Learning to Hash for Visual Big Data

Jiwen Lu

Department of Automation, Tsinghua University, China

http://ivg.au.tsinghua.edu.cn/Jiwen_Lu/

Outline

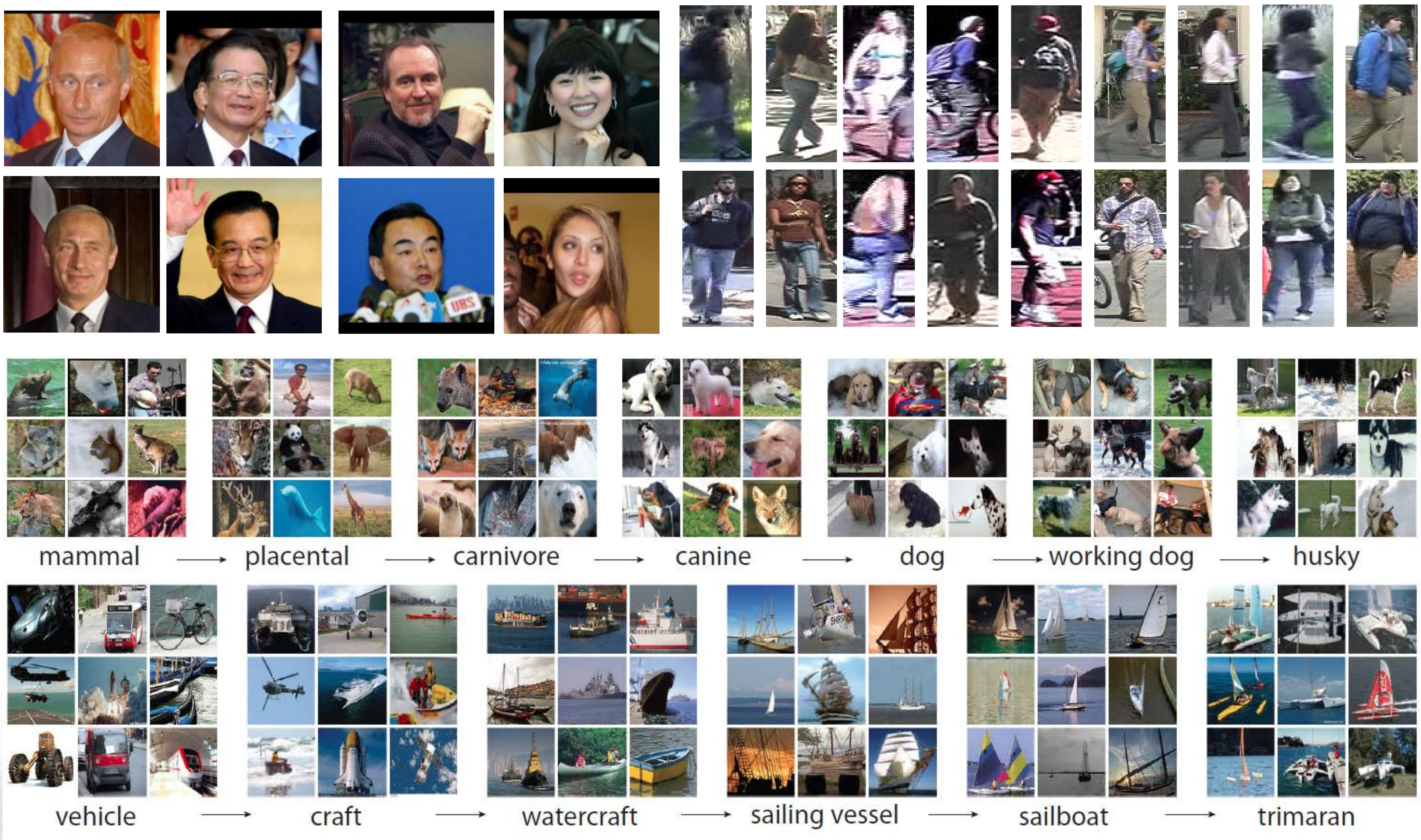
- Introduction
- Learning to Hashing for Visual Recognition
- Learning to Hashing for Visual Search
- Conclusions and Future Works

Visual Big Data

- Visual big data is important and useful in many real-world applications.
- Multi-source/Heterogeneous/Dynamic/Sparse

清華大學 Tsinghua University

Visual Recognition



Visual Search

Learning to Hash for Visual Recognition

- A conventional visual recognition system
  - Offline: training model, gallery feature extraction, storage
  - Online: probe feature extraction, matching

Learning to Hash for Visual Recognition

- Binary descriptors present high storage efficiency and matching speed
- Efficient storage
  - Real-valued descriptors -> Binary codes
- Fast matching
  - Euclidean distance -> Hamming distance

Learning to Hash for Visual Recognition

- Local Binary Feature Descriptor: LBP

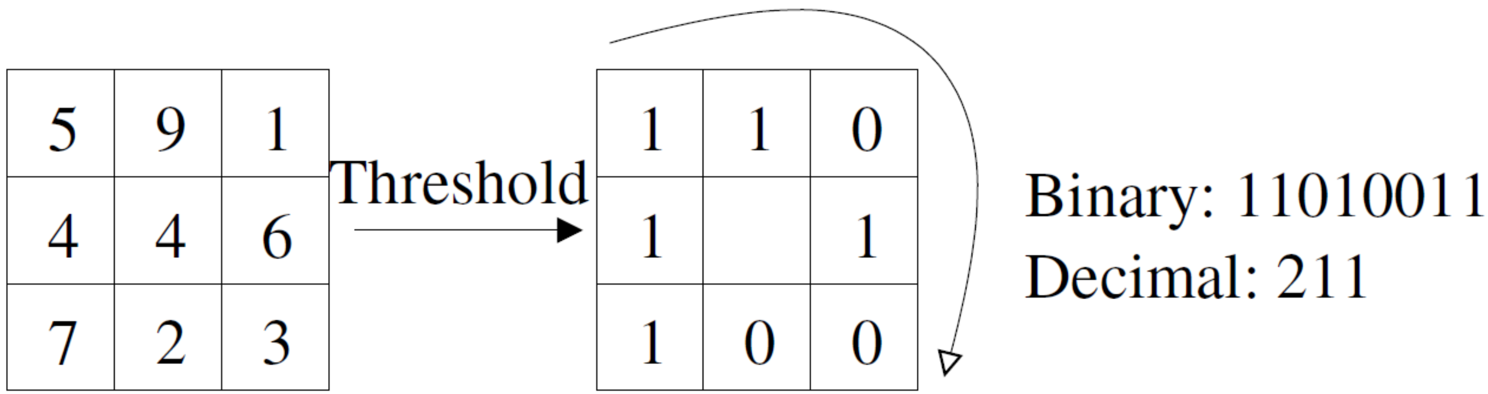
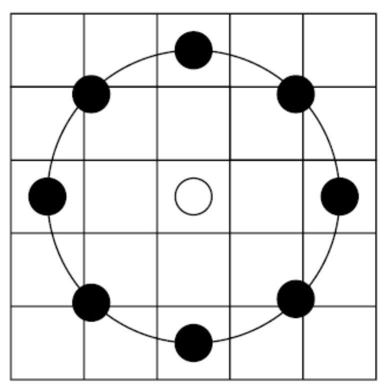


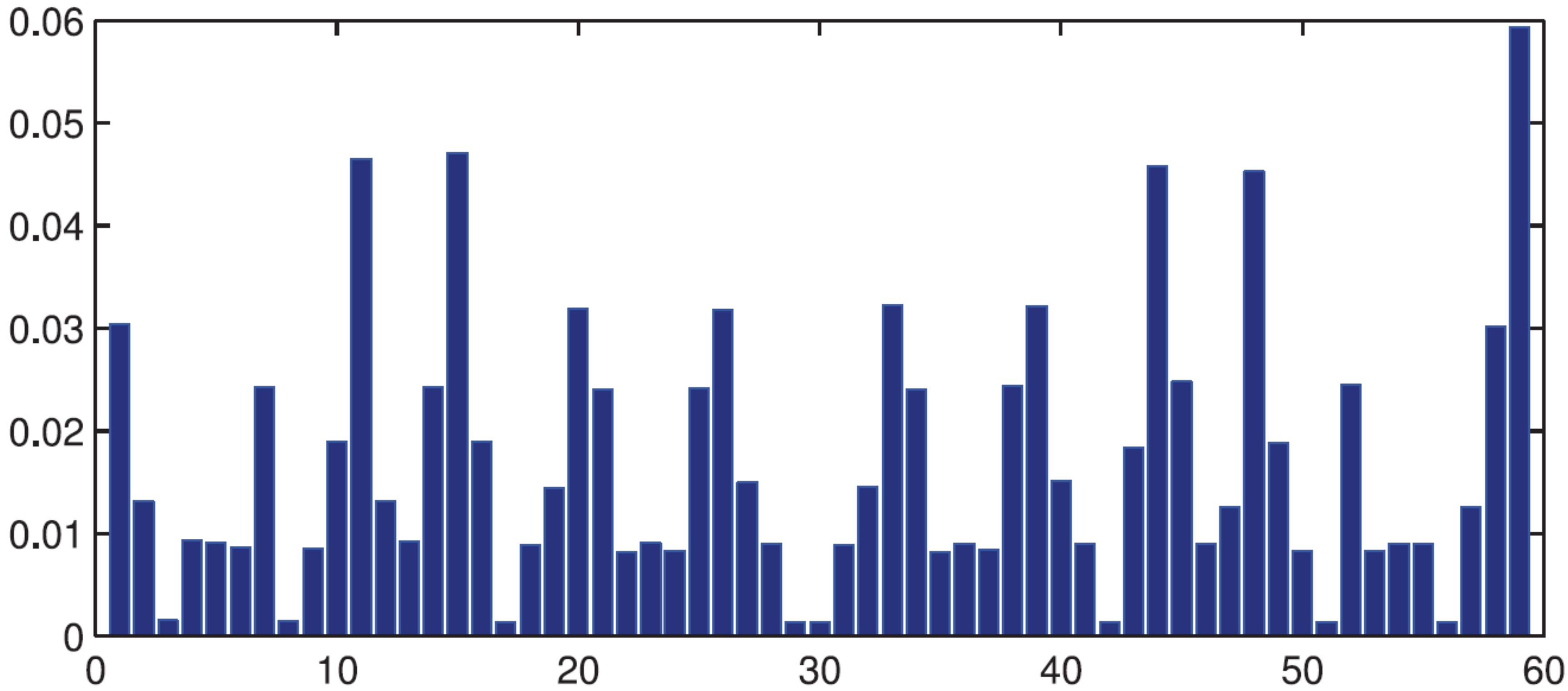
Fig. 1. The basic LBP operator.



