VIREO@INS-TV13

Search of Small Objects by Topology Matching, Context Modeling, and Pattern Mining

Wei Zhang, Chong-Wah Ngo

VIREO: VIdeo REtrieval grOup
City University of Hong Kong

Outlines

- Introduction
- Solutions
  - TC: Topology Checking
  - CM: Context Modeling
  - PM: Pattern Mining
- Conclusion

Outlines

- Introduction
- Solutions
  - TC: Topology Checking
  - CM: Context Modeling
  - PM: Pattern Mining
- Conclusion

General Information

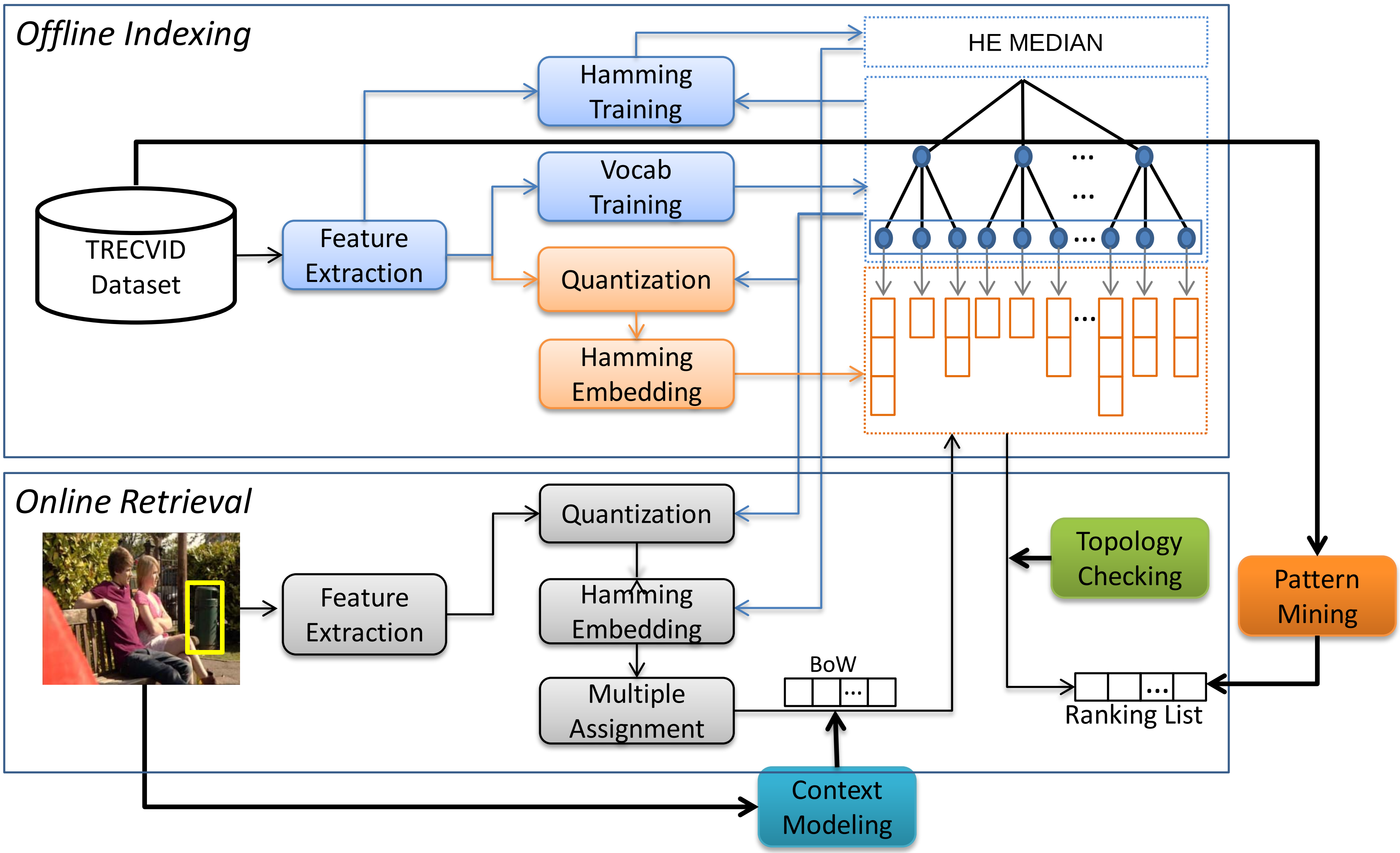
- Reference dataset
  - 464-hours Videos
  - 470k Shots
  - 640k keyframes
    - 1 frame every 4 seconds
    - ≈ 1.36 frames/shot

- Query
  - 30 topics: object(26) + person(4)
  - query image + ROI

9075: a SKOE can

- Our Baseline system
  - BoW model
  - visual matching based on SIFT

Retrieval Framework

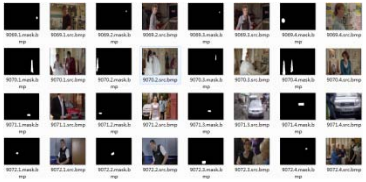


Retrieval Framework

- Time efficiency
  - ~ 300ms/query: time cost for online search
  - ~ 10s/topic, including everything:
    - 4 queries
    - feature extraction, quantization, online search, re-ranking
- Memory cost: ~12 Gbytes
- Source code for the basic framework
  - available as as part of “*VIREO-VH: Video Hyperlinking*”
  - http://vireo.cs.cityu.edu.hk/VIREO-VH/

