ICCV15
SANTIAGO, CHILE DECEMBER 11-18, 2015
International Conference on Computer Vision

Multi-View Complementary Hash Tables for Nearest Neighbor Search

Xianglong Liu, Lei Huang*, Cheng Deng†, Jiwen Lu‡, Bo Lang**

**State Key Lab of Software Development Environment, Beihang University, Beijing, China*

†Xidian University, Xi'an, China *‡Tsinghua University, Beijing, China*

Background and Related Work

Background

- The explosive growth of the vision data motivates the recent studies on hash based nearest neighbor search (NNS)
- Locality-Sensitive Hashing (LSH) is able to achieve compressed storage and efficient computation in NNS
- Building multiple hash tables and probing multiple buckets can boost the overall NNS performance

Related Work

- The most widely-used strategy: random LSH-based multi-table, working like multi-index hashing [27]
- Complementary hash tables: a sequential learning method [37]
- A general multi-table construction strategy: bit selection over existing hashing algorithms [21]

Main Issues

- It often requires a huge number of tables without eliminating the table redundancy
- Hash tables are usually learned only from single type of data source, while adaptively combining them can help learn more informative hash functions

Multi-View Hash Table

Notations

- A set of N training examples with M views
  - $x_n^{(m)} \in R^{d_m \times 1}$: the m-th feature (d_m dimension) of n-th sample
  - $X^{(m)} = \left[x_1^{(m)}, x_2^{(m)}, \dots, x_n^{(m)}\right]$: the m-th feature of all data

Goal

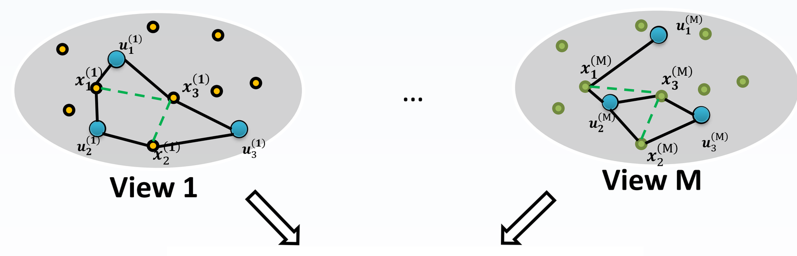
- Build L multi-view complementary hash tables $\{T_l\}_{l=1}^{L}$, with B hash functions $H = \{h_j(\cdot)\}_{j=1}^{B}$, $h_j(\cdot)$: $R^d \to \{-1,1\}$, learnt for each table.

Exemplars based Feature Fusion

- nonlinear feature mapping

$$\left[\mathbf{z}_i^{(m)}\right]_k = \frac{\delta_k^{(m)} \mathcal{K}(\mathbf{x}_i^{(m)}, \mathbf{u}_k^{(m)})}{\sum_{k'=1}^{K} \delta_k^{(m)} \mathcal{K}(\mathbf{x}_i^{(m)}, \mathbf{u}_{k'}^{(m)})}$$

- feature fusion

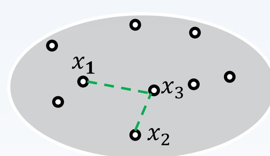


$$\mathbf{z}_i^* = \frac{1}{\lambda}[\mu_1^r \mathbf{z}_i^{(1)}; \dots; \mu_M^r \mathbf{z}_i^{(M)}]$$

Data points · Data points in different views · Exemplars
Similarity to nearest neighbors — Similarity to nearest exemplars

Hash Function Learning

- low-rank similarity



$$\mathbf{S} = \mathbf{Z}^* \mathbf{\Lambda}^{-1} \mathbf{Z}^{*\mathrm{T}}$$

- alternating optimization

$$\max_{\mathbf{Y}} \quad \mathrm{tr}(\mathbf{Y}^{\mathrm{T}} \mathbf{S} \mathbf{Y})$$
$$\text{s.t.} \quad \mathbf{1}^{\mathrm{T}} \mathbf{Y} = 0, \ \mathbf{Y}^{\mathrm{T}} \mathbf{Y} = N \mathbf{I}_{B \times B}$$

step 1.1: $\mathbf{Y}$-update $\quad \mathbf{Y} = \mathbf{Z}^* \mathbf{W}, \ \mathbf{w}_b = \sqrt{N\sigma_b} \mathbf{\Lambda}^{-\frac{1}{2}} \mathbf{v}_b$

step 1.2: μ-update

step 2: $\mathbf{Y}^*$-update $\quad \mathbf{Y}^* = \mathrm{sgn}(\mathbf{Z}^* \mathbf{W}^*)$, $\mathbf{W}^* = \mathbf{W}\mathbf{R}$