[Ahonen et al, ECCV 2004]

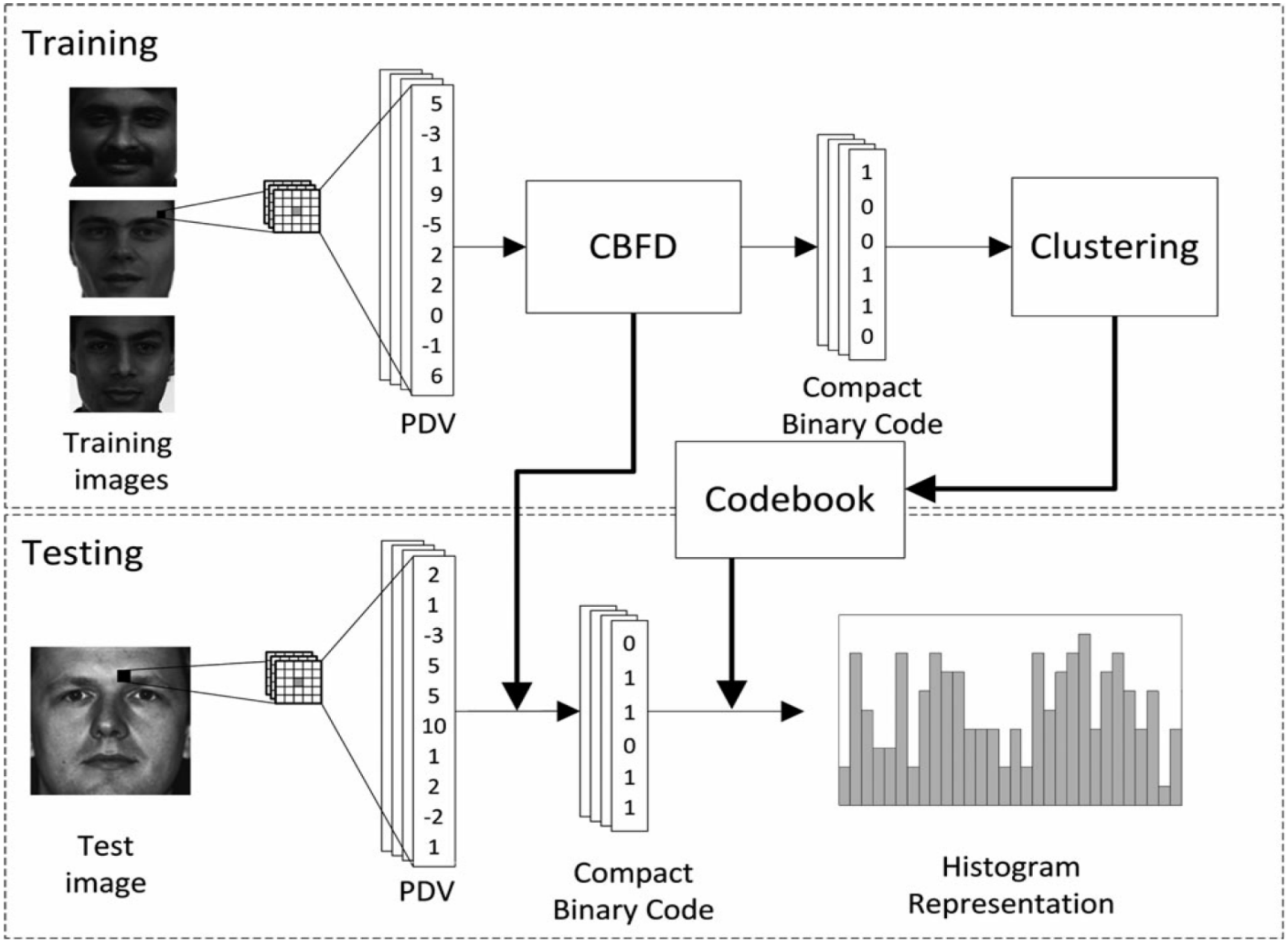
Learning to Hash for Visual Recognition

Bin distribution of LBP



Bin distributions in the LBP histogram in the FERET training set.

Learning to Hash for Visual Recognition



[1] **Jiwen Lu**, Venice Erin Liong, Xiuzhuang Zhou, and Jie Zhou, Learning compact binary face descriptor for face recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 37, no. 10, pp. 2041-2056, 2015.

Learning to Hash for Visual Recognition

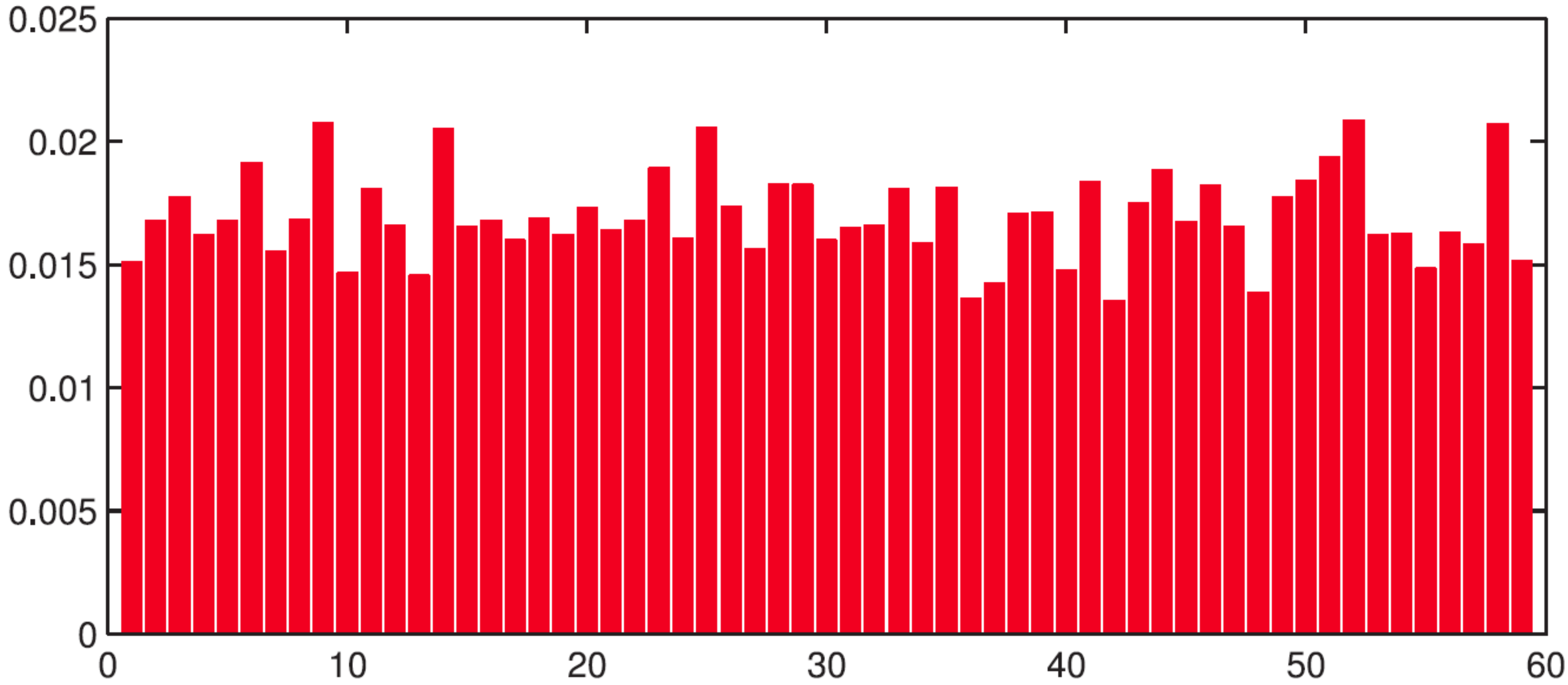
Objective

$$
\begin{aligned}
\min_{w_k} J(w_k) &= J_1(w_k) + \lambda_1 J_2(w_k) + \lambda_2 J_3(w_k) \\
&= -\sum_{n=1}^{N} \|b_{nk} - \mu_k\|^2 \\
&\quad + \lambda_1 \sum_{n=1}^{N} \|(b_{nk} - 0.5) - w_k^T x_n\|^2 \\
&\quad + \lambda_2 \|\sum_{n=1}^{N} (b_{nk} - 0.5)\|^2,
\end{aligned}
$$

- First term: redundancy removing
- Second term: energy preserving
- Third term: balanced bin

Learning to Hash for Visual Recognition

Bin distribution of CBFD



Bin distributions in the CBFD histogram in the FERET training set.

Learning to Hash for Visual Recognition

Comparisons with state-of-the-arts

Method	Accuracy
CSML+SVM, aligned+WPCA [43]	88.00 ± 0.37
PAF [67]	87.77 ± 0.51
SFRD+PMML+WPCA [11]	89.35 ± 0.50
Sub-SML [7]	89.73 ± 0.38
VMRS+WPCA [3]	91.10 ± 0.59
DDML+WPCA [19]	90.68 ± 1.41
CBFD+WPCA(a)	87.33 ± 2.42
CBFD+WPCA(b)	87.57 ± 1.43
CBFD+WPCA(c)	87.23 ± 1.68
CBFD+WPCA(mean: a, b, c)	89.05 ± 1.51
CBFD+WPCA(svm: a, b, c)	$\mathbf{89.07 \pm 1.51}$
CBFD+WPCA(combine)	$\mathbf{92.62 \pm 1.08}$

Method	Accuracy
Combined Joint Bayesian [9]	90.90 ± 1.48
Sub-SML [7]	90.75 ± 0.64
ConvNet - RBM [53]	91.75 ± 0.48
VMRS+WPCA [3]	92.05 ± 0.45
Fisher vector faces+WPCA [52]	93.03 ± 1.05
High-dim LBP [10]	93.18 ± 1.07
CBFD+WPCA(a)	87.87 ± 1.86
CBFD+WPCA(b)	88.90 ± 1.81
CBFD+WPCA(c)	88.35 ± 1.61
CBFD+WPCA(mean: a, b, c)	90.90 ± 1.40
CBFD+WPCA(svm: a, b, c)	$\mathbf{90.75 \pm 1.10}$
CBFD+WPCA(combine)	$\mathbf{93.80 \pm 1.31}$

