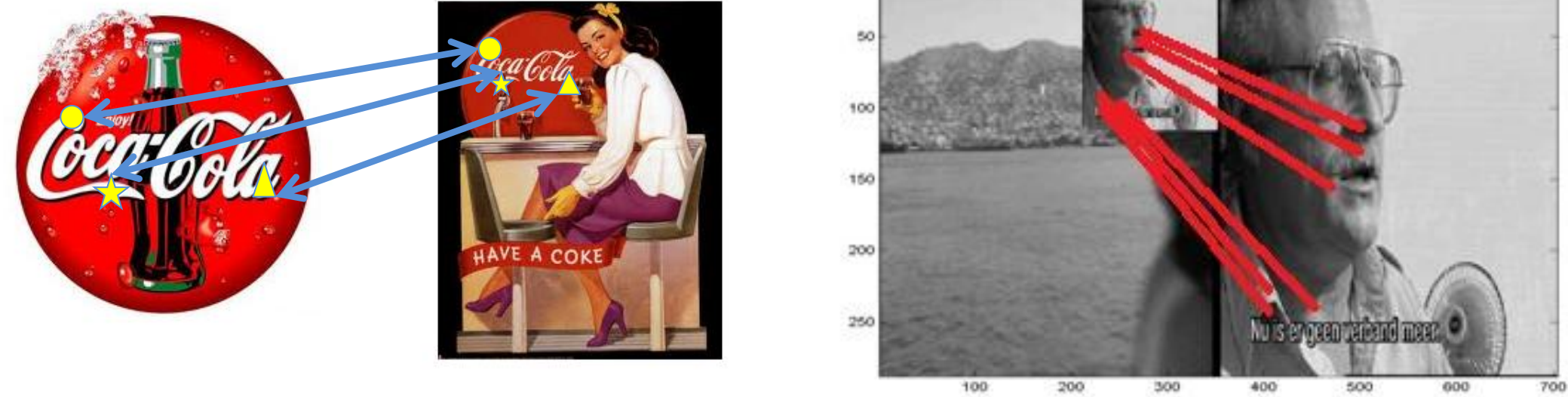


## Abstract

- Picture-in-Picture (PiP) is a useful service to watch two or more videos simultaneously, however it can be exploited to visually hide one (or more) video(s) inside another video
- Today's copy detection techniques can be easily fooled by PiP, which is reflected in the poor results in TRECVID competitions
- Inspired by partial image matching, we propose a novel generalized spatial coding scheme and efficient spatial verification algorithms to address PiP copy detection

## Motivation

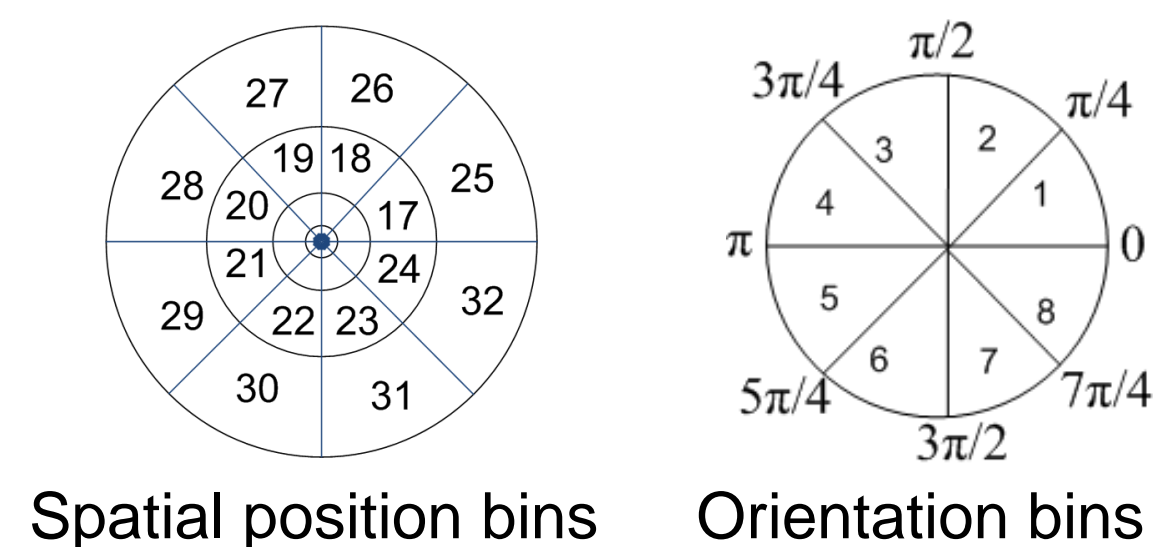
- PiP video key-frame copy detection can be seen as partial-duplicate image patch matching
- Partial duplicate image matching - search for partial-duplicate patches between two images/video keyframes



