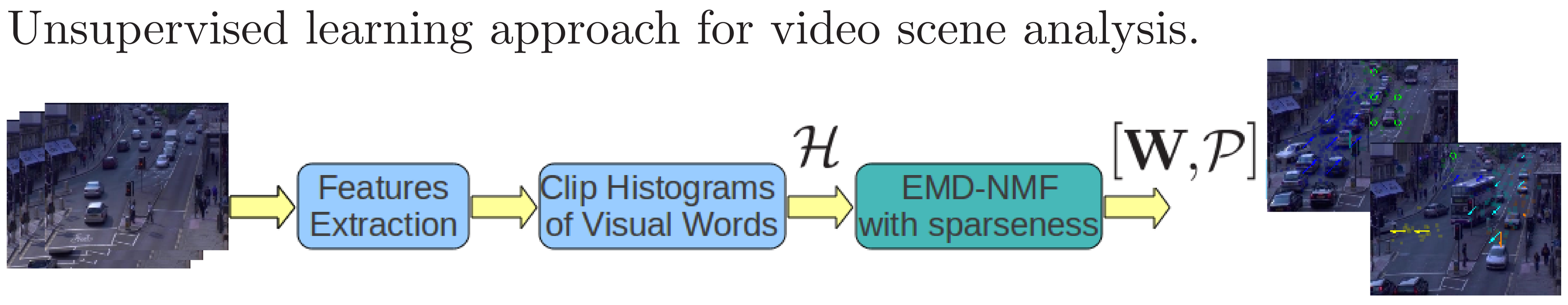
$\mathcal{N}(q)$ is represented by the adjacent bins.



Complexity reduces from O(Q^2) to O(Q).

Our Approach

Unsupervised learning approach for video scene analysis.



Features Extraction

original frame | foreground mask | trajectons | features

EMD NMF with Sparseness Constraints

Given a set of clip histograms $\mathcal{H} = \{\boldsymbol{h}_1,\ \boldsymbol{h}_2, \ldots \boldsymbol{h}_N\}$, discovering high-level activities is modeled as finding a set of bases $\mathcal{P} = \{\boldsymbol{p}^1,\ \boldsymbol{p}^2,\ \ldots \boldsymbol{p}^K\}$, $K << N$, and the coefficients $\mathbf{W} = [\boldsymbol{w}_1\ \boldsymbol{w}_2 \ldots \boldsymbol{w}_N]$, $\boldsymbol{w}_i \in \mathbb{R}^K$, such that:

$$\min_{\boldsymbol{p}_k,\mathbf{W}\geq 0} \sum_{i=1}^{N} \mathcal{D}_{EMD}(\boldsymbol{h}_i, \sum_k w_i^k \boldsymbol{p}^k) \tag{1}$$

$$\text{s.t.} \quad \omega_m \leq \Omega(\boldsymbol{p}^k) \leq \omega_M, \quad \forall\, k = 1\ldots K \tag{2}$$

Choosing $\Omega(\mathbf{x}) = \frac{1}{\sqrt{n-1}}(\sqrt{n} - \|\mathbf{x}\|_1/\|\mathbf{x}\|_2)$, $\mathbf{x} \in \mathbb{R}^N$, (2) enforces sparsity. The resulting optimization problem is:

$$\min_{p_q^k, w_i^k, f_{q,t}^i \geq 0} \sum_{i=1}^{N} \sum_q \sum_{t\in\mathcal{N}(q)} f_{q,t}^i \tag{3}$$

$$\text{s.t.} \quad \sum_{t\in\mathcal{N}(q)} f_{q,t}^i - \sum_{t\in\mathcal{N}(q)} f_{t,q}^i = h_q^i - \sum_k w_i^k p_q^k,$$

$$\sum_k w_i^k = 1, \qquad \sum_q p_q^k = 1, \quad \forall q, \forall i, \forall k$$

$$\|\boldsymbol{p}^k\|_2, \leq \frac{1}{c_M}\mathbf{e}^T\boldsymbol{p}^k \qquad \frac{1}{c_m}\mathbf{e}^T\boldsymbol{p}^k \leq \|\boldsymbol{p}^k\|_2 \quad \forall k \tag{4}$$

with $c_x = \sqrt{Q} - \omega_x(\sqrt{Q} - 1)$, $\mathbf{e} \in \mathbb{R}^Q$ is a vector of ones.