VIREO Video Retrieval Group

Main Challenge



- A target is considered as **small**, if it covers **< 10%** area
- For TV13, **77%** of queries are **small** !

small instance on **query** image

– lack of knowledge on the search target

small instance on **reference** image

– similarity score is easily diluted

more sparse

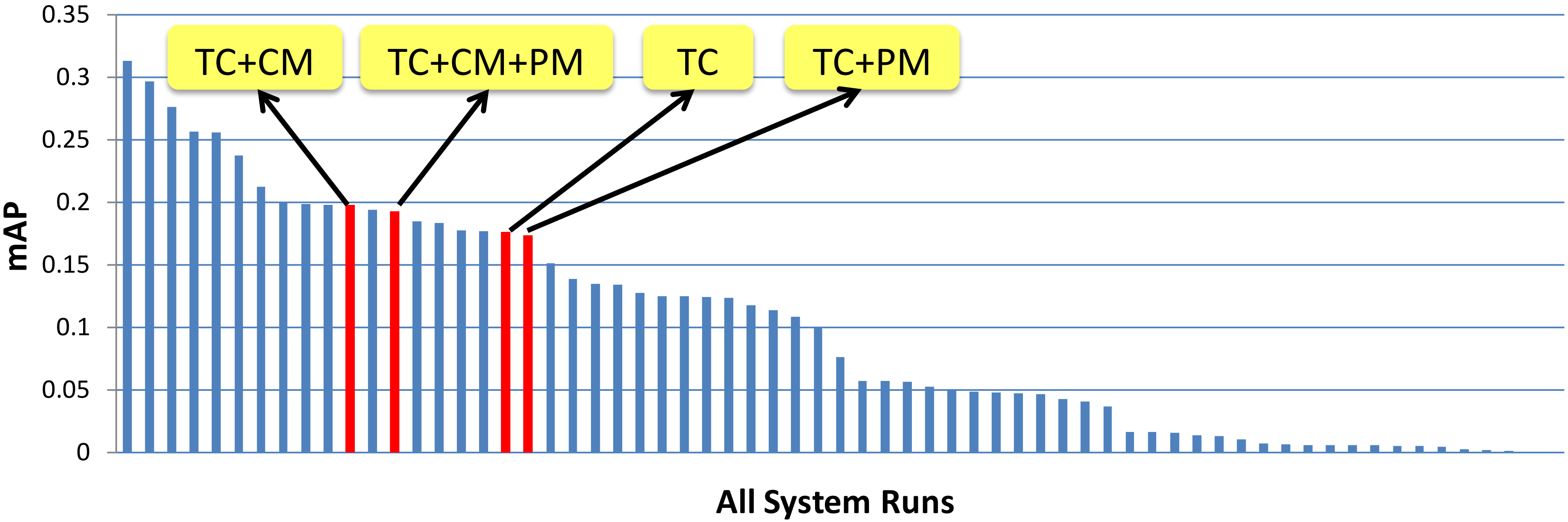
$$\boxed{\mathrm{sim}(\mathbf{q},\mathbf{r})} = \frac{\boxed{\mathbf{q}}\cdot\mathbf{r}}{|\mathbf{q}|\,\boxed{|\mathbf{r}|}}$$

sensitive to noise

- Topology Checking (TC)
  – make better use of limited info by elastic spatial checking
- Context Modeling (CM)
  – increase information quantity by considering background context
- Pattern Mining (PM)
  – link small instances offline

Our Submissions

- Three techniques
  - Topology Checking : TC
  - Context Modeling : CM
  - Pattern Mining : PM



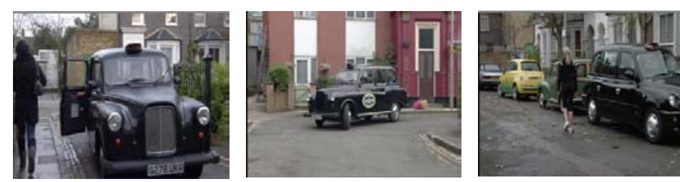
Outlines

- Introduction
- Solutions
  - TC: Topology Checking
  - CM: Context Modeling
  - PM: Pattern Mining
- Conclusion

Topology Checking

- Spatial transformation in INS
  - What we might expect
    - linear transforms (scaling, rotation, translation, shearing)
  - What we actual have
    - much more complex transforms
- The verification model we want
  - tight enough to reject false matches
  - tolerant complex spatial transformations