## Proposed Approach

### Generalized Log-Polar Spatial Coding

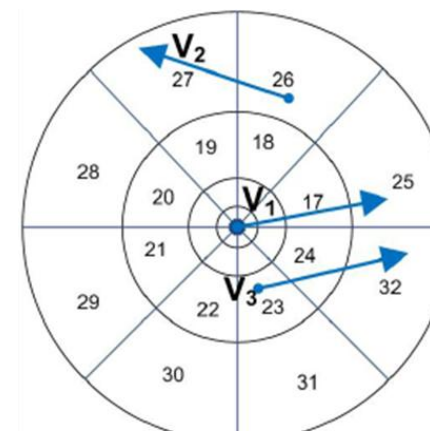
- Visual word is obtained by SIFT descriptor quantization
- Log-polar plot captures 'Shape context' better



## Spatial Coding

- Spatial code – Encode relative spatial position and relative orientation between visual words
- Spatial code of  $V_2$  w.r.t  $V_1$  ( $V_{12}$ ) is given by 01100010
- Spatial Map – spatial codes of all visual words

	$V_1$	$V_2$	$V_3$
$V_1$	0	10011010	00010111
$V_2$	11011110	0	11011110
$V_3$	00010011	10011010	0



## Spatial Verification

- Spatial layout comparison between query and target video key-frames

### Polynomial Algorithms

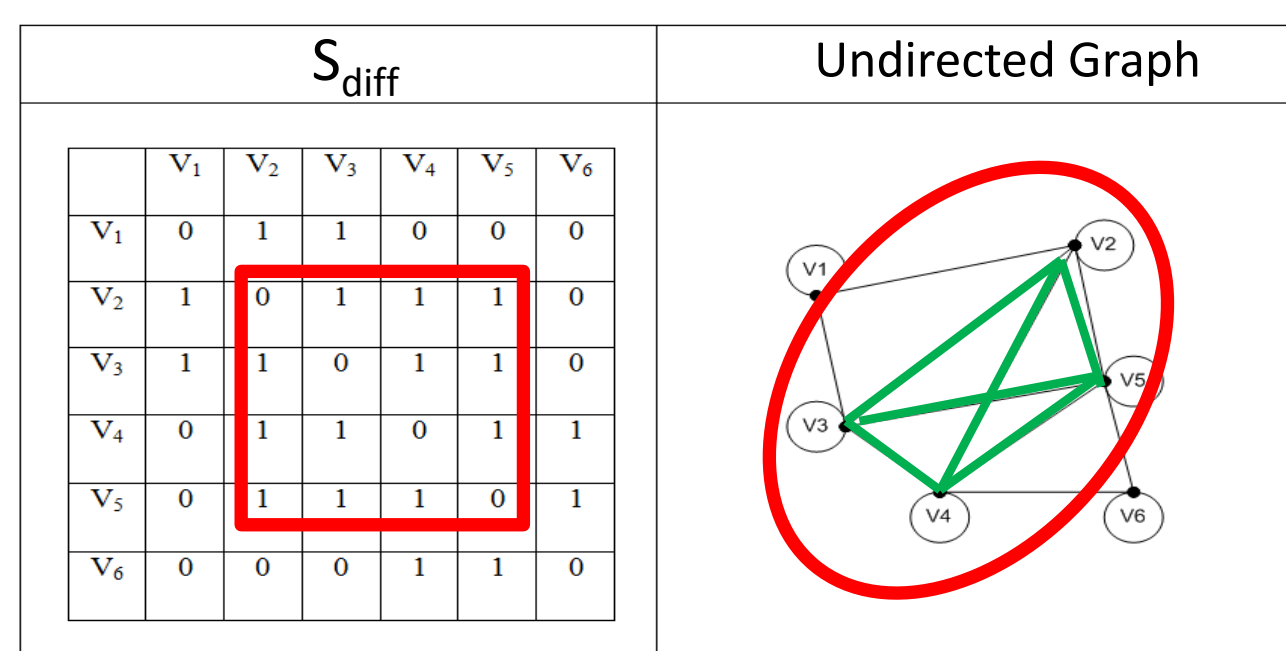
- Inconsistency Sum Method (ISM) Spatial Maps comparison
  - $V(i, j) = M_{query}(i, j) \oplus M_{target}(i, j)$
  - Inconsistency Sum for each feature
    - $S(i) = \sum_j V(i, j)$
    - Removal of false matches
      - $i^* = \arg \max_i S(i) \quad \forall S(i) \neq 0$