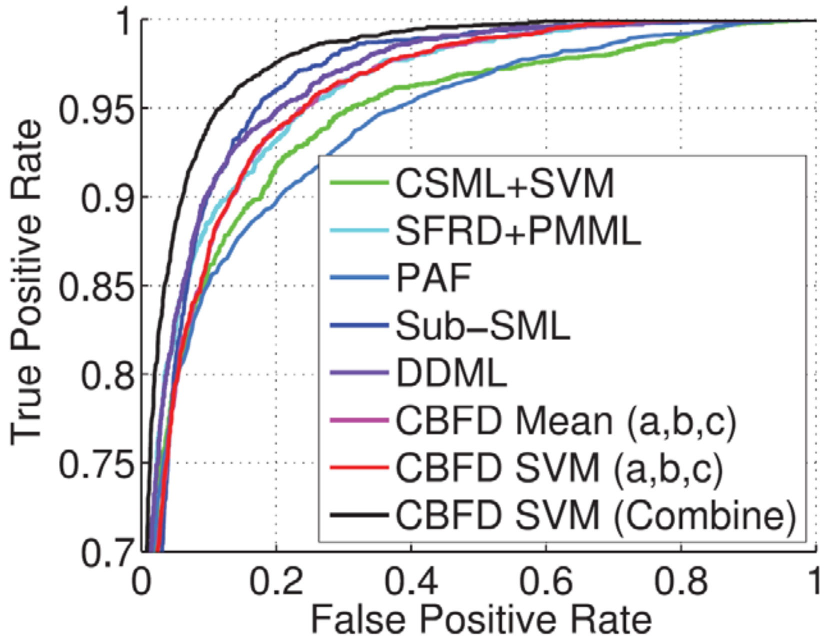


image-restricted setting

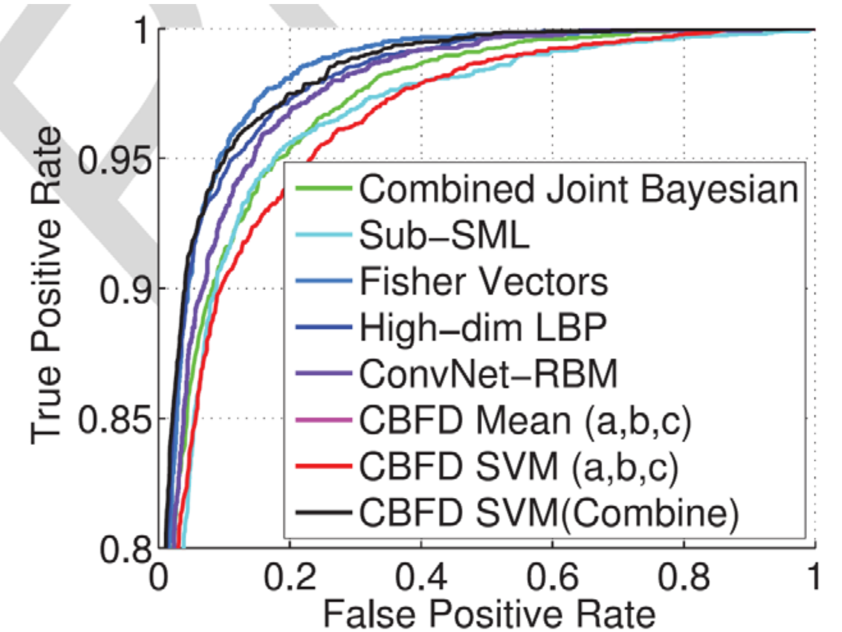
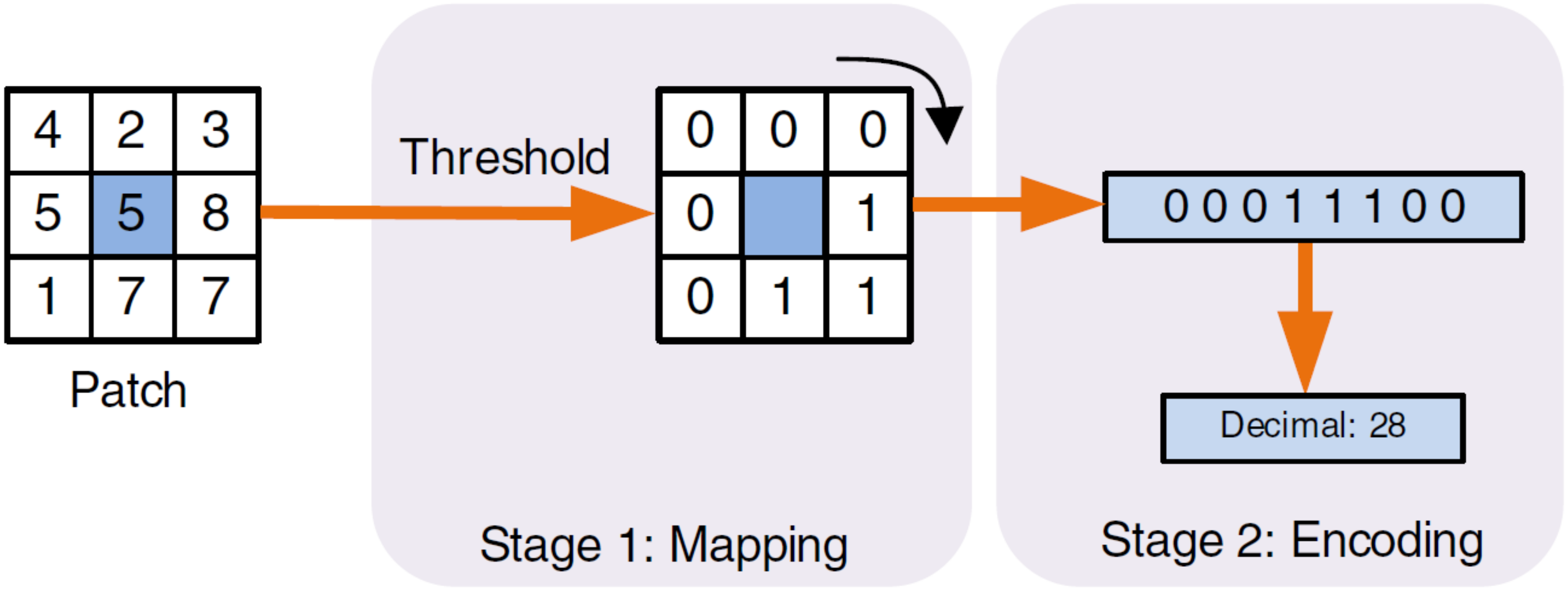


image-unrestricted setting

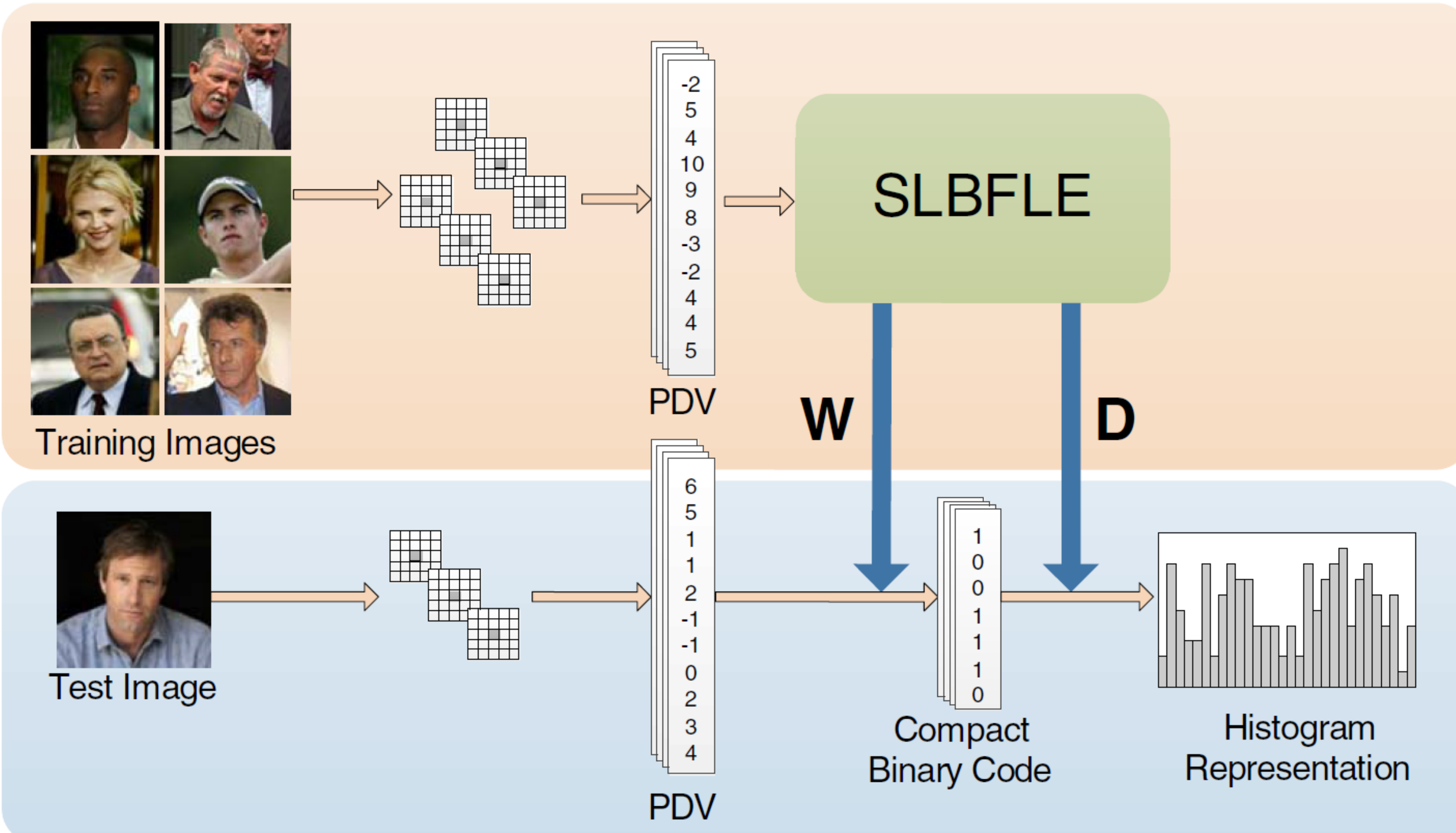
Learning to Hash for Visual Recognition

Two-step procedure in LBP



Learning to Hash for Visual Recognition

SLBFLE

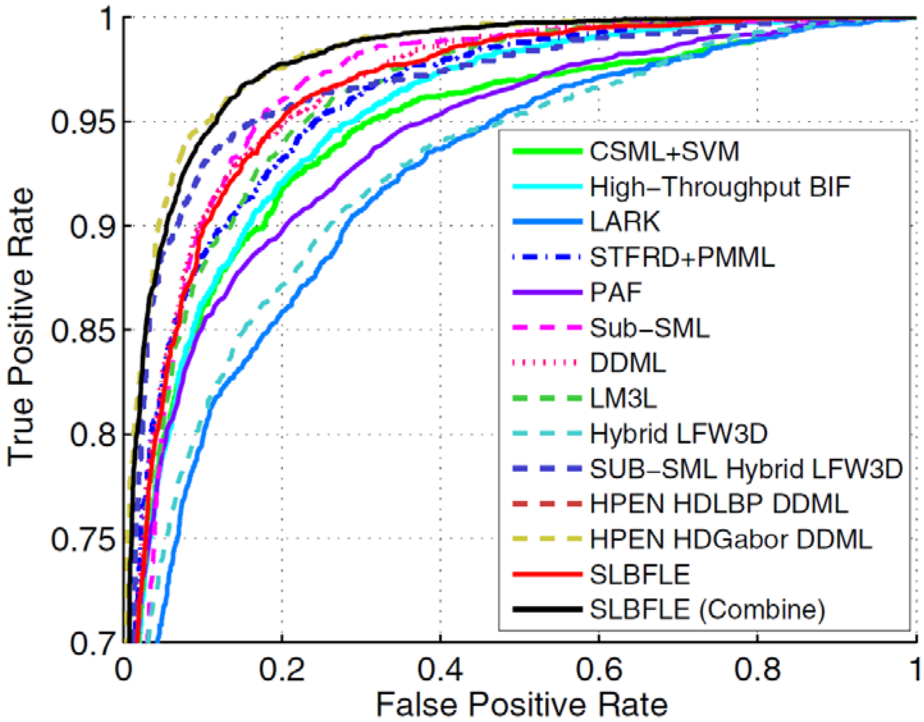


[2] **Jiwen Lu**, Venice Erin Liong, and Jie Zhou, Simultaneous local binary feature learning and encoding for face recognition, *ICCV*, pp. 3721-3729, 2015.

