Queen Mary University of London

School of EECS,
Queen Mary, University of London,
London E1 4NS, UK

Comparing Visual Feature Coding for Learning Disjoint Camera Dependencies

Xiatian Zhu, Shaogang Gong and Chen Change Loy

Queen Mary Vision Laboratory

{xiatian.zhu, sgg}@eecs.qmul.ac.uk
{ccloy}@visionsemantics.com

Problem

Given two disjoint camera views, we wish to estimate:
(1) their inter-camera correlation,
(2) and their spatial-temporal dependencies.
Moreover, we aim to answer the question:
What visual representations are more effective?

Motivation

- ★ Overcome the unreliability of manually selecting visual features from specific datasets;
- ★ Explore high-level structural constraints in coding low-level features for associating objects entities (supervised);
- ★ Employ co-occurrence statistics for constructing more reliable representations (unsupervised).

Contributions

(1) A systematic investigation into the effectiveness of supervised versus unsupervised feature coding methods for learning inter-camera dependencies;
(2) Evaluation of the sensitivity of learning inter-camera time correlation to the size of training data and the quality of scene region decomposition.

Methodology

(i) **Supervised method**: Random Forest (RF) [1] for supervised feature coding;
(ii) **Unsupervised method**: Latent Dirichlet Allocation (LDA) [2] for mapping low-level features to code-words that capture topic distributions;