9088: Tamwar – non-rigid motion



9081: a black taxi – different views of non-planar obj

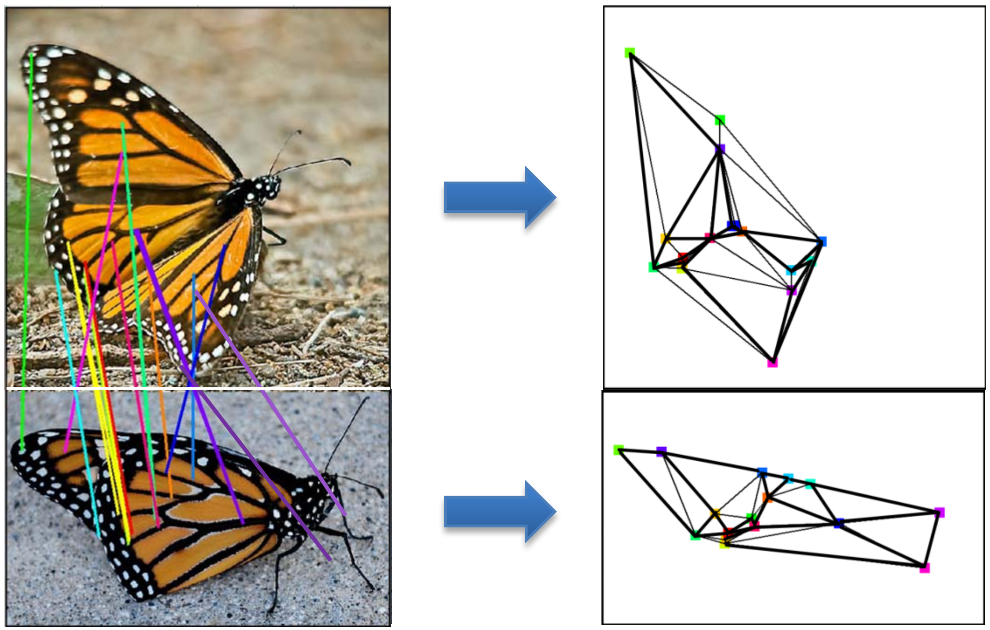
Topology Checking - Illustration

- Sketch - Match

Delaunay Triangulation (DT)

$\mathcal{Q}$

$\mathcal{R}$



$\Delta\mathcal{Q}$

$\Delta\mathcal{R}$

matched points (15)

$\mathbf{E}_{\Delta\mathcal{Q}}$: edges in $\Delta\mathcal{Q}$

$\mathbf{E}_{\Delta\mathcal{R}}$: edges in $\Delta\mathcal{R}$

$|\mathbf{E}_{\Delta\mathcal{Q}}|$= 42

$|\mathbf{E}_{\Delta\mathcal{R}}|$= 42

common edges (28)

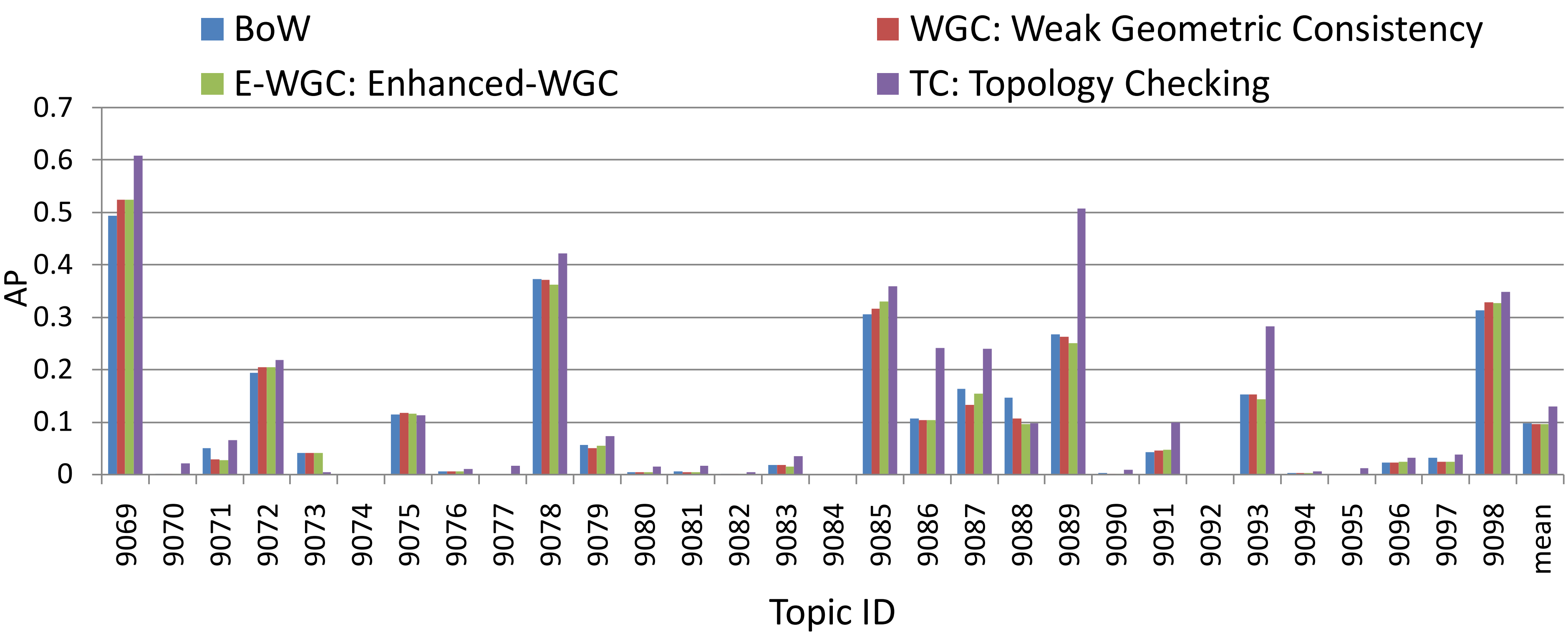
$$\mathrm{BF}(\mathcal{Q},\mathcal{R}) = \|\mathbf{E}_{\Delta\mathcal{Q}} \cap \mathbf{E}_{\Delta\mathcal{R}}\|$$