Learning to Hash for Visual Recognition

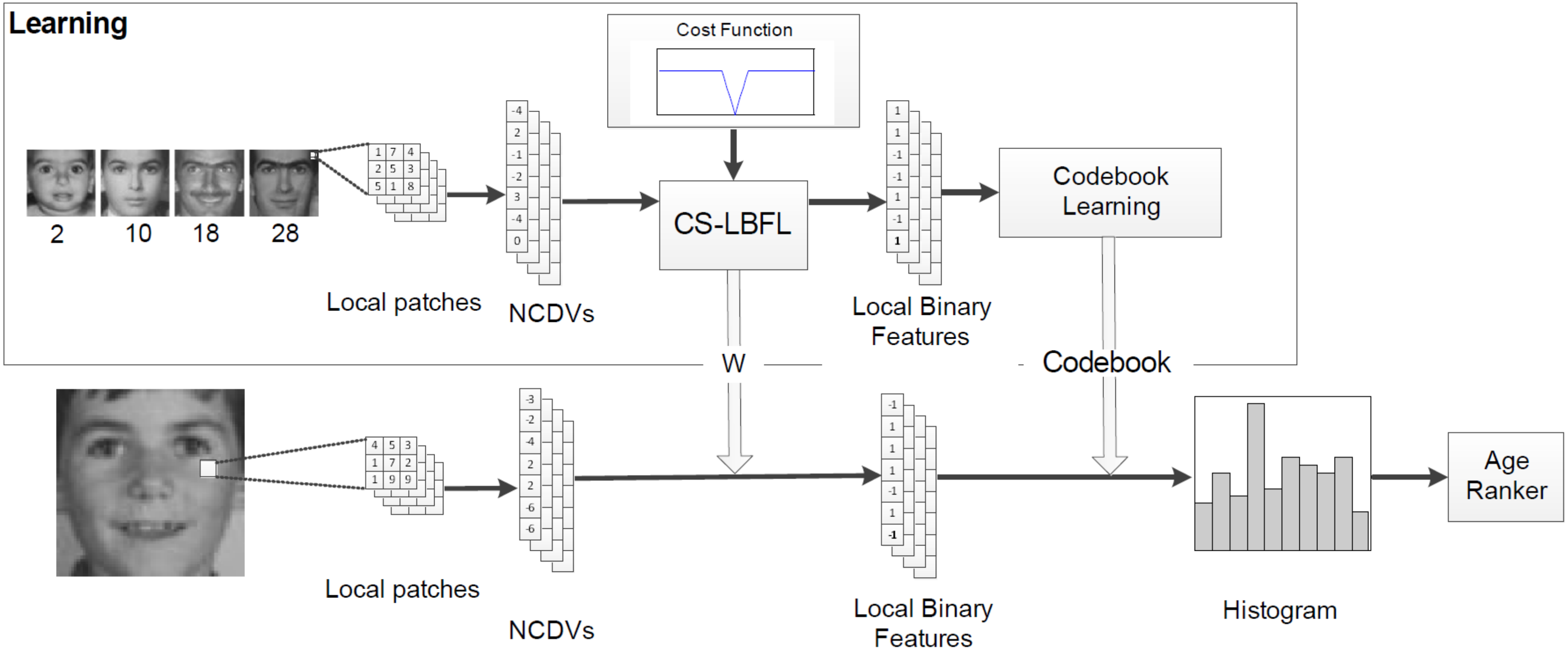
Results on LFW

Method	Accuracy
CSML+SVM [30]	88.00 ± 0.37
High-Throughput BIF [10]	88.13 ± 0.58
LARK supervised [32]	85.10 ± 0.59
DML-eig combined [51]	85.65 ± 0.56
Covolutional DBN [18]	87.77 ± 0.62
STFRD+PMML [11]	89.35 ± 0.50
PAF [50]	87.77 ± 0.51
Sub-SML [7]	89.90 ± 0.38
VMRS [3]	91.10 ± 0.59
DDML [15]	90.68 ± 1.41
LM3L [16]	89.57 ± 0.02
Hybrid on LFW3D [13]	85.63 ± 0.005
Sub-SML + Hybrid on LFW3D [13]	91.65 ± 0.01
HPEN + HD-LBP + DDML [56]	92.57 ± 0.003
HPEN + HD-Gabo + DDML [56]	92.80 ± 0.005
SLBFLE (R=2)	85.62 ± 1.41
SLBFLE (R=3)	86.57 ± 1.65
SLBFLE (R=4)	87.45 ± 1.28
SLBFLE (R=2+3+4)	90.18 ± 1.89
SLBFLE (All combined)	**92.97 ± 1.20**



Learning to Hash for Visual Recognition

CS-LBFL

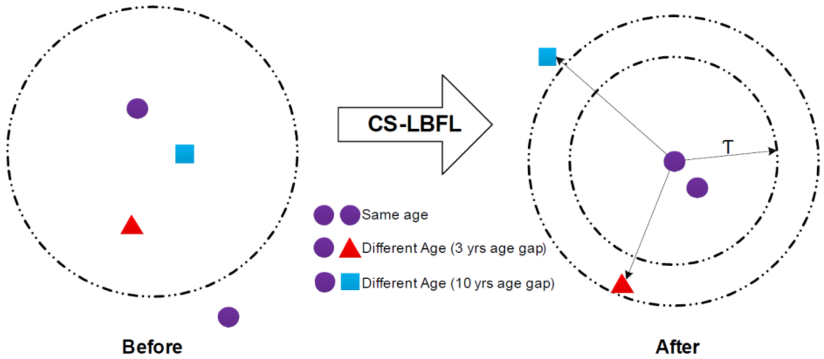


[3] **Jiwen Lu**, Venice Erin Liong, and Jie Zhou, Cost-sensitive local binary feature learning for facial age estimation, *IEEE Transactions on Image Processing,* vol. 24, no. 12, pp. 5356-5368, 2015.

Learning to Hash for Visual Recognition

CS-LBFL

$$\min_{W} J = \sum_{m=1}^{M}(1-\ell_m(\tau - d(b(x_{1m}), b(x_{2m}))) \times Q(y_{1m}, y_{2m})),$$



- First term: large margin
- Second term: cost-sensitive

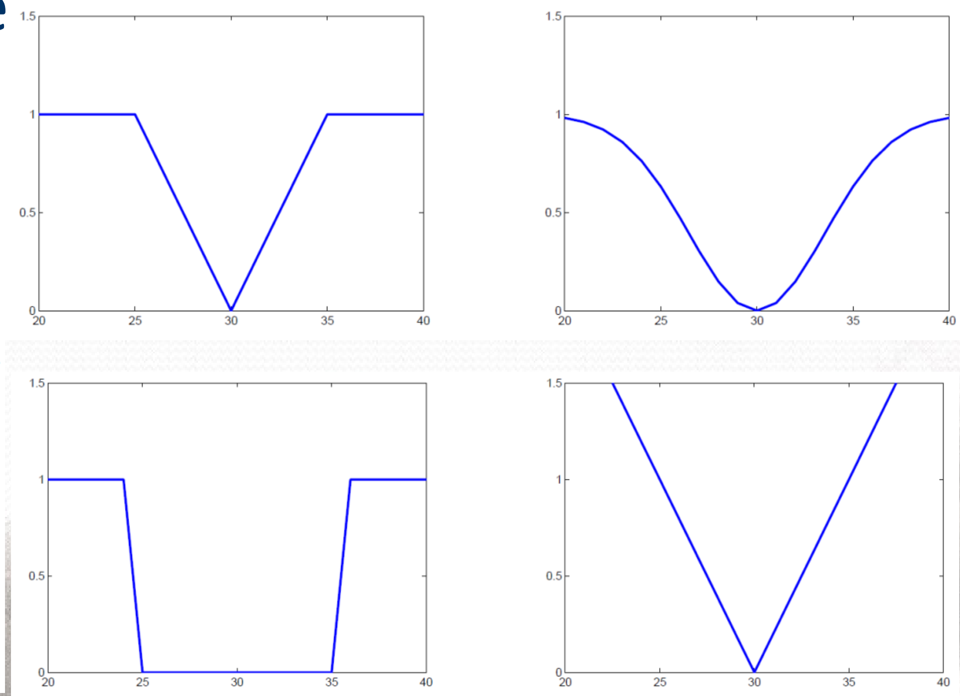
Cost function

$$Q(c_1, c_2) = \begin{cases} \frac{|c_2-c_1|}{L}, & |c_1-c_2| \leq L \\ 1, & \text{otherwise} \end{cases}$$

$$Q(c_1, c_2) = \begin{cases} 1-\exp^{\frac{-(c_2-c_1)^2}{L^2}}, & |c_1-c_2| \leq L \\ 1, & \text{otherwise} \end{cases}$$

$$Q(c_1, c_2) = \frac{|c_2-c_1|}{L}$$

$$Q(c_1, c_2) = \begin{cases} 0, & |c_1-c_2| \leq L \\ 1, & \text{otherwise} \end{cases}$$



