NAGOYA UNIVERSITY

GRADUATE PROGRAM FOR REAL-WORLD DATA CIRCULATION LEADERS
PROGRAM FOR LEADING GRADUATE SCHOOLS
NAGOYA UNIVERSITY

Nagoya University at TRECVID 2014: the Instance Search Task

Cai-Zhi Zhu [1], Yinqiang Zheng[2], Ichiro Ide [1],
Shin'ichi Satoh [2], Kazuya Takeda [1]
[1] Nagoya U. Japan [2] NII, Japan
November 10, 2014

Outline

- A Practical Spatial Re-ranking for Instance Search from Videos

- Our submissions at TRECVID INS 2014

Outline

- A Practical Spatial Re-ranking for Instance Search from Videos *

- Our submissions at TRECVID INS 2014

* *X. Zhou, C.-Z. Zhu, Q. Zhu, S. Satoh, Y.-T. Guo, A practical spatial re-ranking method for instance search from videos, In ICIP, 2014.*

The problem

- Classical RANSAC spatial re-ranking is proved to be effective for image retrieval purpose.
- Yet no work on classical spatial re-ranking has been systematically reported for instance search from videos so far
  - Efficiency is a big concern, as

    videos are composed of multiple frames, and frame-by-frame spatial verification is too prohibitive
  - Effectiveness is also unclear *

* *W. Zhang and C.-W. Ngo. Searching visual instances with topology checking and context modeling. In ICMR, 2013.*

Our efforts

- To efficiency
  - A representative image/frame selection scheme
    - ➢ Avoids verifying all images of a video/topic
  - VQ based tentative matching
    - ➢ Avoids expensive NN search and accessing raw SIFT.
- To effectiveness
  - A ROI-originated RANSAC scheme
    - ➢ Regards ROI as a priori in transformation computation and the background a posteriori in the voting phase.

Methodology

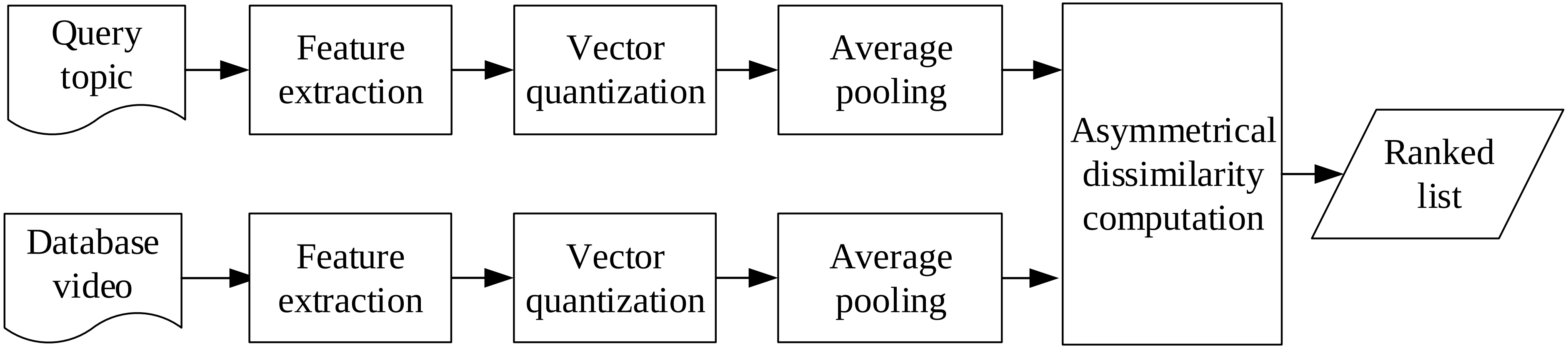
1. Representative selection:

$$(i_{rep}, j_{rep}) = \arg \max_{i \in \mathcal{I}, j \in \mathcal{J}} \|\min(\mathbf{Q}_i, \mathbf{V}_j)\|_{\ell_1}$$

2. VQ based tentative matching
   - Raw features quantized to the same visual words are considered as matches.

3. ROI-originated RANSAC
   - Transformation is estimated from the ROI while verified on the full image

Baseline system *



** C.-Z. Zhu, H. Jégou, S. Satoh, Query-adaptive asymmetrical dissimilarities for visual object retrieval, In ICCV, 2013.*

TRECVID INS2013 dataset

Experiments on INS2013

Table 1: Comparison with the baseline and different configurations of the spatial re-ranking method on the INS2013 dataset.