## Spatial Verification

### Non-Polynomial Algorithms

#### Maximum Clique Problem (MCP) Approach

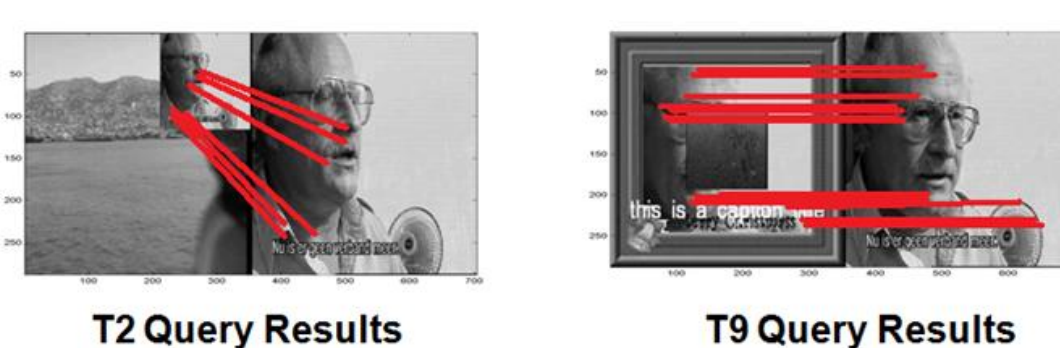
- Formulate spatial verification as finding the maximum clique in an undirected graph
- Find spatial maps ( $S_P$  and  $S_Q$ ) of *target* ( $P$ ) and *query* ( $Q$ ) images
- Convert  $S_{diff} = S_P \oplus S_Q$  into an adjacency matrix
- $S_{diff}$  can be visualized as a graph, with '1' representing edges and visual words representing vertices
- All true matches implies edges in a graph, therefore, maximum clique implies all consistent true matches