Benefits of Topology Checking (TC)

- Edge of the graph
  - encode relative positioning / spatial nearness
- # common edges depicts the topology similarity
- Avoid using noisy local features’ scale/orientation
  - local features’ orientation / scale are biased
  - only location is used

- Get evidence from multiple <u>*local*</u> consistent sub-regions
  - robust to small viewpoint change / motion

Results for spatial checking – *ROI Only*



Outlines

- Introduction
- Solutions
  - TC: Topology Checking
  - CM: Context Modeling
  - PM: Pattern Mining
- Conclusion

Full-Image v.s. ROI search

- Full-Image is mostly better, since:
  - limited info inside small ROI
  - high correlation between *ROI* and its *background*
    - they appear/disappear together

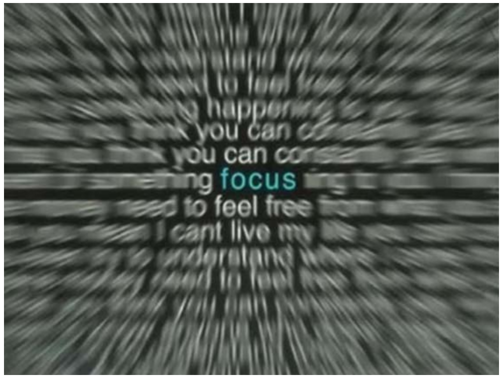
9070: small red obelisk
<obelisk, this painting>
<obelisk, this room>
<obelisk, this woman>

- Sometimes, ROI is better, when:
  - low correlation ➔ instances that could appear anywhere

Context Modeling

- Observation
  - Feature ∈ ROI: highly correlated with the target
  - Feature ∉ ROI: correlation degenerates quickly.

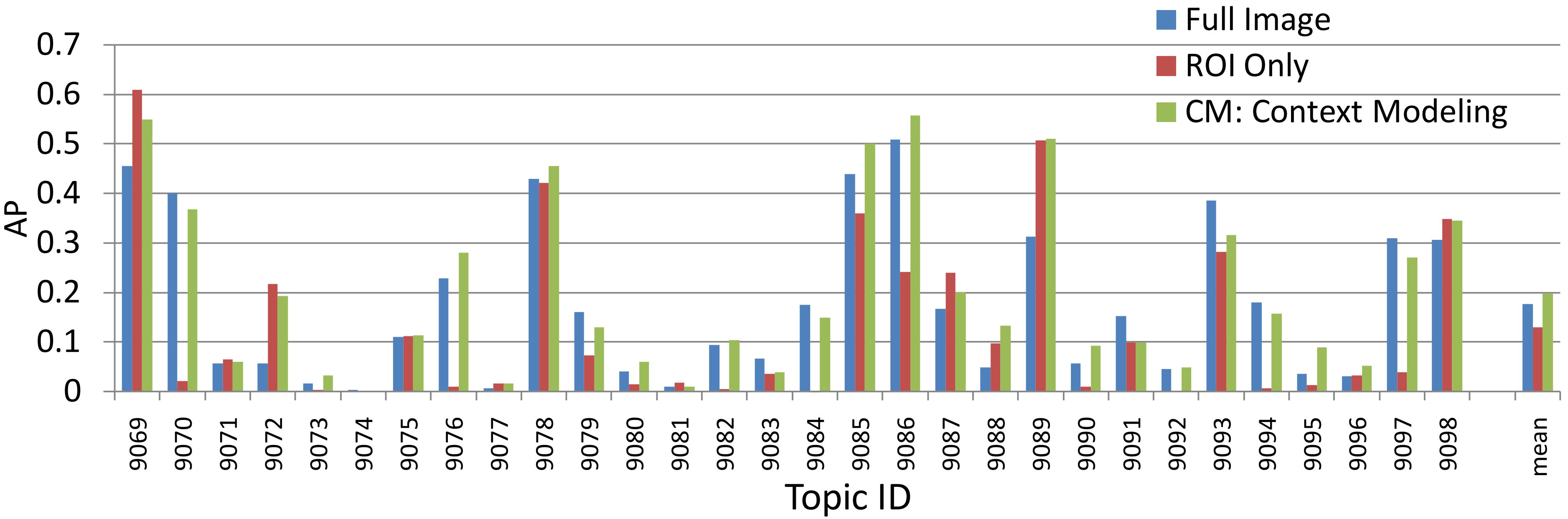
- Context modeling
  - weight background context
  - simulate the behavior of "[stare]"
  - blur things away from the [focus]



$$k(x) = \begin{cases} 1, & \text{if } x \in \mathbf{ROI}, \\ \exp(-\frac{\|x-f\|^2}{2\delta^2}), & \text{otherwise}, \end{cases} \quad \text{with } \delta^2 = -\frac{diag^2}{8\ln 0.1}$$

Results - Context Modeling

- Tradeoff between two extremes
- Avoids zero-performance, when one of them does not work
- Improves overall performance