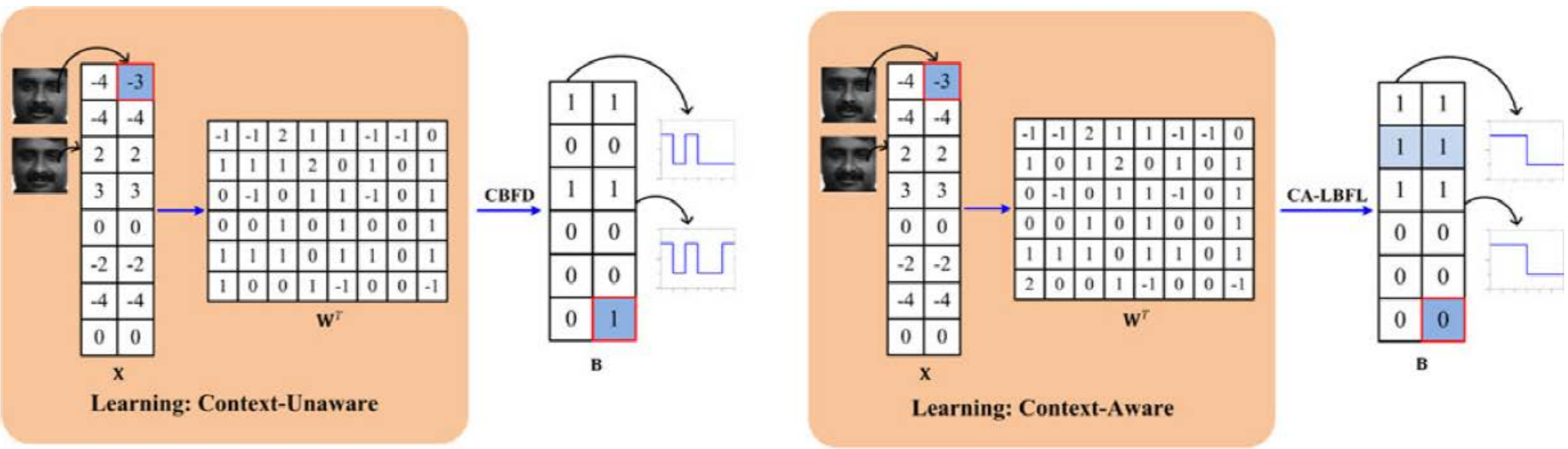
Learning to Hash for Visual Recognition

Comparisons with state-of-the-arts

Method	MAE	Method description	year
KNN	8.24		
SVM	7.25		
MLP	6.95		
RUN [60]	5.78	AAM + RUN	2007
AGES [16]	6.77	AAM + Aging pattern subspace	2007
LARR [18]	5.07	AAM + Locally adjusted robust regression	2008
PFA [19]	4.97	AAM + Probabilistic fusion approach	2008
KAGES [14]	6.18	AAM + Kernel AGES	2008
MSA [13]	5.36	AAM + Multilinear subspace analysis	2009
SSE [59]	5.21	AAM + submanifold embedding	2009
mKNN [57]	5.21	AAM + metric learning	2009
MTWGP [63]	4.83	AAM + Multi-task warped GPR	2010
RED-SVM [4]	5.21	AAM + Red SVM	2010
OHRank [5]	4.48	AAM + Ordinal hyperplanes ranker	2011
PLO [36]	4.82	Feature selection + OHRank	2012
IIS-LLD [15]	5.77	AAM/BIF+learning from label distribution	2013
CPNN [15]	4.76	AAM/BIF+learning from label distribution	2013
CA-SVR [6]	4.67	AAM+cumulative/joint attribute learning	2013
CS-LBFL	**4.43**	Feature learning + OHRank	
CS-LBMFL	**4.36**	Multiple feature learning + OHRank	

Learning to Hash for Visual Recognition

- Motivation
  - Exploit contextual information of binary codes as strong prior knowledge to enhance the robustness



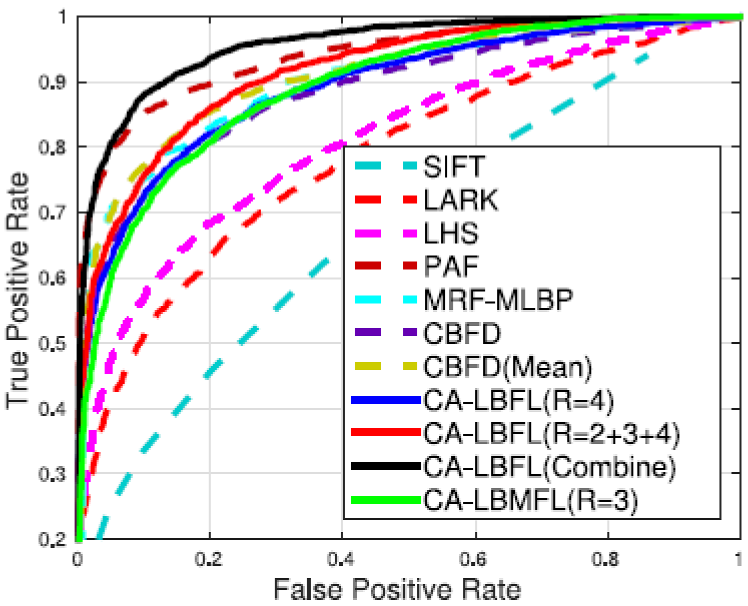
(a) (b)

[4] Yueqi Duan, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou, Context-aware local binary feature learning for face recognition, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2017, accepted.

Learning to Hash for Visual Recognition

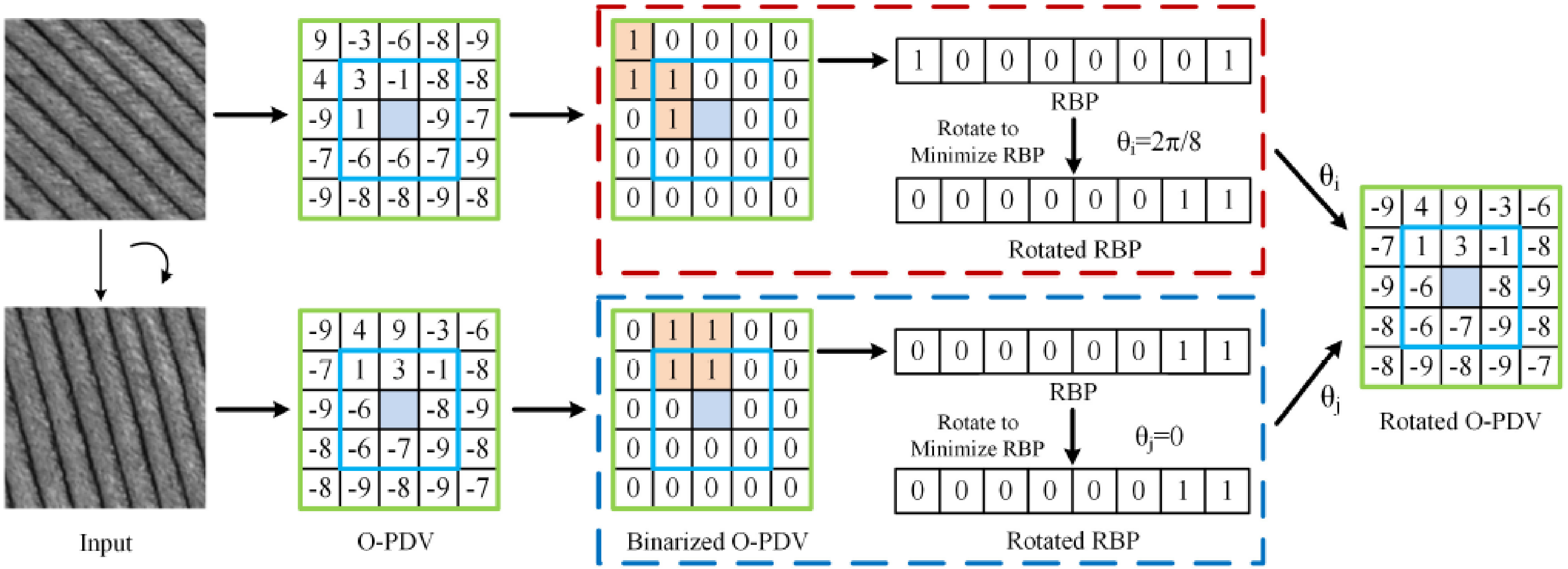
- LFW

Method	VR	AUC
LBP [1]	69.45	75.47
SIFT [41]	64.10	54.07
LARK [57]	72.23	78.30
POEM [72]	75.22	-
LHS [59]	73.40	81.07
MRF-MLBP [3]	80.08	89.94
PEM (LBP) [36]	81.10	-
PEM (SIFT) [36]	81.38	-
DFD [34]	84.02	-
High-dim LBP [13]	84.08	-
PAF [77]	-	94.05
CBFD [43]	-	88.65
CA-LBFL (R=2)	81.50	86.44
CA-LBFL (R=3)	82.97	88.92
CA-LBFL (R=4)	83.30	89.24
CA-LBFL (R=2+3+4)	84.72	91.66
CA-LBFL (combine)	**86.57**	**95.67**
CA-LBMFL (R=3)	83.22	89.26



Learning to Hash for Visual Recognition

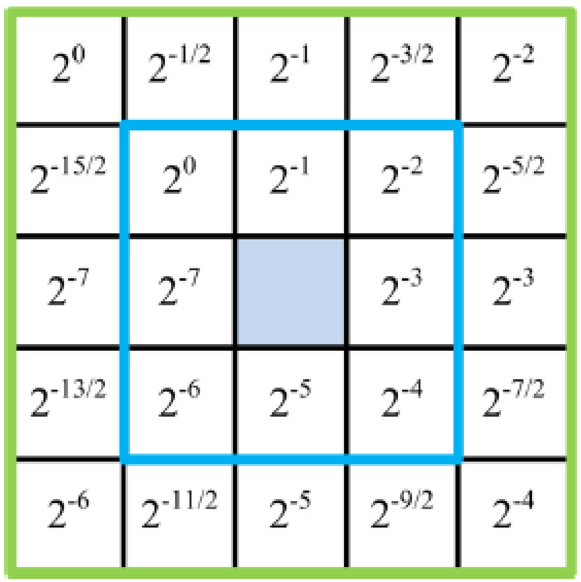
- Rotation-invariance

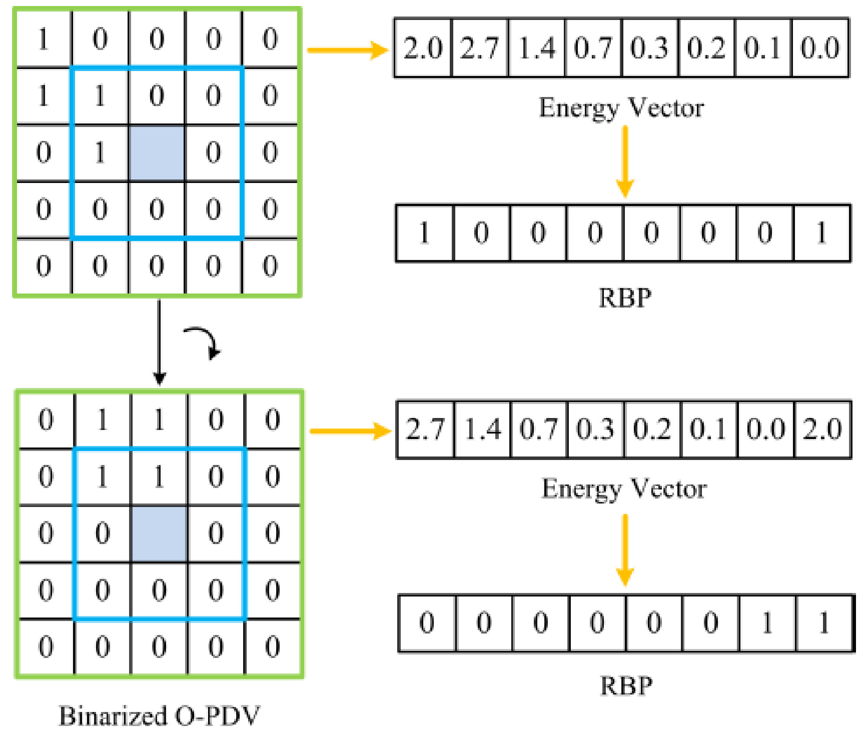


[5] Yueqi Duan, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou, Learning rotation-invariant local binary descriptor, *IEEE Trans. on Image Processing,* vol. 26, no. 8, pp. 3636-3651, 2017.