- out-sample-extension

Theorem 1: $z(\mathbf{x}) = [\mu_1^r \mathbf{z}^{(1)}(\mathbf{x}), \dots, \mu_M^r \mathbf{z}^{(M)}(\mathbf{x})]^{\mathrm{T}}$

$$\mathbf{y}^* = sgn(\mathbf{W}^{*\mathrm{T}} z(\mathbf{x}))$$

Complementary Multi-View Tables

Table Complementarity

- sequential learning: for each view the similarities on the inconsistent neighbor pairs will be amplified at next round

$$\hat{\mathbf{S}}_{ij}^{(m)} = \hat{\mathbf{S}}_{ij}^{(m)} \exp(-\alpha^{(m)} \mathbf{P}_{ij})$$

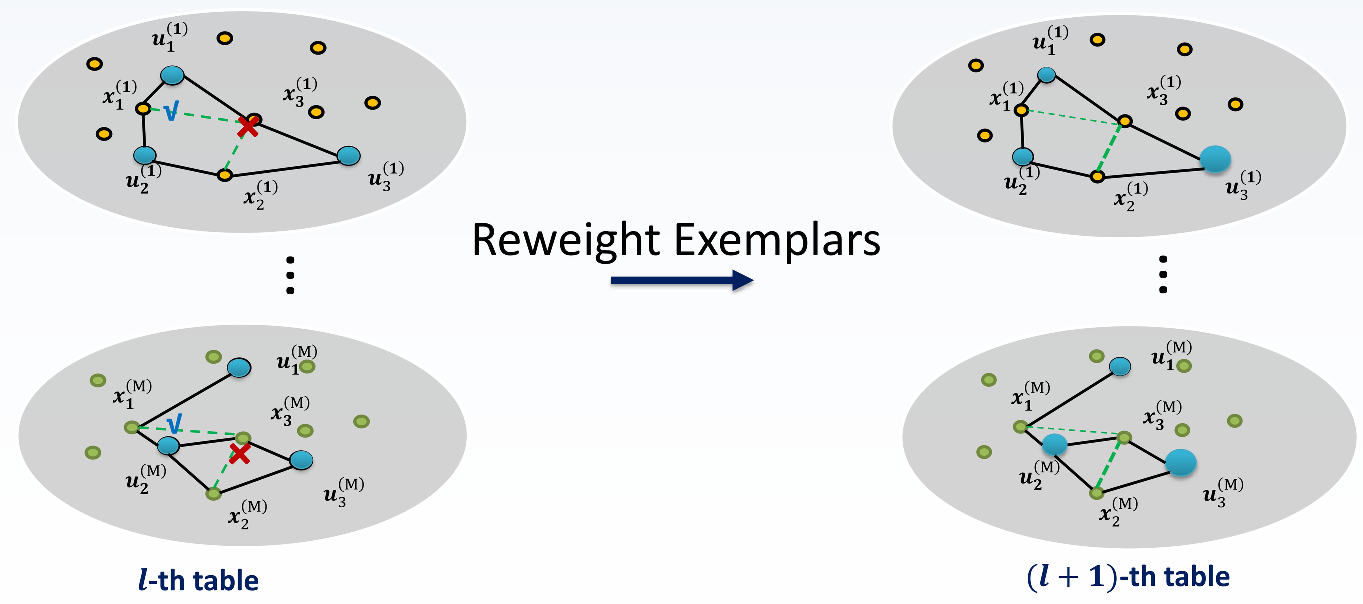
Exemplar Reweighting

- pursue the table complementarity and meanwhile preserve the low-rank similarity: calibrate the role of each exemplar

$$\hat{\mathbf{Z}}^{(m)} = \mathbf{\Gamma}^{-1} \mathbf{Z}^{(m)} \mathbf{\Pi}^{(m)}$$

Proposition 1: Theorem 1 still holds when using the nonlinear feature map based on exemplar reweighing

$$\mathbf{z}^{(m)}(\mathbf{x}) = \frac{[\delta_1^{(m)} \pi_1^{(m)} \mathcal{K}(\mathbf{x}^{(m)}, \mathbf{u}_1^{(m)}), \dots, \delta_K^{(m)} \pi_K^{(m)} \mathcal{K}(\mathbf{x}^{(m)}, \mathbf{u}_K^{(m)})]}{\sum_{k=1}^{K} \delta_k^{(m)} \pi_k^{(m)} \mathcal{K}(\mathbf{x}^{(m)}, \mathbf{u}_k^{(m)})}$$