Method	Query	Video	Match	CReg	SReg	infAP	Mins
M1	RMD	RMD	VQ1	ROI	FUL	31.61	19
M2	REP	REP	VQ1	ROI	FUL	33.49	20
M3	ALL	ALL	VQ1	ROI	FUL	34.29	1,440
M4	ALL	REP	VQ1	ROI	FUL	33.77	80
M5	ALL	REP	VQ2	ROI	FUL	34.39	96
M6	ALL	REP	VQ3	ROI	FUL	34.58	128
M7	ALL	REP	NN	ROI	ROI	30.28	800
M8	ALL	REP	VQ1	ROI	ROI	33.72	80
M9	ALL	REP	VQ1	FUL	ROI	33.32	200
M10	ALL	REP	VQ1	FUL	FUL	28.00	200
BL	–	–	–	–	–	31.33	–

Examples

Conclusions

- The representative selection scheme does take effect: which significantly outperforms the random method, while slightly inferior to exhausting matching all image pairs.
- Surprisingly, the VQ based method is remarkably better than the NN based method: burstness, reference…
- In general, larger *k* better performance for VQ based method.
- The ROI-originated RANSAC method really works, that's to say, transformation is better to be computed from the ROI (as a priori information). In contrast, the verification should be fulfilled on the full image (background helps in voting).

1. *C.-Z. Zhu, X. Zhou, and S. Satoh. Bag-of-words against nearest-neighbor search for visual object retrieval. In ACPR, 2013.*
2. *X. Zhou, C.-Z. Zhu, Q. Zhu, S. Satoh, Y.-T. Guo, A practical spatial re-ranking method for instance search from videos, In ICIP, 2014.*

Outline

- A Practical Spatial Re-ranking for Instance Search from Videos

- Our submissions at TRECVID INS 2014

Our submissions

- Experimental settings (same as last year)
  - Sample 5 frames/sec.
  - SIFT only
    - 3 detectors: Hessian-affine, Harris-Laplace and MSER.
    - 2 descriptors: Root-SIFT and color SIFT.
- Our submissions
  - NU_1: The proposed re-ranking method on NU_2.
  - NU_2: Asymmetrical δ_2 dissimilarity with multiple SIFTs.
  - NU_3: Asymmetrical δ_1 dissimilarity with multiple SIFTs.
  - NU_4: Our own implementation of HE*

* *H. Jegou, M. Douze, and C. Schmid. Hamming embedding and weak geometric consistency for large scale image search. In ECCV, 2008.*

Performance table

Table 2: Our submissions for the TRECVID INS2014.

Method ID	x=A	x=B	x=C	x=D	x=E	Method Description
F_x_NU_1	–	–	–	30.44^4	–	δ_2+re-ranking
F_x_NU_2	19.11^1	24.56^1	26.50^1	28.77^5	28.99^2	δ_2 [16]
F_x_NU_3	18.33^2	21.95^2	23.71^2	25.56^6	21.42^3	δ_1 [16]
F_x_NU_4	16.00^3	18.75^3	22.27^3	24.34^7	–	An implementation of HE [2]

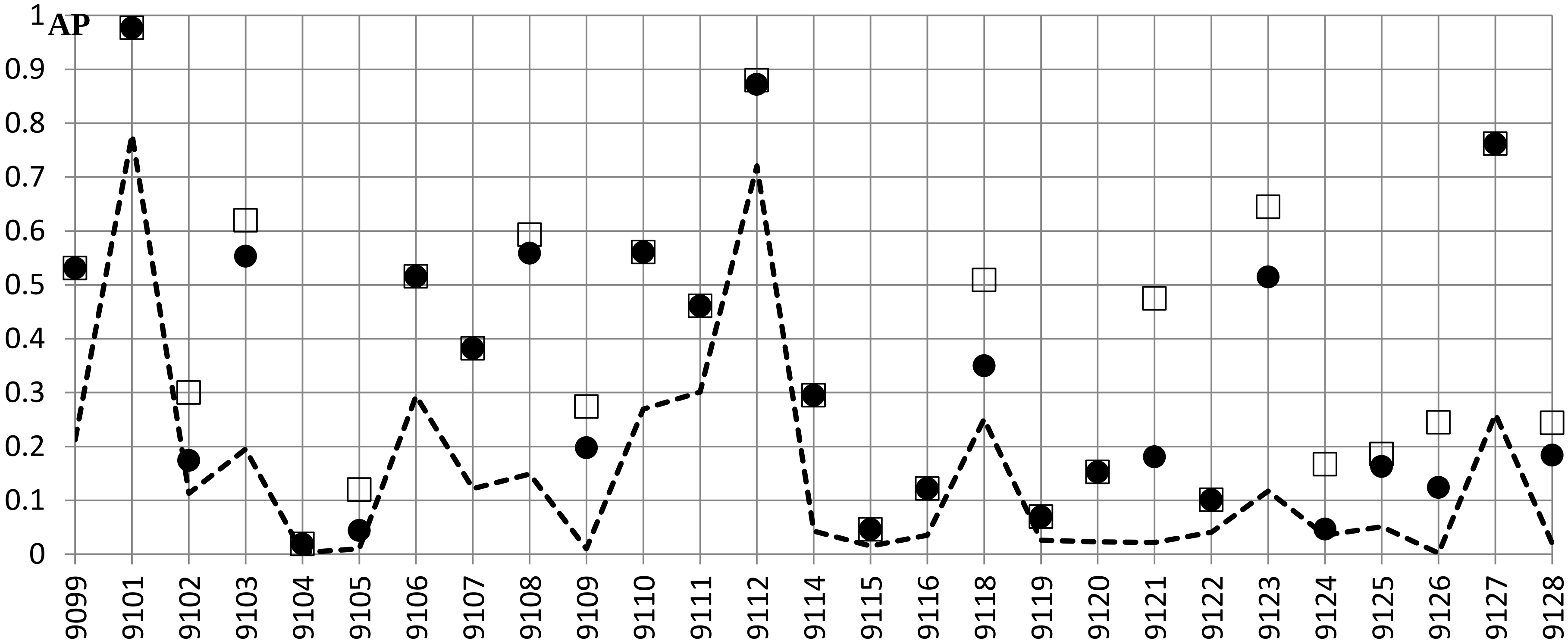
Note: The official performance were reported on 27 query topics after three topics (9100, 9113 and 9117) being excluded. The superscript digits are the ranking in the corresponding track.