Learning to Hash for Visual Recognition

- RBP: Describe the circular changing tendency of a local patch

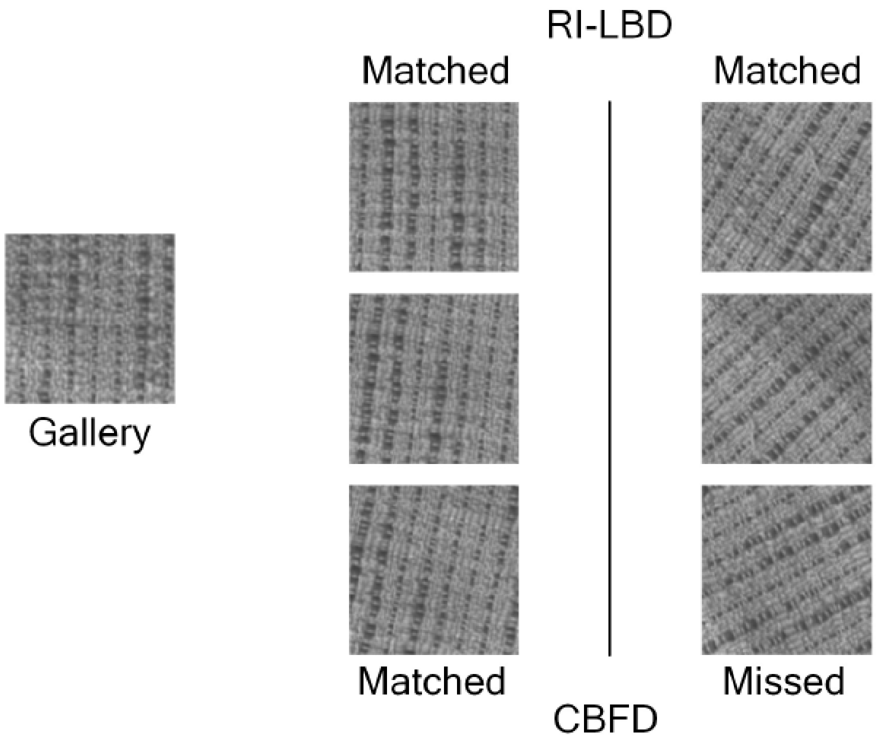




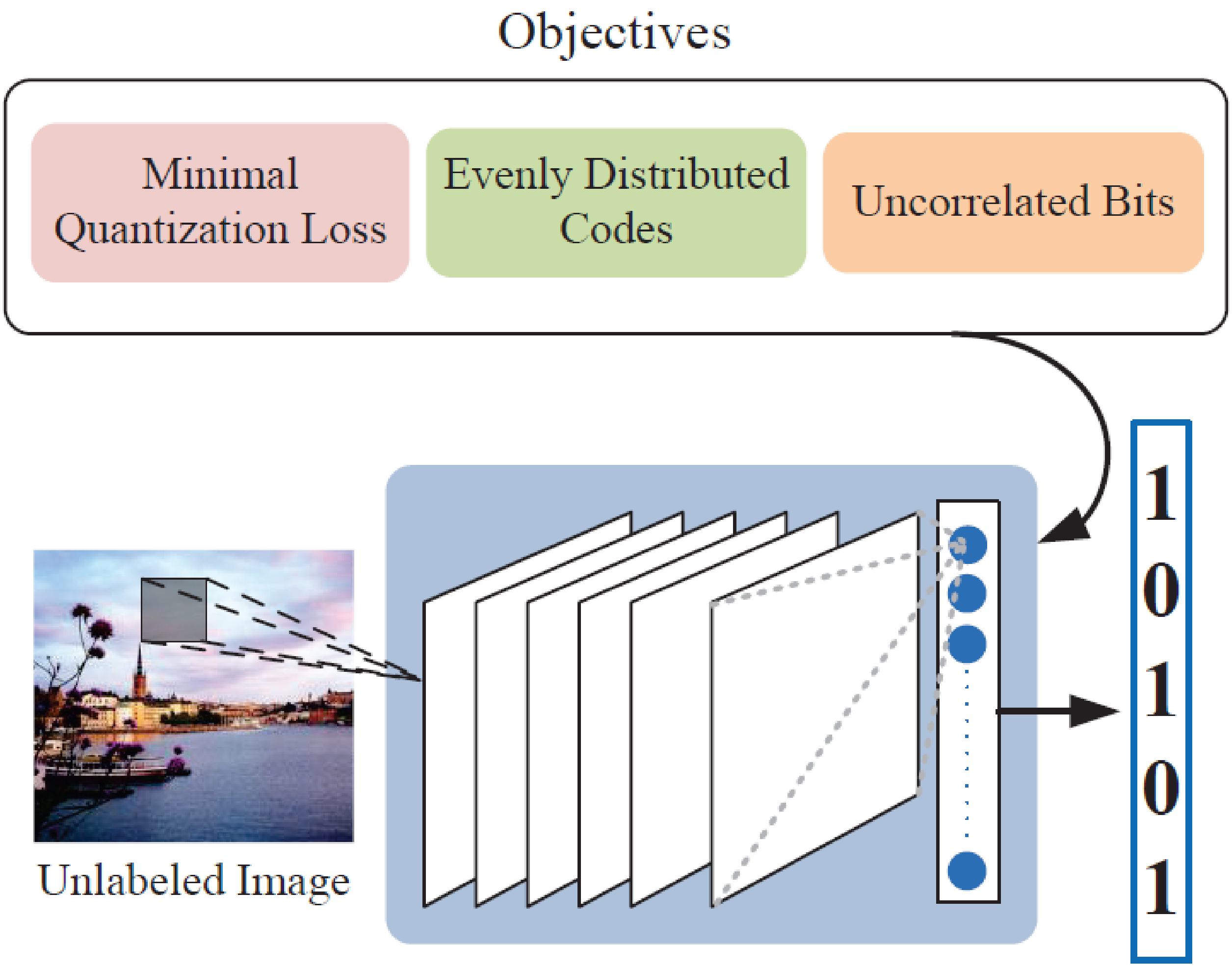
Learning to Hash for Visual Recognition

- Outex-TC12

Method	Outex_TC12		
	5×5	7×7	9×9
LBP [6]	82.07	86.79	89.64
CBFD [2]	63.27	68.84	72.66
LTP [15]	86.46	90.88	92.08
RLBP [50]	83.45	88.01	90.73
NTLBP [51]	82.21	88.28	91.61
DLBP [52]	74.11	81.24	85.53
CLBP [48]	94.48	95.67	95.78
CLBC [53]	93.98	95.17	95.75
BRINT [54]	94.29	96.28	97.16
LBPV [16]	89.59	94.16	95.80
MRELBP [8]	**96.24**	-	**99.03**
RI-LBD	90.06	92.85	94.12
RI-LBD (Combine)	95.04	**96.90**	97.77
TRICo-LBD	92.38	94.96	95.63

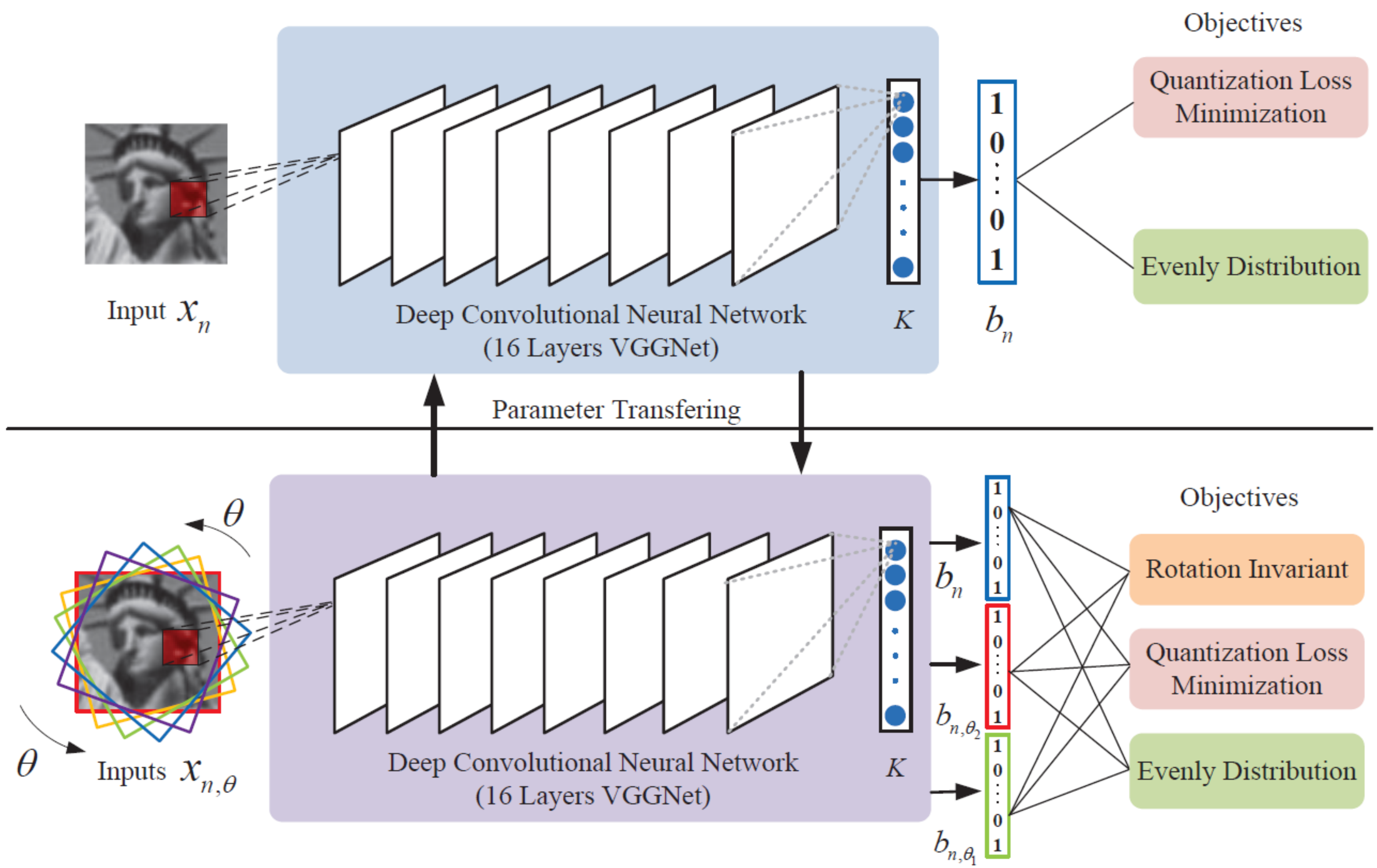


Learning to Hash for Visual Recognition



[6] Kevin Lin, **Jiwen Lu**, Chu-Song Chen, and Jie Zhou, Learning compact binary descriptors with unsupervised deep neural networks, *CVPR*, pp. 1183-1192, 2016.