√ Correct neighbor prediction ✗ Wrong neighbor prediction

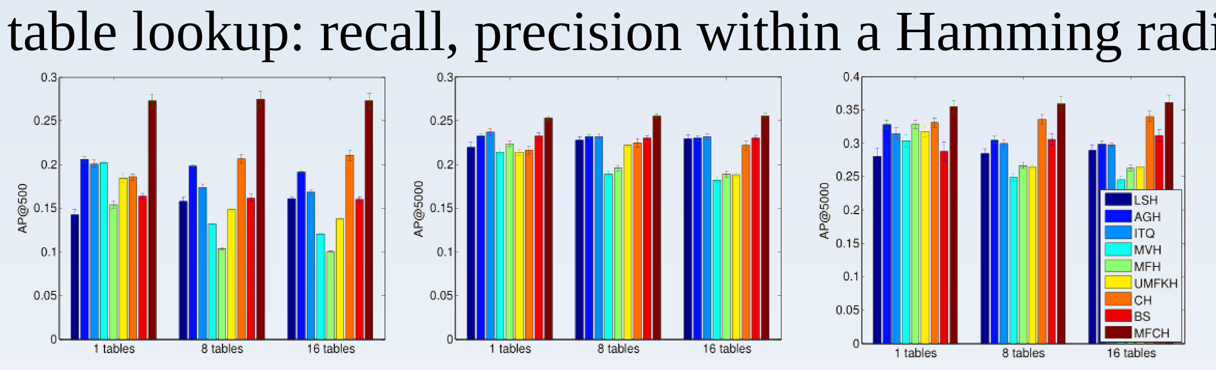
Experiments

Datasets

- CIFAR-10: 60K, 384D GIST + 300D SIFT BoW
- TRECVID: 250K, 512-D GIST + 1000-D spatial pyramid SIFT BoW
- NUS-WIDE: 270K, 128D texture + 225D color + 500-D SIFT BoW

Results

- Hash table lookup: recall, precision within a Hamming radius



(a) AP @500 on CIFAR-10 (b) AP @5000 on TRECVID (c) AP @5000 on NUS-WIDE

- Hamming distance ranking: average precision of the top results

Table 1. Hash table lookup performance of different multi-table methods on CIFAR-10, TRECVID and NUS-WIDE.