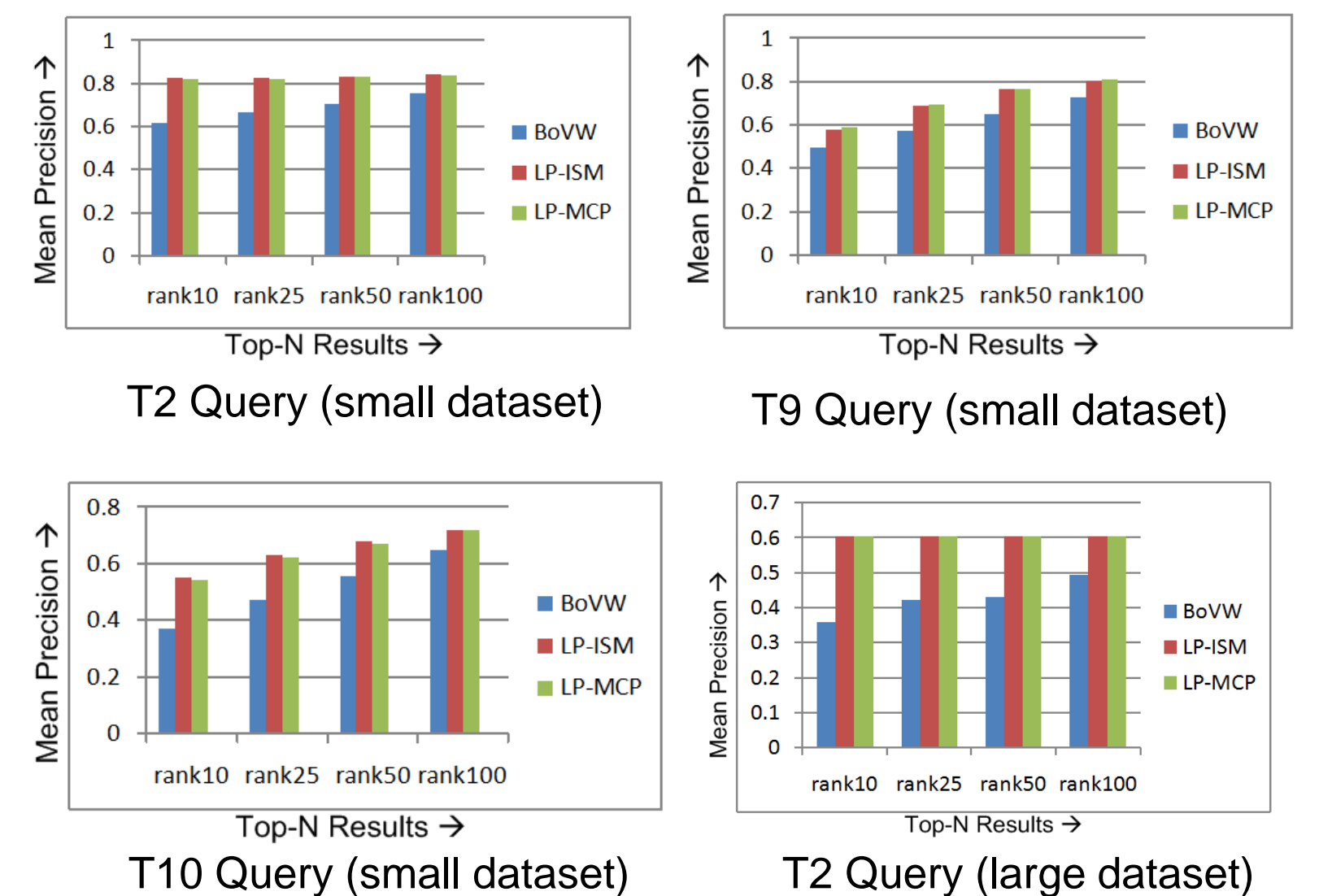
## Experiments

### Experimental setup

- Datasets
  - Small -120 TRECVID 2010 videos indexed (10-30s)
  - Large (MSRA v2.0 + TRECVID 2010) – 12,650+ videos (~10s to 4 min)
- Testing dataset (10-30 s) - T2 query, T9 query, T10 query – 120 videos each
- Comparison: Bag-of-visual-words using vocabulary Tree
- Metrics: (a) mAP, (b) mean processing time for query frame