Outlines

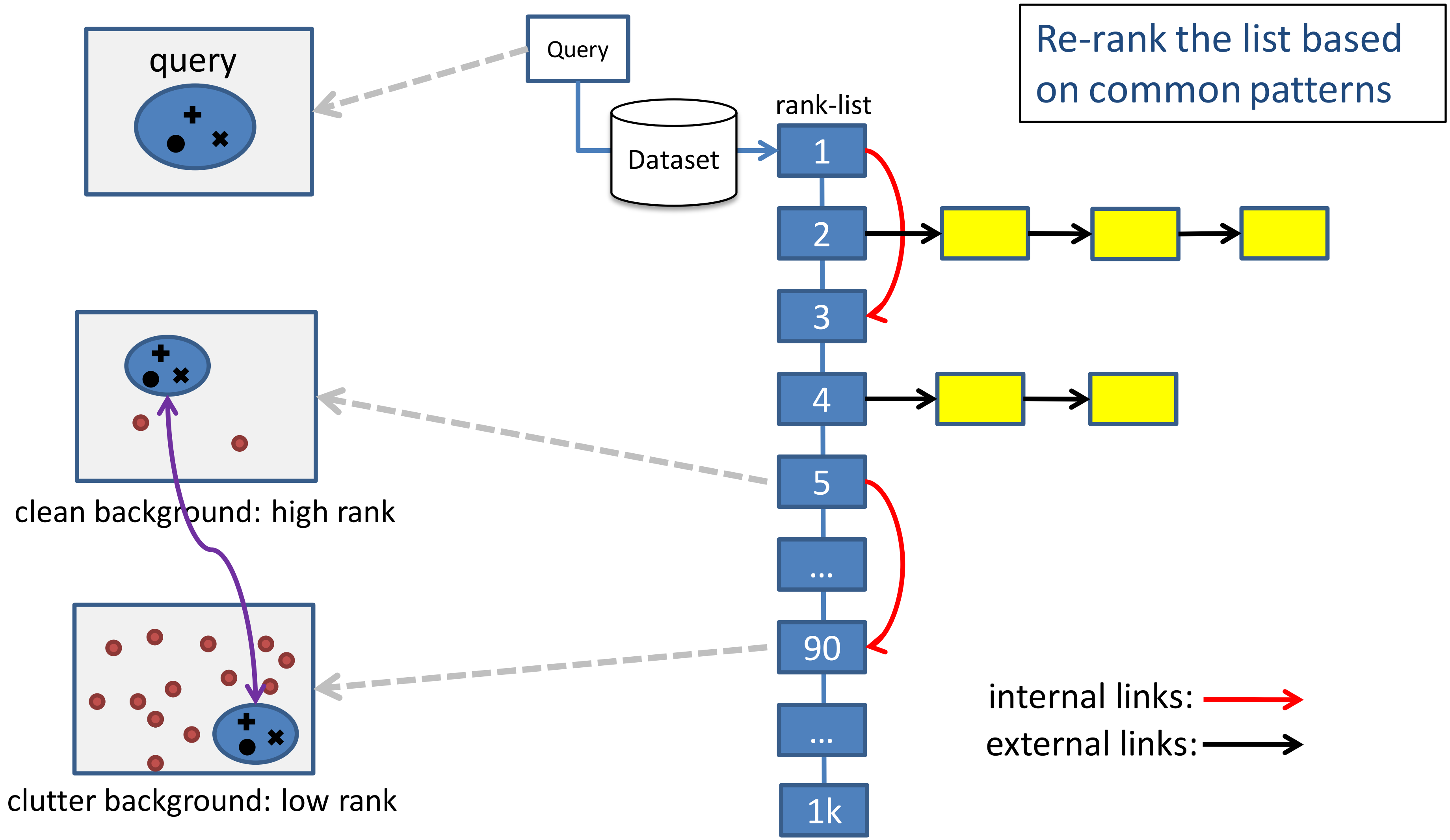
- Introduction
- Solutions
  - TC: Topology Checking
  - CM: Context Modeling
  - PM: Pattern Mining
- Conclusion

Common patterns

- “BBC Easterenders” dataset
  - repetitions of {characters, scenes, objects}
  - hyperlink shots with common patterns