	Methods	RH2 $L=1$	RH2 $L=4$	RH2 $L=8$	RH2 $L=16$	PH2 $L=1$	PH2 $L=4$	PH2 $L=8$	PH2 $L=16$
CIFAR-10	LSH	0.71 ±0.11	2.67 ±0.16	5.06 ±0.48	9.07 ±0.25	14.67 ±0.51	15.25 ±0.27	15.01 ±0.35	14.71 ±0.19
	AGH	3.15 ±0.17	4.53 ±0.17	5.63 ±0.16	7.59 ±0.18	20.10 ±0.32	18.28 ±0.14	16.66 ±0.13	14.97 ±0.11
	ITQ	3.49 ±0.19	5.86 ±0.28	7.68 ±0.20	10.81 ±0.15	19.60 ±0.38	16.98 ±0.33	15.92 ±0.28	15.04 ±0.19
	MFH*	4.68 ±0.49	-	-	-	22.67 ±0.12	-	-	-
	MVH	1.58 ±0.03	1.43 ±0.02	2.31 ±0.02	4.03 ±0.02	20.54 ±0.07	15.08 ±0.08	12.94 ±0.04	11.72 ±0.03
	UMFKH	1.10 ±0.07	1.98 ±0.18	3.02 ±0.07	4.50 ±0.04	19.24 ±0.70	15.53 ±0.44	14.15 ±0.17	12.76 ±0.07
	CH	0.71 ±0.03	3.46 ±0.15	5.66 ±0.16	8.73 ±0.33	20.06 ±0.38	19.33 ±0.41	19.06 ±0.37	18.47 ±0.40
	BS	3.05 ±0.55	5.80 ±0.56	7.90 ±0.57	10.68 ±0.59	16.18 ±0.26	15.31 ±0.34	15.01 ±0.32	14.44 ±0.23
	MVCH	**5.55** ±0.20	**9.18** ±0.28	**11.19** ±0.31	**13.32** ±0.39	**26.21** ±0.78	**25.02** ±0.75	**24.38** ±0.69	**23.76** ±0.59
TRECVID	LSH	1.23 ±0.38	4.78 ±0.52	7.64 ±0.40	12.25 ±0.19	22.93 ±0.92	22.67 ±0.93	22.36 ±0.88	22.06 ±0.75
	AGH	4.75 ±0.64	5.67 ±0.70	6.15 ±0.66	7.09 ±0.59	22.97 ±0.42	22.70 ±0.44	22.39 ±0.47	21.77 ±0.48
	ITQ	4.26 ±0.23	8.02 ±0.61	10.90 ±0.69	11.66 ±0.77	23.99 ±0.44	22.89 ±0.55	22.22 ±0.36	21.71 ±0.31
	MFH	1.96 ±0.23	3.73 ±0.26	6.57 ±0.52	9.93 ±0.66	23.11 ±0.27	20.75 ±0.34	19.19 ±0.35	18.42 ±0.40
	MVH	0.73 ±0.03	0.55 ±0.03	0.82 ±0.03	1.36 ±0.03	22.77 ±0.79	21.16 ±0.62	19.47 ±0.49	18.39 ±0.43
	UMFKH	0.36 ±0.03	0.99 ±0.07	1.61 ±0.12	2.00 ±0.06	23.81 ±0.52	22.23 ±0.24	21.46 ±0.19	19.44 ±0.13
	CH	0.43 ±0.01	3.61 ±0.44	6.10 ±0.45	8.92 ±0.61	24.13 ±0.83	22.89 ±0.26	22.14 ±0.26	21.45 ±0.39
	BS	2.66 ±0.39	5.80 ±0.64	8.87 ±1.26	**13.02** ±0.93	23.65 ±0.33	22.75 ±0.28	22.40 ±0.30	21.99 ±0.30
	MVCH	**5.07** ±1.20	**9.09** ±0.77	**10.94** ±0.86	**12.77** ±0.94	**24.46** ±0.62	**23.86** ±0.56	**23.59** ±0.57	**23.34** ±0.58
NUS-WIDE	LSH	0.38 ±0.07	1.29 ±0.08	2.82 ±0.26	5.46 ±0.29	30.32 ±1.01	30.23 ±1.15	30.15 ±1.53	29.71 ±1.52
	AGH	1.31 ±0.04	1.79 ±0.07	2.12 ±0.07	2.82 ±0.09	34.34 ±2.21	32.27 ±2.14	30.62 ±1.89	28.82 ±1.65
	ITQ	2.01 ±0.30	**4.51** ±0.68	**6.25** ±0.27	8.11 ±0.43	33.36 ±1.36	30.05 ±1.00	28.63 ±1.65	28.46 ±1.29
	MFH	0.32 ±0.01	0.57 ±0.01	0.87 ±0.01	1.44 ±0.01	34.50 ±0.79	30.55 ±0.53	28.14 ±0.48	26.51 ±0.45
	MVH	0.15 ±0.01	0.32 ±0.01	0.60 ±0.01	1.15 ±0.01	32.63 ±1.08	26.82 ±0.67	25.36 ±0.58	24.57 ±0.53
	UMFKH	0.31 ±0.01	0.94 ±0.18	1.35 ±0.17	1.63 ±0.04	**36.46** ±0.73	29.71 ±0.28	28.03 ±0.40	26.26 ±0.44
	CH	0.22 ±0.00	3.14 ±0.03	5.18 ±0.21	7.33 ±0.22	36.23 ±1.63	32.96 ±1.52	33.12 ±1.28	32.69 ±1.43
	BS	0.77 ±0.02	1.84 ±0.01	3.01 ±0.04	5.44 ±0.18	33.17 ±1.75	30.25 ±1.02	29.48 ±0.84	29.30 ±0.87
	MVCH	**2.17** ±0.17	4.18 ±1.03	**6.02** ±1.07	**8.95** ±1.15	35.67 ±1.06	**34.88** ±1.03	**34.60** ±1.45	**34.29** ±1.59

* We didn't get a reasonable performance by tuning parameters of MFH on CIFAR-10 when using multiple hash tables. Similar results are shown in Figure 2.

Conclusion

- The First Multi-View Complementary Multi-Table Method
  - Exemplar-based feature fusion: adaptively exploit multi-view information and guarantee the fast computation
  - Exemplar reweighting: eliminate table redundancy in a fast boosting manner

北京航空航天大學 BEIHANG UNIVERSITY
西安電子科技大學 XIDIAN UNIVERSITY
清華大學 Tsinghua University
SKL 国家重点实验室 STATE KEY LABORATORY