Alternating Optimization

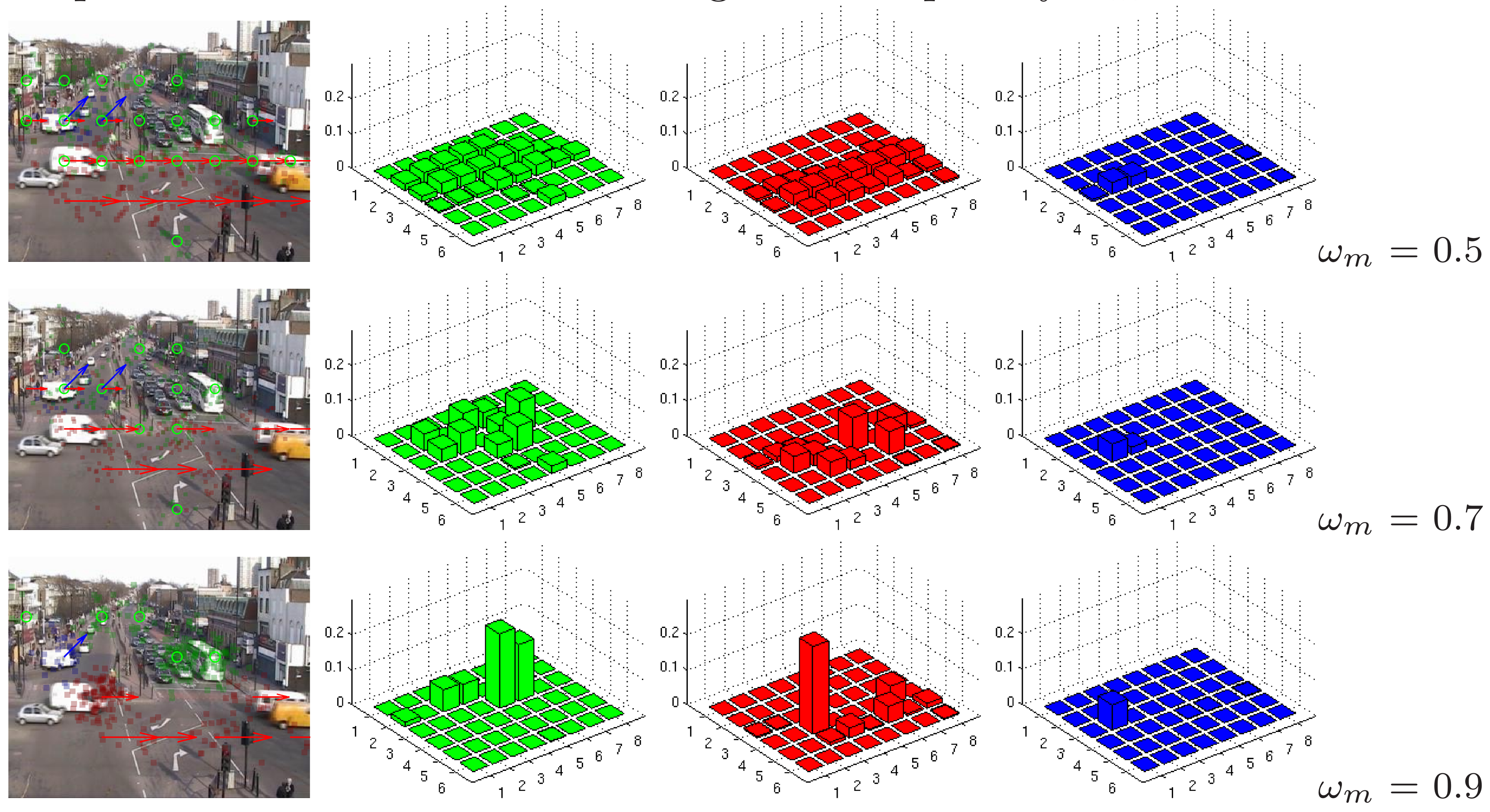
The EMD learning problem is solved using alternating optimization:

Input: Original clips histograms $\mathcal{H} = \{\boldsymbol{h}_1,\ \boldsymbol{h}_2, \ldots \boldsymbol{h}_N\}$.
  Initialize $\mathbf{W} = [\boldsymbol{w}_1\ \boldsymbol{w}_2 \ldots \boldsymbol{w}_N]$ with positive random values.
  Normalize the columns of $\mathbf{W}$ such that $\sum_k w_i^k = 1$, $\forall i = 1, \ldots N$.
  while not converged
    Solve (3) s.t. (4) with respect to $\boldsymbol{p}^k, \boldsymbol{f}$ using TPC method.
    Solve the LP (3) with respect to $\mathbf{W}, \boldsymbol{f}$.
  end
Output: $\mathbf{W}$, $\mathcal{P} = \{\boldsymbol{p}^1,\ \boldsymbol{p}^2,\ \ldots \boldsymbol{p}^K\}$.

The **Tangent Plane Constraint** (**TPC**) method [3] consists in approximating the non convex cone constraints by linear constraints.

Results - High Level Activities Extraction

Computed basis matrix at increasing level of sparsity ω_m:



Nice grouping effect of combining EMD-L_1 with sparsity constraints!

Results - Anomaly detection

(Left)Anomaly score and temporal segmentation bar for a 40 clips video sequence.
(Right) Representative frames are extracted from clips detected as anomalous.

Results - Quantitative Evaluation

Clustering accuracy at varying level of sparsity ω_m:

ω_m	0	0.1	0.3	0.5	0.7	0.9
Junction2 (48 clips)	89.58%	89.58%	89.58%	89.58%	**91.67%**	89.58%
Roundabout (60 clips)	88.33%	88.33%	**90.00%**	**90.00%**	**90.00%**	**90.00%**
Junction (39 clips)	**89.74%**	**89.74%**	**89.74%**	**89.74%**	**89.74%**	84.62%

Comparison with previous approaches:

	std pLSA [4]	hrc pLSA [4]	DDP-HMM [5]	EMP [6]	our approach ($\omega_m = 0.7$)
Roundabout (60 clips)	81.67%	75.00%	85.00%	86.67%	**90.00%**
Junction	89.74%	76.92%	87.18%	**92.31%**	89.74%
Roundabout (148 clips)	84.46%	72.30%	85.14%	**86.40%**	*85.81%*

Datasets are publicly available[1]. Code is available at our website[2].

References

[1] http://www.eecs.qmul.ac.uk/~ jianli/Dataset_List.html
[2] http://disi.unitn.it/~ zen/sparse_emd.html

[1] E. Ricci, G. Zen, N. Sebe, S. Messelodi. A Prototype Learning Framework using EMD: Application to Complex Scenes Analysis In *PAMI*, 2012.

[2] H. Ling, K. Okada. An efficient Earth Mover Distance algorithm for robust histogram comparison In *PAMI*, 2006.

[3] H. Tuy. Convex programs with an additional reverse convex constraint J. of Optim. Theory and Applic, 1987

[4] J. Li, S. Gong, T. Xiang. Global behaviour inference using probabilistic latent semantic analysis In *BMVC*, 2008.

[5] D. Kuettel, M.D. Breitenstein, L.V. Gool, V. Ferrari. What's going on? Discovering spatio-temporal dependencies in dynamic scenes In *CVPR*, 2010.

[6] G. Zen, E. Ricci. Earth mover's prototypes: a convex learning approach for discovering activity patterns in dynamic scenes. In *CVPR*, 2011.