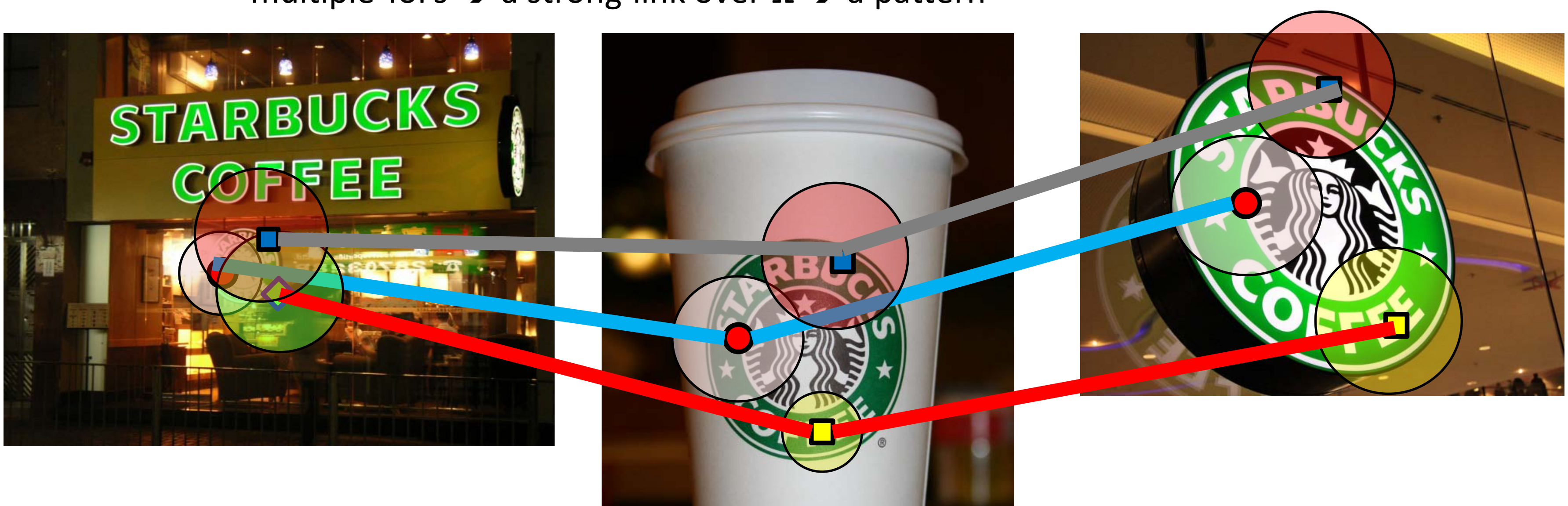
- Are these patterns useful for INS?
  - large patterns ➔ no harm
    - Near Duplicates
    - already easy to retrieve
  - small patterns ➔ potentially helpful
    - small objects
    - difficult to retrieve

Improve INS with Common Patterns



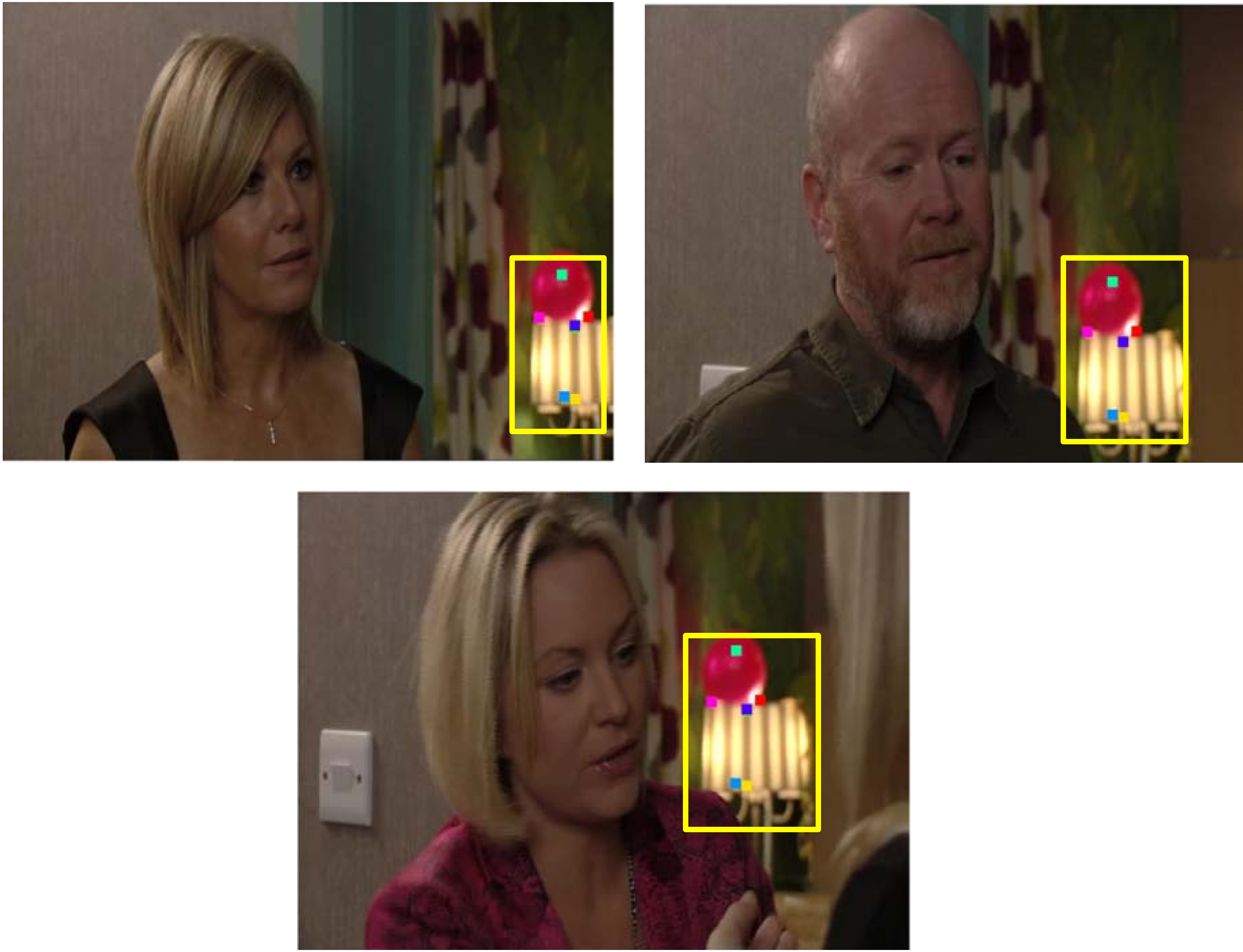
How to mine Common Patterns

- Extract ToF (Thread of Feature)
  - a ToF is a set of consistent patches across images
  - represented as a set of image ids
- Cluster ToF
  - min-Hash is adopted for efficient clustering
  - clustered ToFs
    - each ToF ➔ a link over a set of images Ω
    - multiple ToFs ➔ a strong link over Ω ➔ a pattern



