# Video semantic clustering with sparse and incomplete tags

*Published in:* The Thirtieth AAAI Conference on Artificial Intelligence, 2016

Authors: Jingya Wang, Xiatian Zhu, Shaogang Gong

Centre for Intelligent Sensing
Queen Mary University of London

CIS centre for intelligent sensing

Queen Mary University of London

# Introduction

Visual Information + Textual Information


• • •


How to cluster tagged videos into semantic groups ?

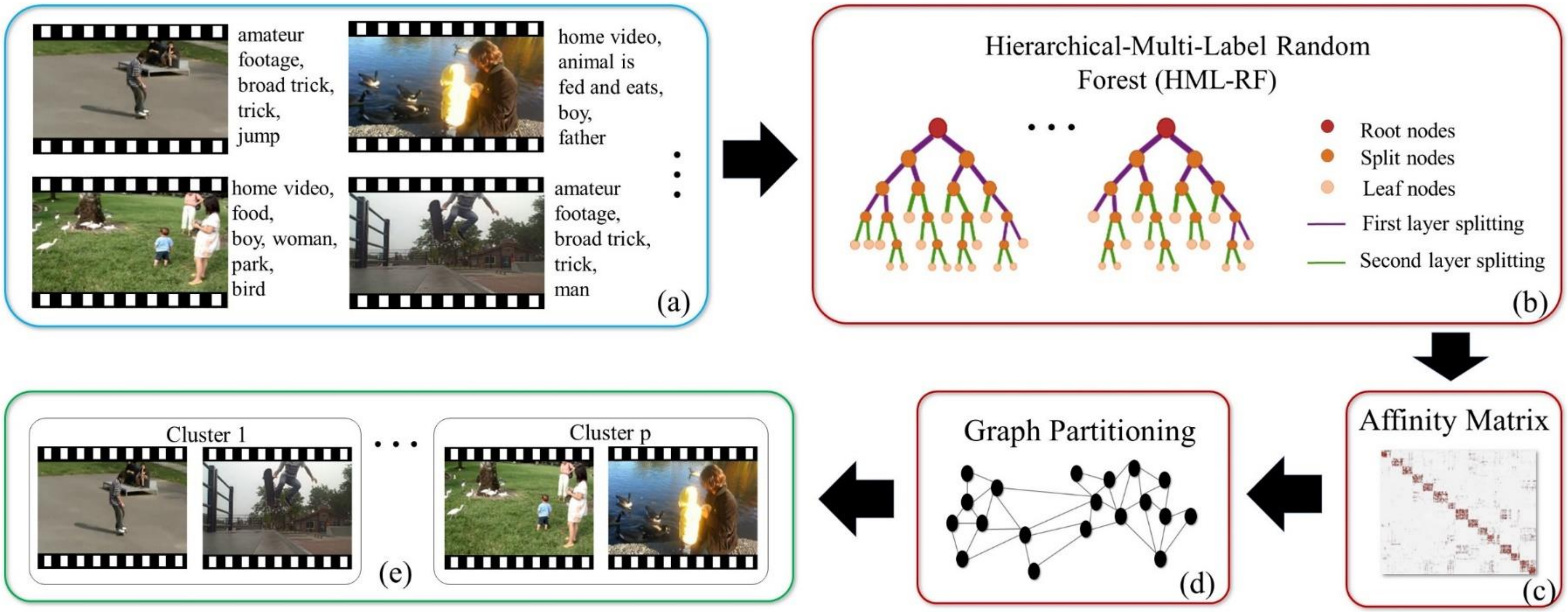
**Difficulty:**

- Visual and textual data significantly differ in representation and distribution
- Visual features: ambiguous and unreliable
- Tag: inherently sparse and incomplete

## Introduction

- Existing methods
  - Tags are organised and used in a flat structure
  - Tag statistical correlations are not fully exploited
- Contributions
  - A novel tag-based video clustering method capable of effectively fusing information from ambiguous visual features and sparse textual tags
  - A unified tag correlation based algorithm to cope with tag sparseness/incompleteness

# Methods




➢ **Limitations of existing random forests**

× Conventional clustering forest:
(a) fully concatenated representation
(b) two modalities are not balanced