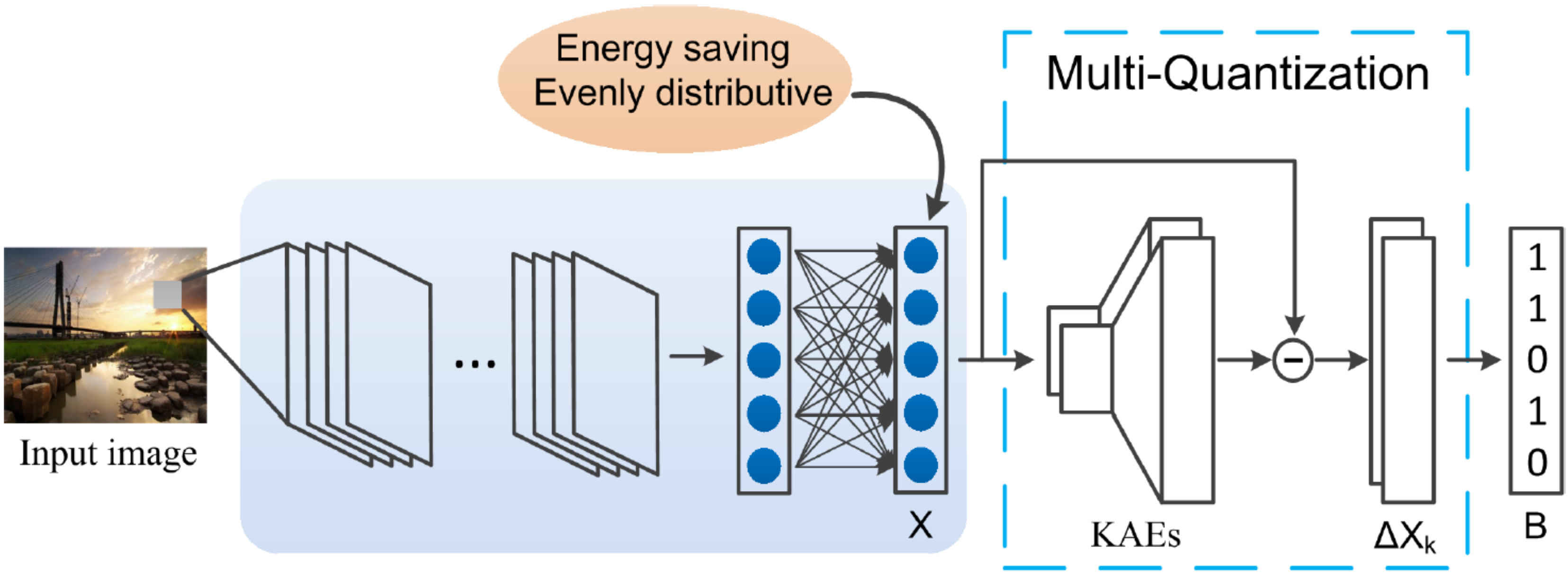
Learning to Hash for Visual Recognition



Learning to Hash for Visual Recognition

Train	Test	Real-valued	Binary						
		SIFT [26] 128 bytes	Boosted SSC [35] 16 bytes	BRISK [22] 64 bytes	ORB [33] 32 bytes	BRIEF [6] 32 bytes	LDAHash [38] 16 bytes	D-BRIEF [41] 4 bytes	DeepBit 32 bytes
Yosemite	Notredame	28.09	72.20	74.88	54.57	54.57	51.58	43.96	**29.60**
Yosemite	Liberty	36.27	71.59	79.36	59.15	59.15	49.66	53.39	**34.41**
Notredame	Yosemite	29.15	76.00	73.21	54.96	54.96	52.95	**46.22**	63.68
Notredame	Liberty	36.27	70.35	79.36	59.15	59.15	49.66	51.30	**32.06**
Liberty	Notredame	28.09	72.95	74.88	54.57	54.57	51.58	43.10	**26.66**
Liberty	Yosemite	29.15	77.99	73.21	54.96	54.96	52.95	**47.29**	57.61
Average 95% ERR		31.17	73.51	75.81	56.23	56.23	51.40	47.54	**40.67**

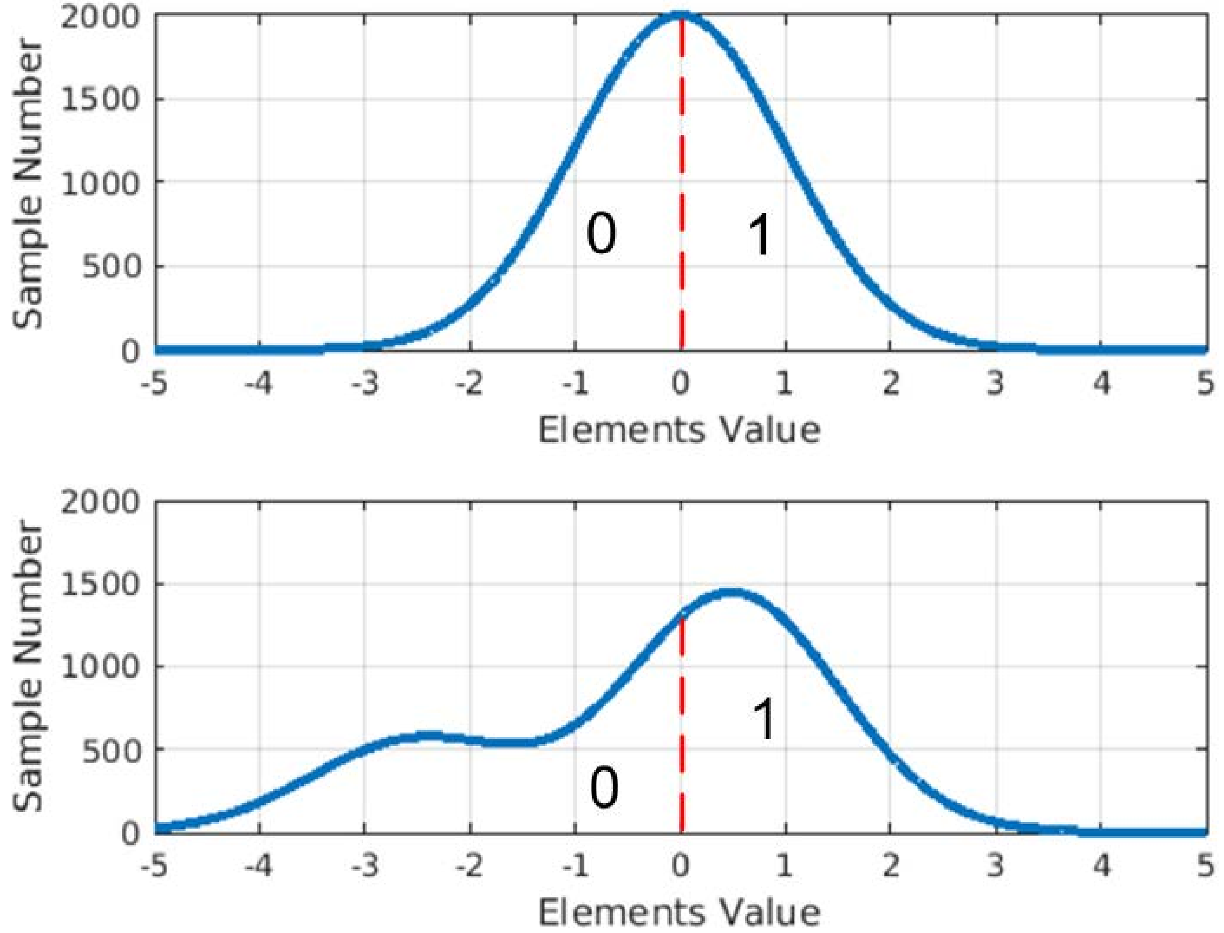
Learning to Hash for Visual Recognition



[7] Yueqi Duan, **Jiwen Lu**, Ziwei Wang, Jianjiang Feng, and Jie Zhou, Learning deep binary descriptor with multi-quantization, *CVPR*, 2017, accepted.

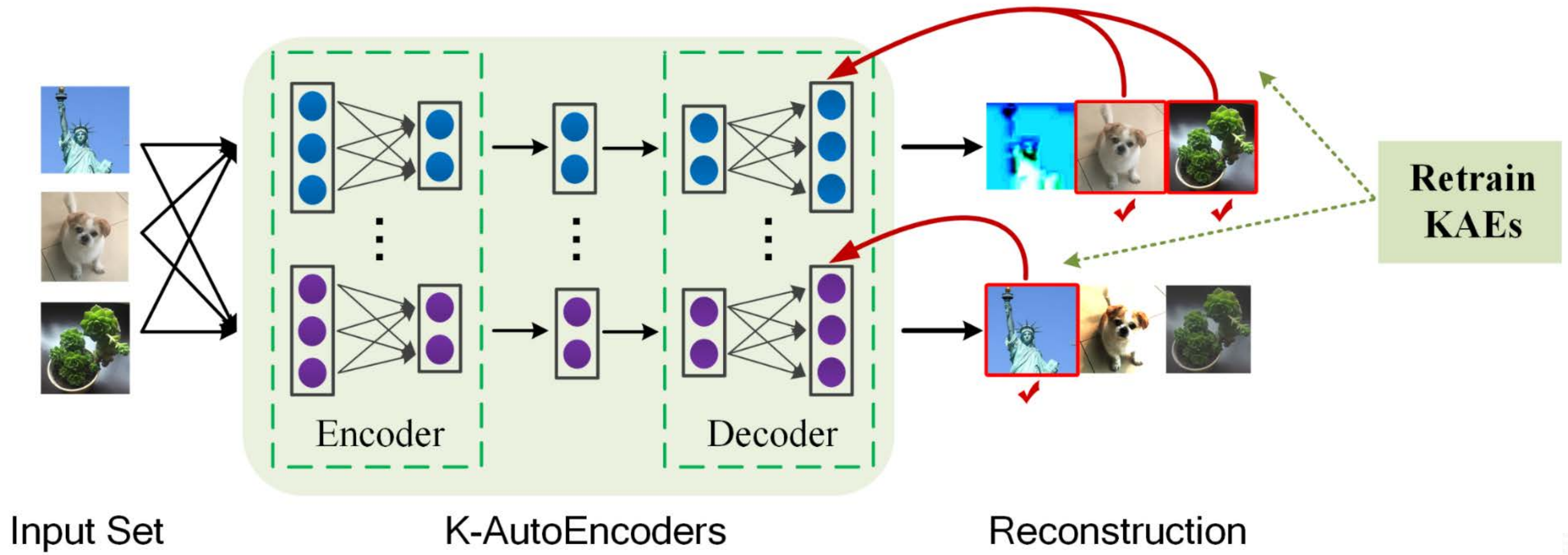
Learning to Hash for Visual Recognition

- Sign function ignores data distributions



Learning to Hash for Visual Recognition

- Train K-autoencoders (KAEs) with an iterative two-step procedure



Learning to Hash for Visual Recognition

- Brown

Train Test	Yosemite Noter Dame	Yosemite Liberty	Notre Dame Yosemite	Notre Dame Liberty	Liberty Notre Dame	Liberty Yosemite	Average ERR
SIFT [27] (128 bytes)	28.09	36.27	29.15	36.27	28.09	29.15	31.17
Boosted SSC [40] (16 bytes)	72.20	71.59	76.00	70.35	72.95	77.99	73.51
BRISK [25] (64 bytes)	74.88	79.36	73.21	79.36	74.88	73.21	75.81
BRIEF [6] (32 bytes)	54.57	59.15	**54.96**	59.15	54.57	**54.96**	56.23
DeepBit [26] (32 bytes)	29.60	34.41	63.68	32.06	26.66	57.61	40.67
LDAHash [42] (16 bytes)	51.58	49.66	52.95	49.66	51.58	52.95	51.40
D-BRIEF [47] (4 bytes)	43.96	53.39	46.22	51.30	43.10	47.29	47.54
BinBoost [45] (8 bytes)	14.54	21.67	18.96	20.49	16.90	22.88	19.24
RFD [10] (50-70 bytes)	11.68	19.40	14.50	19.35	13.23	16.99	15.86
DBD-MQ (32 bytes)	**27.20**	**33.11**	57.24	**31.10**	**25.78**	57.15	**38.59**