Patterns Mined from TV13 dataset

- Near Duplicates (ND)
  - easiest pattern to mine
  - many similar shots in TV series
- Objects/scenes
- Only a few is related with the 30 topics
- Some examples …




KIDS ALONE
BBC
BRAND NEW
TONIGHT 9.0
LAST MAN STANDING


Resident
permit
holders
only


NO SMOKING.
It is against the law to smoke in these premises
CALL


It is against the law to


It is against the law to

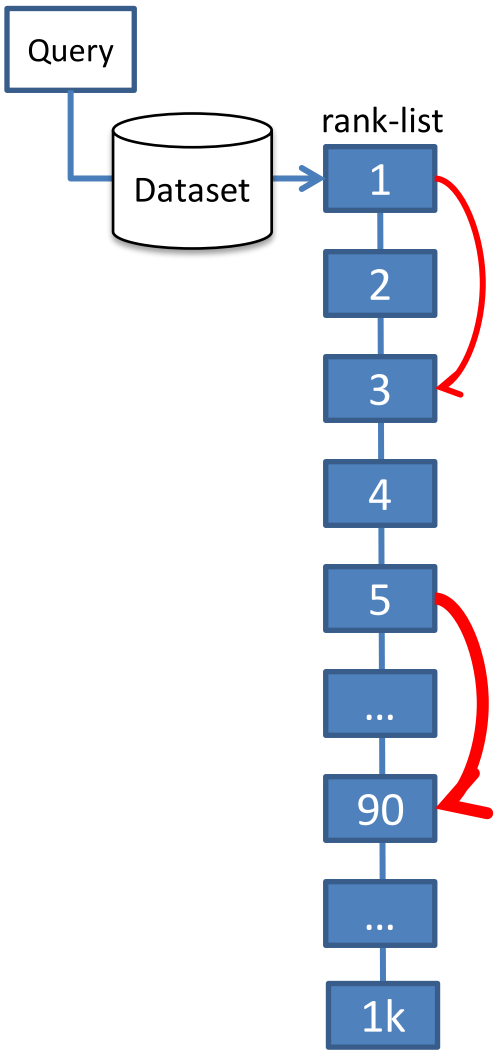

Central London
THERE'S ONLY ONE
It is against the law to
BACK IN
5 MINS
RACING POST
OPEN


SARKAR


Kathy's
P

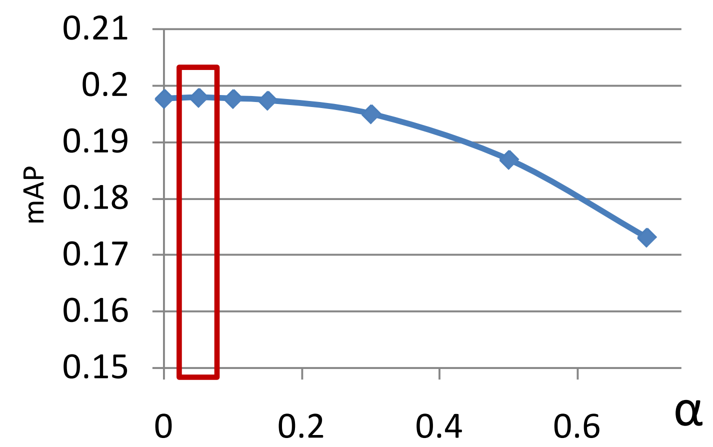
Approach-1: Frame-level linking

- Re-rank results using patterns
  - Random Walk
  - nodes: top 1k images in rank-list
  - initial weights: retrieval scores
  - link: mining results
  - link strength:
    - # patterns containing the image pair

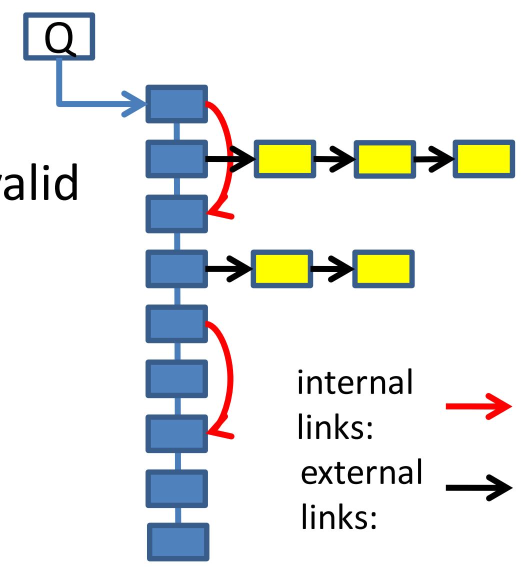


Results – Frame-level Linking

- Results
  - weight for mining result : α
  - weight for retrieval score : 1 - α
  - best performance: α ≈ 0