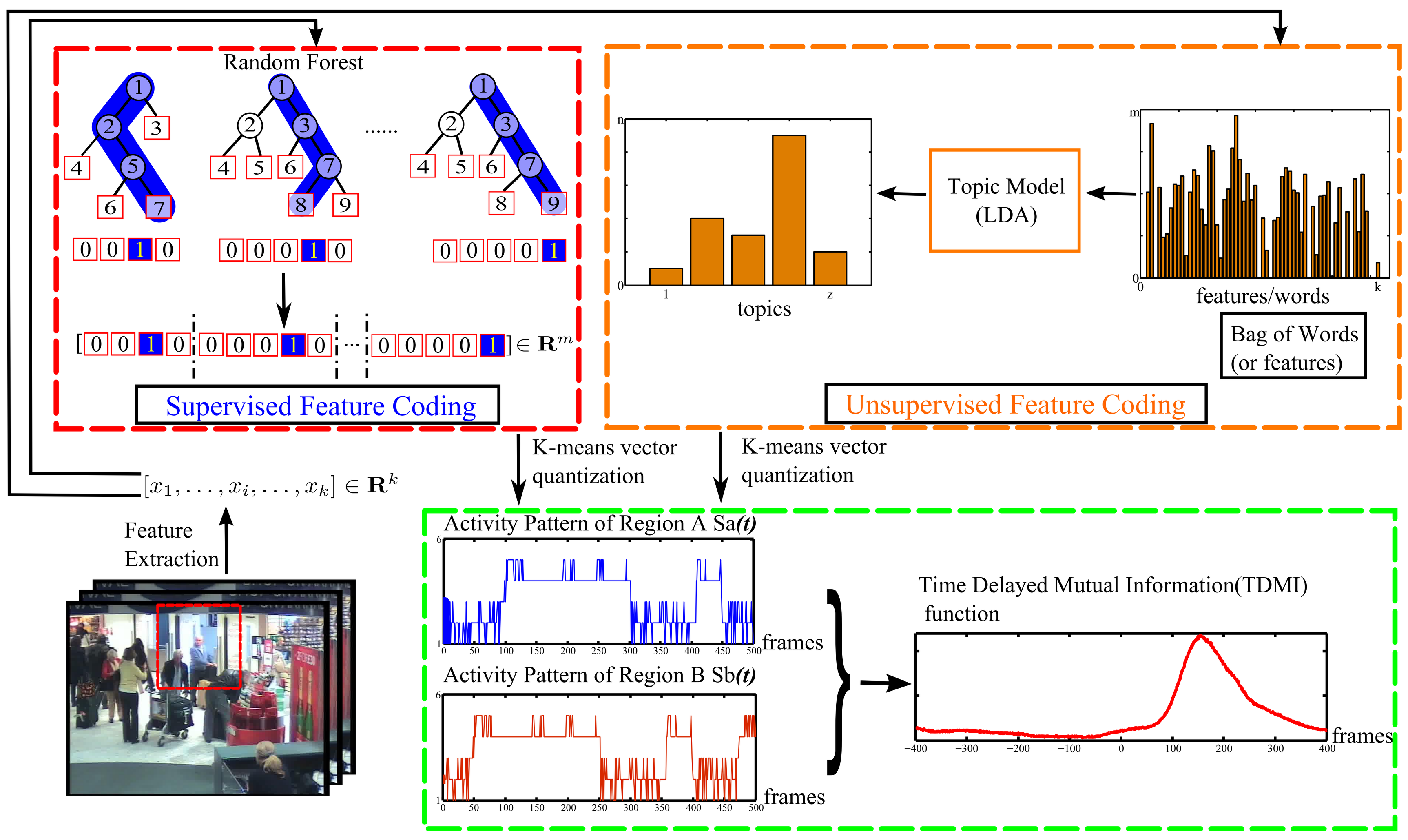

Figure 1: An overview of feature coding comparison for learning inter-camera dependencies.

(iii) **Time Delayed Dependency Inference**: Time Delayed Mutual Information (TDMI) [3] for learning inter-camera dependencies with the aforementioned feature codes;
(iv) A new metrics called **Mutual Information Margin(MIM)** proposed for evaluating different feature coding methods:

$$\triangle\mathcal{I} = \frac{\delta(\mathcal{I}_{\text{con}}) - \delta(\mathcal{I}_{\text{uncon}})}{\delta(\mathcal{I}_{\text{con}})}, \delta(\mathcal{I}) = \max(\mathcal{I}) - \min(\mathcal{I}), \quad (1)$$

where $\mathcal{I}_{\text{con}}$ and $\mathcal{I}_{\text{uncon}}$ denote the TDMI function yielded by the connected and unconnected pairs of regions.

Experiments

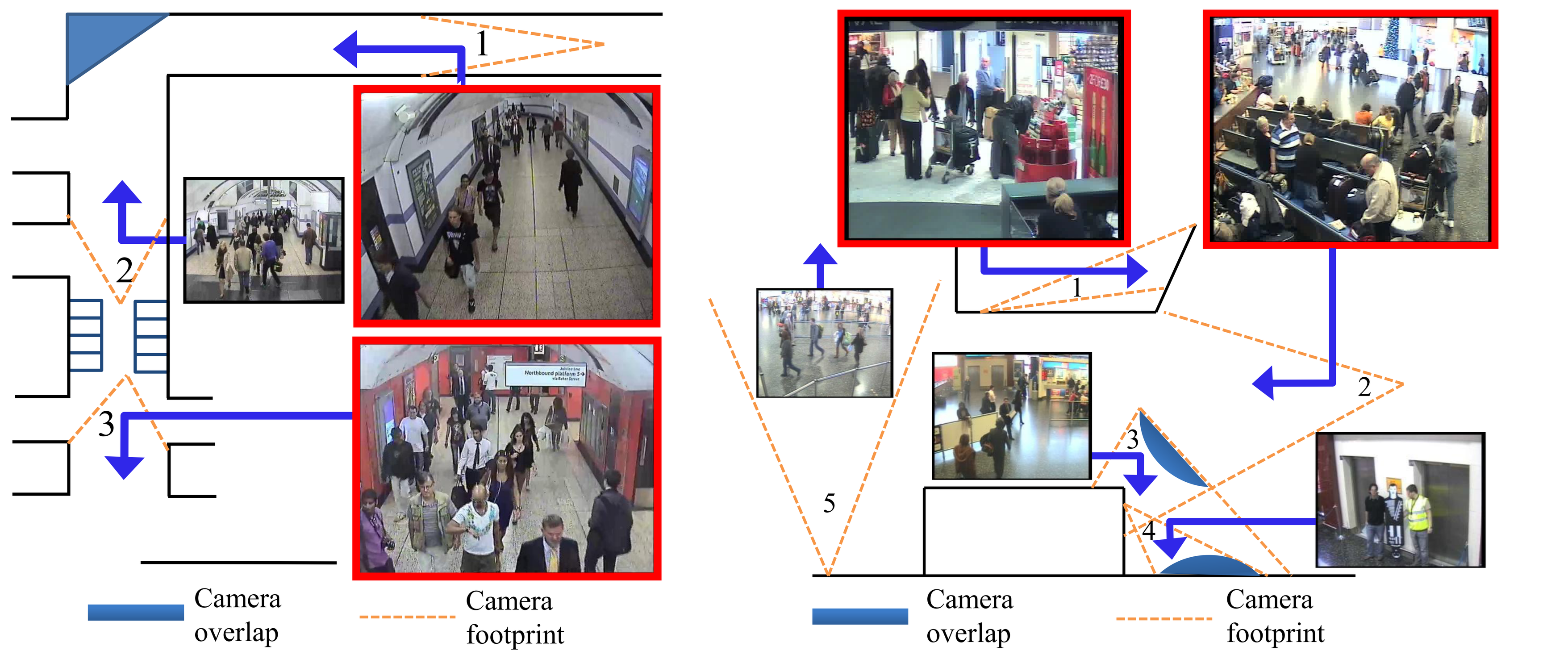

Figure 2: The layout and example views of an Underground Station (US) dataset (left) and the i-LIDS (right) dataset.