× Conventional classification forest:
(a) supervised

➢ **Merits of the proposed HML-RF model**

✓ A hybrid model of classification and clustering forests

✓ Jointly exploiting heterogeneous visual and tag data

✓ Integrating abstract-to-specific tag topology

✓ Handling tag sparseness and incompleteness

# Methods

## Hierarchical-Multi-Label Random Forest (HML-RF)

$$\Delta\psi_{\mathrm{sl}} = \psi_s - \frac{|L|}{|S|}\psi_l - \frac{|R|}{|S|}\psi_r,$$

Conventional individual information gain

where $L$ and $R$ denote the data set routed into the left $l$ and right $r$ children,
The uncertainty $\psi$ over the tag distribution was computed as Gini impurity

$$\Delta\psi_{\mathrm{ml}} = \sum_{i=1}^{m} \Delta\psi_{\mathrm{sl}}^{i}$$

Multi-tags information gain

where $\Delta\psi_{\mathrm{sl}}^{i}$ is individual information gain computed in the $i$-th tag.

$$\Delta\psi_{\mathrm{hml}} = \sum_{k=1}^{\mu} \left( \prod_{j=1}^{k-1} (1-\alpha_j)\alpha_k \sum_{i\in Z_k} \Delta\psi_{\mathrm{sl}}^{i} \right)$$

Adaptive hierarchical multi-label information gain

where $Z_k$ denotes the tag index set of the $k$-th layer in the tag hierarchy (totally $\mu$ layers), with $\cup_{k=1}^{\mu} Z_k = Z$, and $\forall_{j\neq k} Z_j \cap Z_k = \emptyset$. Binary flag $\alpha_k \in \{0,1\}$ indicates the impurity of the $k$-th tag layer, $k \in \{1,\ldots,\mu\}$, i.e. $\alpha_k = 0$ when tag values are identical/pure across all the training samples $S$ of split node $s$ in any tag $i \in Z_k$, $\alpha_k = 1$ otherwise. The target layer is $k$ in case that $\alpha_k = 1$ and $\forall\alpha_j = 0, 0 < j < k$.

# Methods

## ➢ Handling tag sparseness and incompleteness

**Idea:** Estimate soft tag labels (real valued) with two tag statistical correlations mined from available tag data

(I) Co-occurrence $\varrho_{i,j} = co_{i,j}/o_j,$

where $co_{i,j}$ denotes the co-occurrence frequency of tags $i$ and $j$.$o_j$ denotes the number of occurrences of tag $j$ over all samples.

once $\varrho_{i,j}$ is obtained, for a potentially missing tag $i \in Z_k$ of $\mathring{\mathbf{x}} \in S_{\text{miss}}$, we estimate its positive score $\hat{y}^{+}_{\cdot,i}$ via:

$$\hat{y}^{+}_{\cdot,i} = \sum_{j \in \{Z_{k+1},\ldots,Z_\mu\}} \varrho_{i,j} y_{\cdot,j}$$

where $y_{\cdot,j}$ refers to the $j$-th tag value of $\mathring{\mathbf{x}}$.

(II) Mutual-exclusion $\epsilon_{i,j} = \max(0, r^{-+}_{i,j} - r^{-}_{i})/(1 - r^{-}_{i}),$

where $r^{-}_{i}$ refers to negative sample percentage on tag $i$ across all samples, and $r^{-+}_{i,j}$ negative sample percentage on tag $i$ over samples with positive tag $j$.

we predict the negative score $\hat{y}^{-}_{\cdot,i}$ for $\mathbf{x}$ on tag $i$ with:

$$\hat{y}^{-}_{\cdot,i} = \sum_{j \in \{Z_{k+1},\ldots,Z_\mu\}} \epsilon_{i,j} y_{\cdot,j},$$