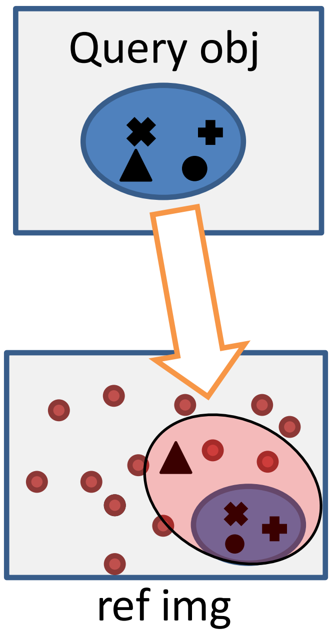


- Problems
  – only internal links are considered
  – transitivity propagation at frame-level is not valid
  – most links has nothing to do with the query
  – emphasize Near Duplicates
    – NDs always have strong links



Approach-2: Instance-level linking

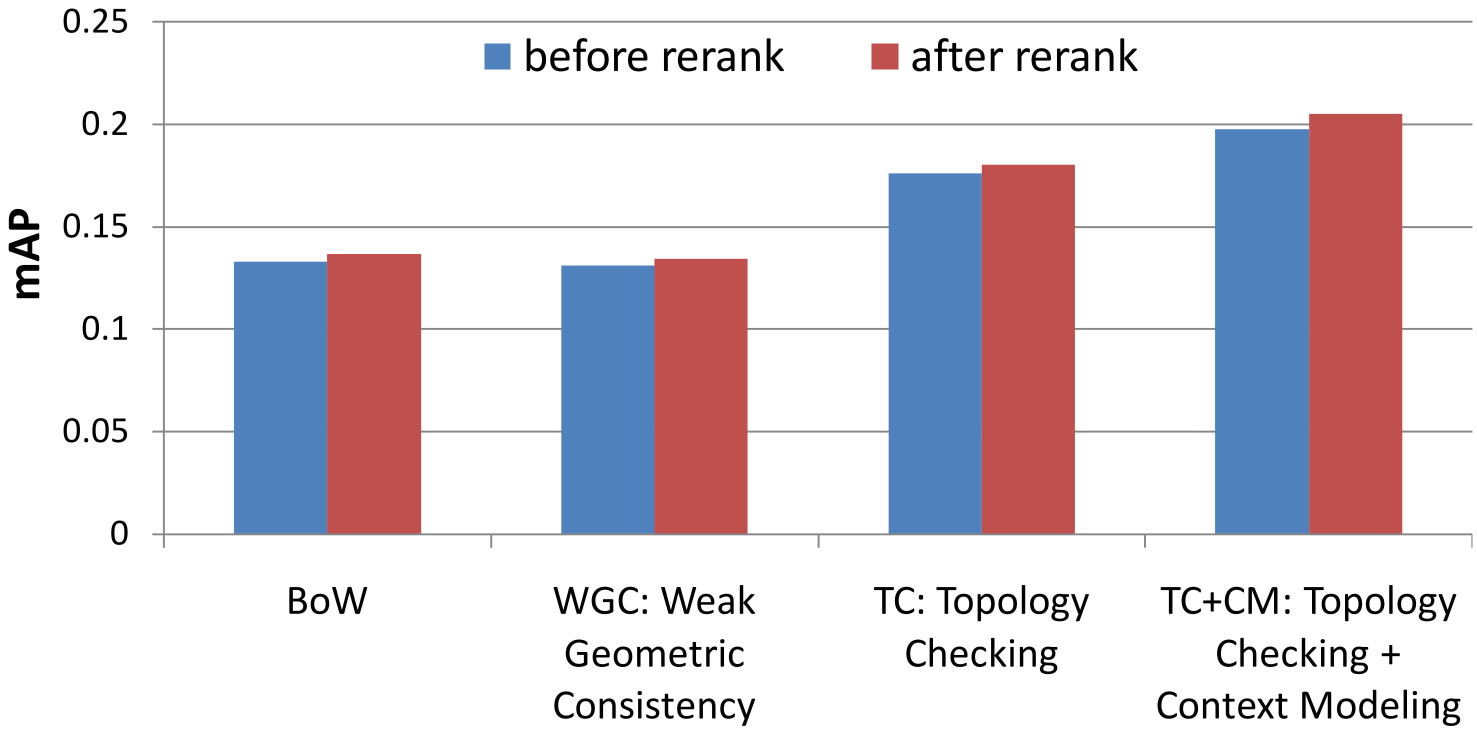
- Encode locations of matched points via (μ, σ^2)
  - μ: the centroid of matched points
  - σ^2: the variance of the location
  - Z-test for region overlapping
    - two sets of points overlap, if $Z = \frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}} < t$



- Rank strategy
  - no distinction on link strength (binary strength)
  - give a bonus score to the linked images (both in/external links)

Results – Instance-level Linking

- Mining improves corresponding results consistently
  - invalid transitivity is prevented
  - only a few links are related with the 30 topics



Outlines

- Introduction
- Solutions
  - TC: Topology Checking
  - CM: Context Modeling
  - PM: Pattern Mining
- Conclusion

Conclusion

- Visual matching is mostly enough, despite low sampling rate
- Small objects are still difficult to search

- Complex spatial configuration in INS
  - Topology suits better

- ROI v.s. full-image search
  - tradeoff between precision and recall
  - generally, full-image search performs better, and
  - proper weighting is even better

- Pattern mining
  - many patterns can be linked offline
  - large fraction is near duplicates
  - low overlap with the query is the major problem