1. We take the average of scores of querying by each image, which performs similarly as searching after pooling all images
2. In general, more query images better performance*

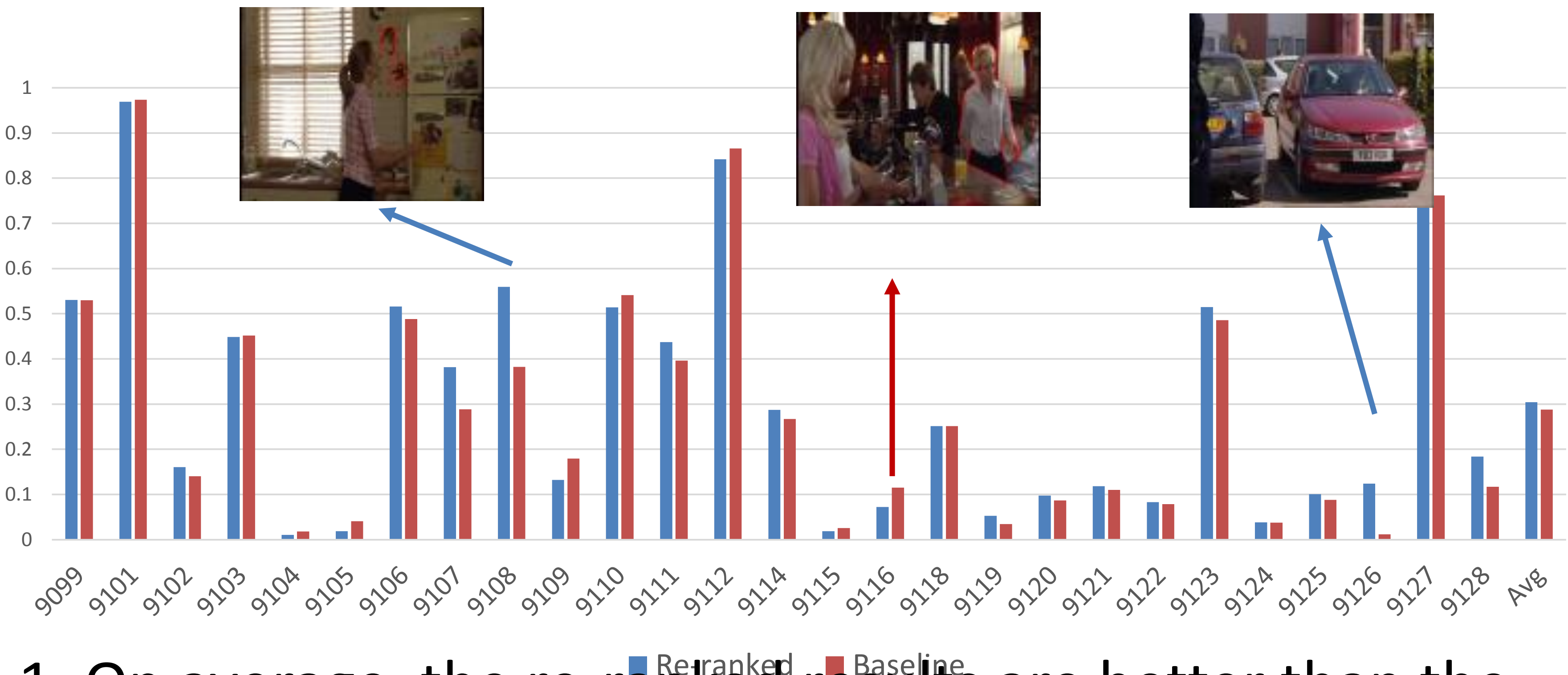
* *C.-Z. Zhu, Y.-H. Huang, and S. Satoh. Multi-image aggregation for better visual object retrieval. In ICASSP, 2014.*

Performance per topic of the best run

- In total we won 14 out of 27 topics.



Compare results on topics



1. On average, the re-ranked results are better than the baseline, quite significantly on some topics.
2. While on some topics, the re-ranking method performs even worse. Therefore, a fusion algorithm, like NII method, could further improve the overall performance.


9099
9100
9101
9102
9103
9104
9105
9106
9107
9108
9109
9110
9111
9112
9113
9114
9115
9116
9117
9118
9119
9120
9121
9122
9123
9124
9125
9126
9127
9128

Thank you!