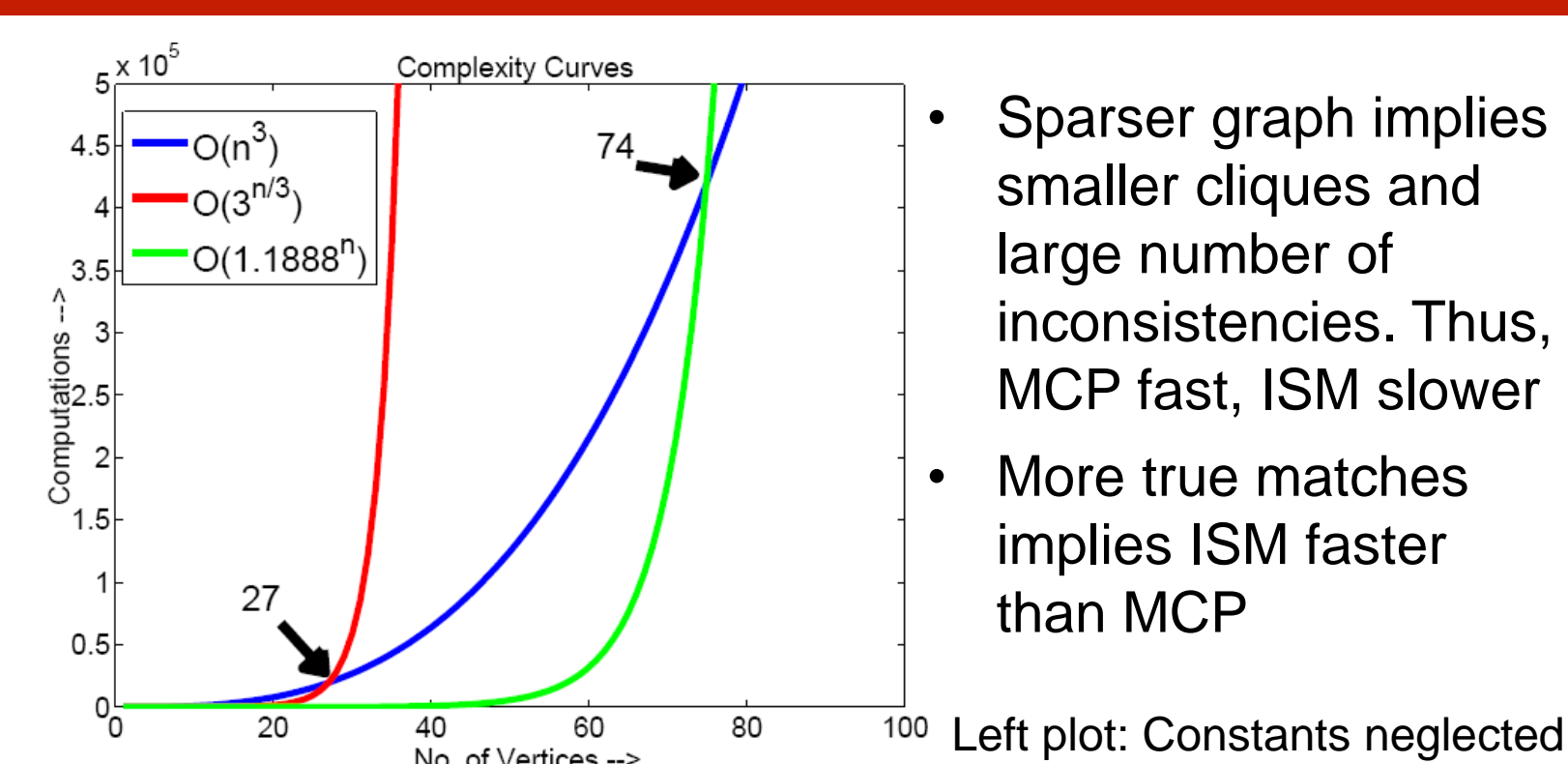


### Results



Spatial Verification	Complexity	Avg. time
Inconsistency Sum Method (ISM)	$O(n^3)$	2 seconds
MCP approach: Bron-Kerbosch algorithm	$O(3^{n/3})$	0.5 seconds

## Discussion



- Sparser graph implies smaller cliques and large number of inconsistencies. Thus, MCP fast, ISM slower
- More true matches implies ISM faster than MCP

## Conclusions

### Contribution

- Generalized Log-Polar Spatial Codes to jointly encode relative spatial positions & orientations
- Spatial Verification Algorithms: Efficient spatial layout comparison

### Future work

- Feasibility of MCP approach for large graphs
- Video-level comparison instead of key-frame comparison

- S. Purushotham, Q. Tian, C.-C.J. Kuo, "Picture-in-Picture Copy Detection Using Spatial Coding Techniques", ACM MM, AIEMPro Workshop, Scottsdale, Arizona, USA
- W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian, "Spatial Coding for Large Scale Partial-Duplicate Web Image Search", ACM Multimedia, Firenze, Italy, October 25-29, 2010