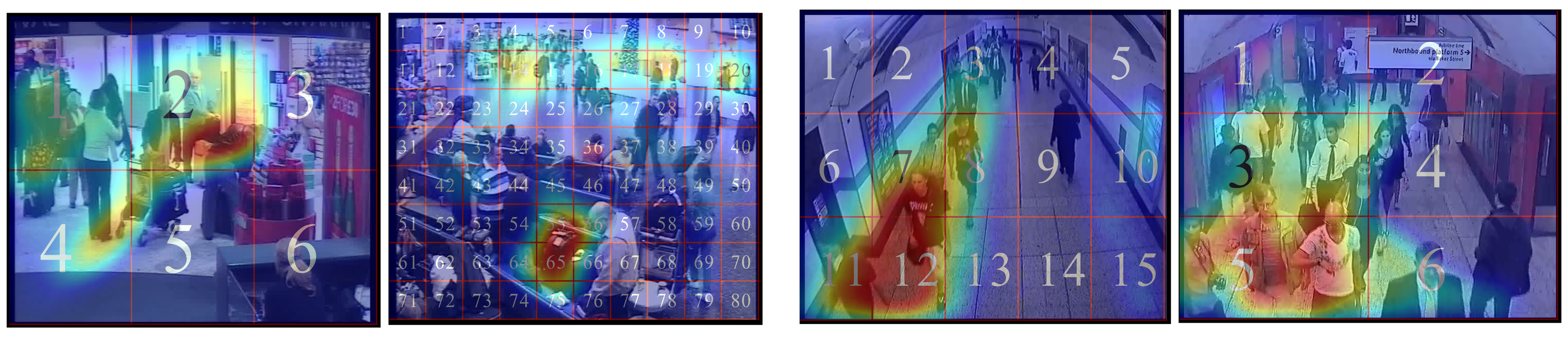

Figure 3: Motion Saliency Maps obtained on the US and the i-LIDS datasets. The selected regions are labelled by black digits.

Feature Codings	MI-MIM (US)	MI-MIM (i-LIDS)
RF pred	5.1530	7.8577
tree code	-1.7979	-1.7847
RF pred + tree code	-2.3839	-1.0335
topic code	**9.9057**	**16.6349**

Table 1: **Sensitivity to the length of the training sequence**: the average improvement in MIM of different feature coding methods over the k-means vector quantisation based representation. Mean improved MIM (MI-MIM) was computed by averaging individual percentage of improvement over the testing range.

Feature Codings	MI-MIM (US)	MI-MIM (i-LIDS)
RF pred	10.7670	**13.1541**
tree code	7.8714	2.0040
RF pred + tree code	7.6564	3.5522
topic code	**14.3076**	4.1265

Table 2: **Sensitivity to region decomposition:** Mean Improved MIM was computed following the same steps as explained in Table 1.

Experiment 1: sensitiveness to the size of training data
(1) Topic code gave the most favourable results (see Table 1);
(2) Suggest that feature coding methods can suppress noisy dependencies between unconnected region pairs.

Experiment 2: sensitiveness to the quality of region decomposition
(1) Topic code shows the best performance for the US dataset while RF pred for the i-LIDS dataset (see Table 2);
(2) Suggest that person count and topic clusters can be useful cues for inter-camera dependency learning.

Conclusion:
(1) Investigate the effectiveness of supervised (RF) and unsupervised (LDA) feature coding methods for learning inter-camera correlations;
(2) RF and LDA coding schemes outperform the k-means vector quantisation in robustness to small training data size;
(3) The coded features are more reliable to poor scene region decomposition;
(4) Feature coding can suppress noisy dependencies while capture inherent correlations between camera views.

References

[1] Breiman. Machine Learning, 45(1):5–32, 2001.
[2] Blei, Ng, Jordan. J. Machine Learning Research, 3:993–1022, 2003.
[3] Loy, Xiang, Gong. IEEE Trans PAMI, 34(9):1799-1813, 2012.