# Evaluations

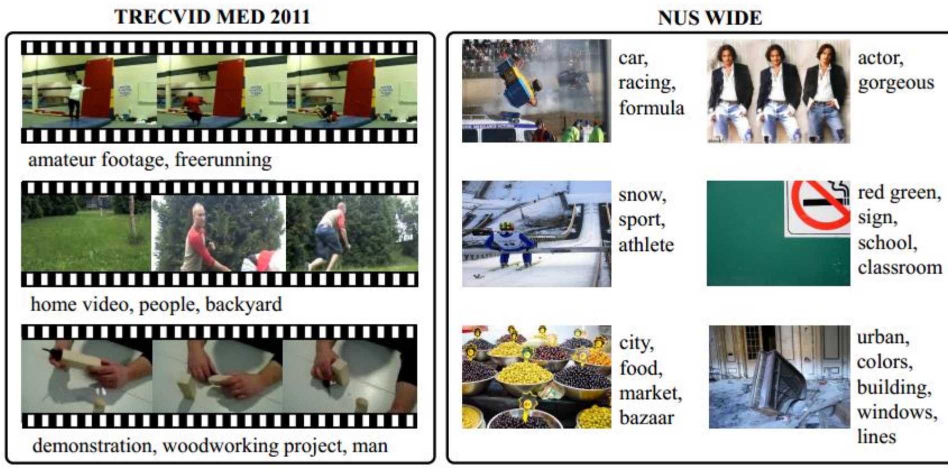



Figure 3: Examples from the TRECVID MED 2011 [56] and NUS-WIDE [9] dataset.

Table 1: Comparing clustering methods on TRECVID MED 2011 [56].

| Input mode | Method | Purity | NMI | RI | F1 | ARI |
|---|---|---|---|---|---|---|
| ViFeat | K-means[2] | 0.26 | 0.19 | 0.88 | 0.14 | 0.08 |
| | SpClust[14] | 0.25 | 0.20 | 0.88 | 0.15 | 0.07 |
| | ClustRF[48] | 0.23 | 0.17 | 0.87 | 0.14 | 0.08 |
| | AffProp[62] | 0.23 | 0.16 | 0.87 | 0.14 | 0.07 |
| | MMC[63] | 0.25 | 0.19 | 0.88 | 0.14 | 0.09 |
| BiTag | K-means[2] | 0.51 | 0.52 | 0.86 | 0.30 | 0.23 |
| | SpClust[14] | 0.71 | 0.73 | 0.93 | 0.56 | 0.60 |
| | ClustRF[48] | 0.77 | 0.81 | 0.94 | 0.64 | 0.60 |
| | AffProp[62] | 0.50 | 0.44 | 0.87 | 0.28 | 0.21 |
| | MMC[63] | 0.76 | 0.72 | 0.95 | 0.64 | 0.60 |
| DetScore | K-means[2] | 0.63 | 0.60 | 0.93 | 0.50 | - |
| | SpClust[14] | 0.82 | 0.76 | 0.96 | 0.69 | - |
| | MMC[63] | 0.83 | 0.78 | 0.96 | 0.73 | - |
| | L-MMC[3] | 0.86 | 0.82 | 0.97 | 0.79 | - |
| ViFeat&BiTag-cmb | K-means[2] | 0.51 | 0.49 | 0.90 | 0.34 | 0.24 |
| | SpClust-cmb[14] | 0.76 | 0.74 | 0.94 | 0.62 | 0.66 |
| | ClustRF[48] | 0.23 | 0.17 | 0.87 | 0.15 | 0.08 |
| | AffProp[62] | 0.51 | 0.46 | 0.86 | 0.29 | 0.21 |
| ViFeat&BiTag-bln | SpClust-bln[14] | 0.75 | 0.72 | 0.95 | 0.62 | 0.59 |
| | CCA+SpClust[19] | 0.85 | 0.81 | 0.97 | 0.77 | 0.75 |
| | 3VCCA+SpClust[21] | 0.86 | 0.86 | 0.97 | 0.78 | 0.77 |
| | CC-Forest[47] | 0.41 | 0.33 | 0.89 | 0.41 | 0.19 |
| | AASC[13] | 0.30 | 0.15 | 0.87 | 0.13 | 0.06 |
| | MMC[63] | 0.79 | 0.72 | 0.95 | 0.66 | 0.66 |
| | DCCA[24] | 0.84 | 0.80 | 0.96 | 0.74 | 0.72 |
| | DCCAE[25] | 0.84 | 0.80 | 0.97 | 0.75 | 0.73 |
| | S-MMC[4] | 0.87 | 0.84 | 0.97 | 0.79 | - |
| | F-MMC[4] | 0.90 | 0.88 | **0.98** | 0.84 | - |
| | HML-RF(Ours) | **0.94** | **0.90** | **0.98** | **0.88** | **0.87** |

# Evaluations

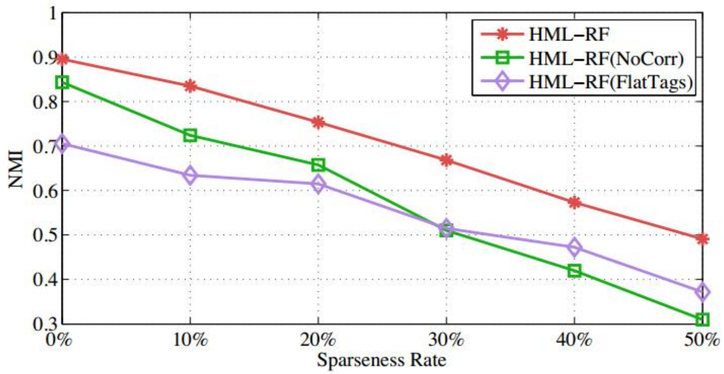
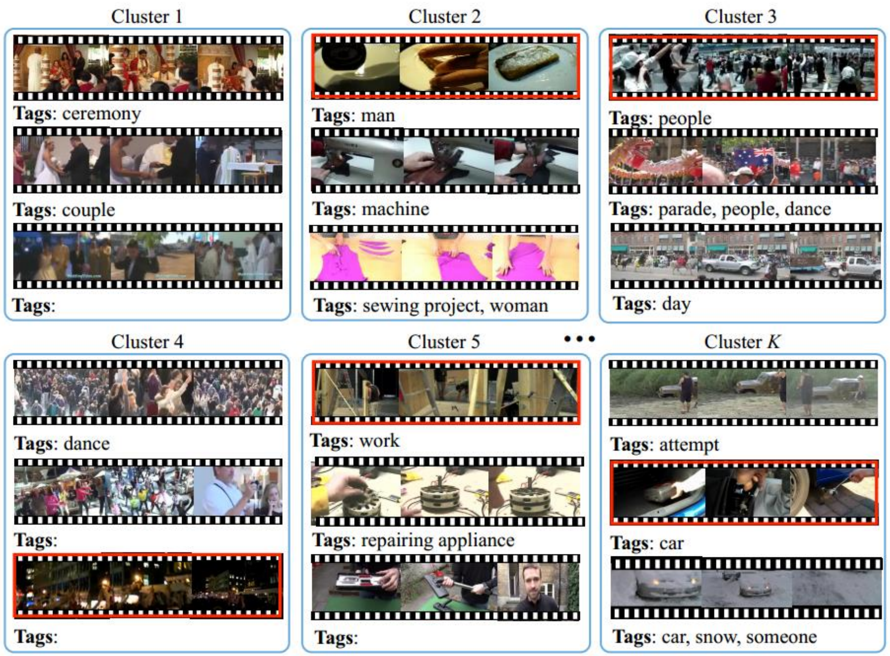



Figure 7: Evaluating the effectiveness of specific HML-RF components on TRECVID MED 2011 [56].



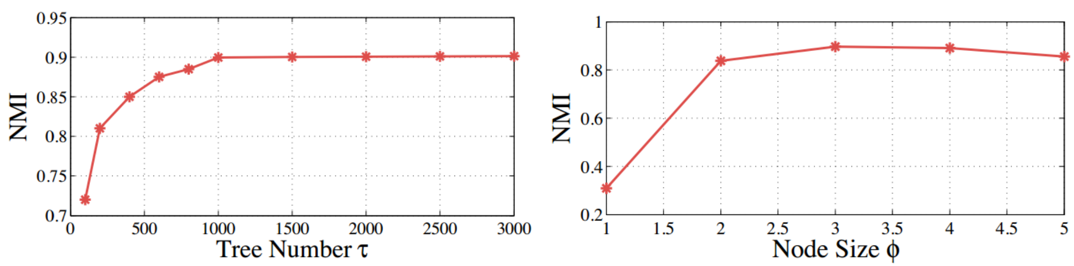




Figure 8: Clustering performance in NMI of HML-RF over different forest sizes ($\tau$) and node size ($\phi$) on TRECVID MED 2011 [56].

Thank you