Outline

- Introduction
- Learning to Hashing for Visual Recognition
- Learning to Hashing for Visual Search
- Conclusions and Future Works

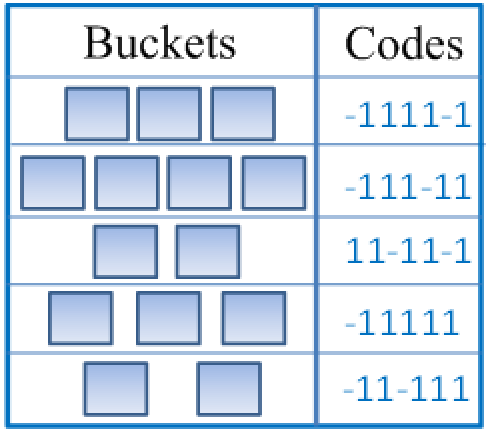
Learning to Hashing for Visual Search

Tsinghua University

- Image and Video search
  - Find most similar images/videos
- Search engine
- Collaborative filtering
- Product search
- Medical search
- Person re-identification

Learning to Hashing for Visual Search

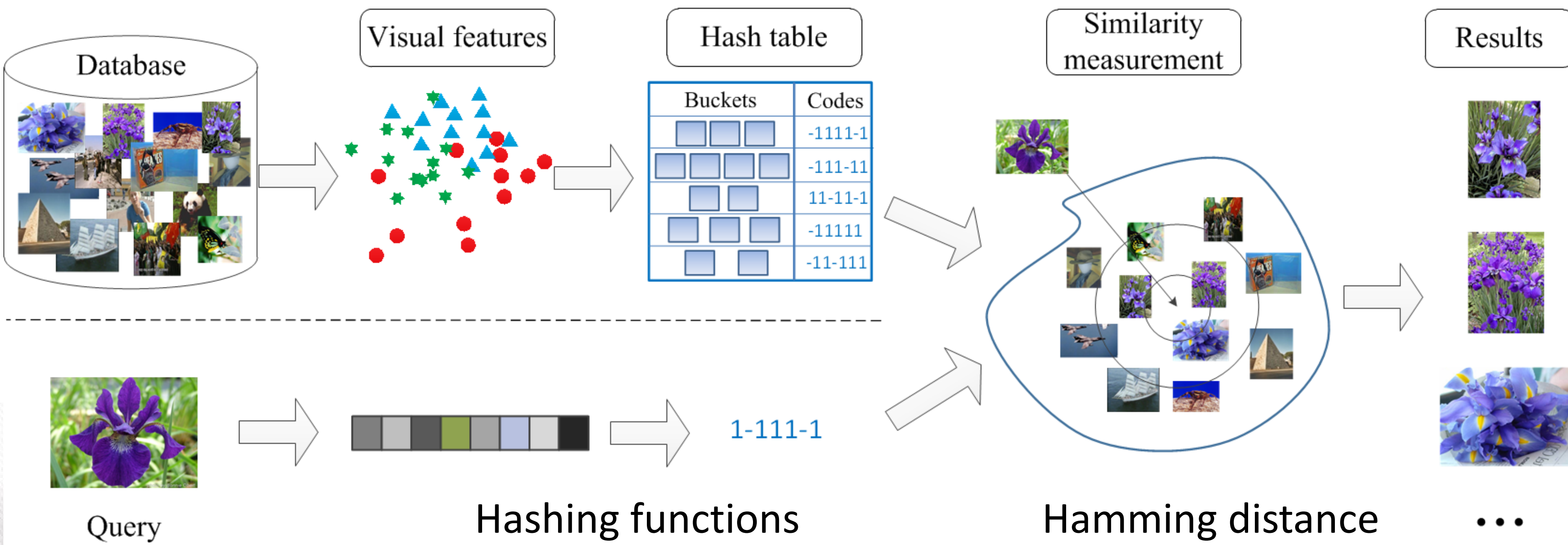
- Similarity measurement
  - Hamming distance
- Storage
  - Short binary codes



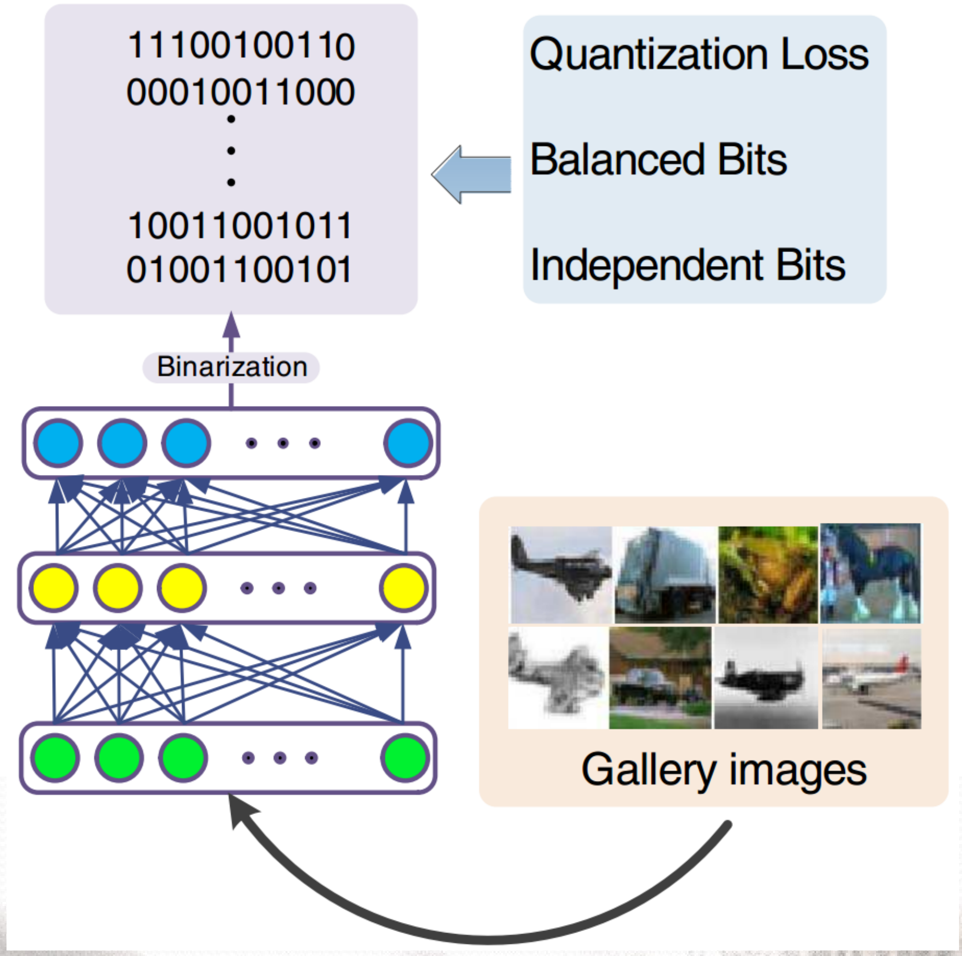
- Encoding strategy
  - Hashing functions $H=[h_1, h_2, \ldots, h_n]$
  - Binary code for sample x_1, $B_1=[h_1(x_1), h_2(x_1), \ldots, h_n(x_1)]$

Learning to Hashing for Visual Search

- Design of hashing function is crucial for effective search.
- Goal: Compact yet discriminative binary codes.



Learning to Hashing for Visual Search



[8] Venice Erin Liong, **Jiwen Lu,** Gang Wang, Pierre Moulin, and Jie Zhou, Deep hashing for compact binary codes learning, *CVPR*, pp. 2475-2483, 2015.

Learning to Hashing for Visual Search

Method	Hamming ranking (mAP, %)			precision (%) @ sample = 500			precision (%) @ r=2	
	16	32	64	16	32	64	16	32
PCA-ITQ [6]	15.67	16.20	16.64	22.46	25.30	27.09	22.60	14.99
KMH [8]	13.59	13.93	14.46	20.28	21.97	22.80	22.08	5.72
Spherical [9]	13.98	14.58	15.38	20.13	22.33	25.19	20.96	12.50
SH [36]	12.55	12.42	12.56	18.83	19.72	20.16	18.52	**20.60**
Semantic [26]	12.95	14.09	13.89	14.79	17.87	18.27	11.49	13.78
PCAH [34]	12.91	12.60	12.10	18.89	19.35	18.73	21.29	2.68
LSH [1]	12.55	13.76	15.07	16.21	19.10	22.25	16.73	7.07
DH	**16.17**	**16.62**	**16.96**	**23.79**	**26.00**	**27.70**	**23.33**	15.77
SPLH [34]	17.61	20.20	20.98	25.32	29.43	32.22	23.05	30.47
MLH [21]	18.37	20.49	21.89	24.43	29.60	33.01	23.52	28.72
BRE [15]	14.42	15.14	15.88	20.68	22.86	25.14	20.89	20.29
SDH	**18.80**	**20.83**	**22.51**	**26.32**	**30.42**	**33.60**	**23.26**	**31.48**

Learning to Hashing for Visual Search

- Multi-label extension
  - Re-formulate the between-class and within-class scatter matrix of SDH for multi-label samples

$$\Sigma_w^{(l)} = \sum_{i=1}^{N} \delta_{il} (\mathbf{h}_i^M - \mu_l)(\mathbf{h}_i^M - \mu_l)^\top$$

$$\Sigma_b^{(l)} = \sum_{i=1}^{N} \delta_{il} (\mu_l - \mu)(\mu_l - \mu)^\top$$

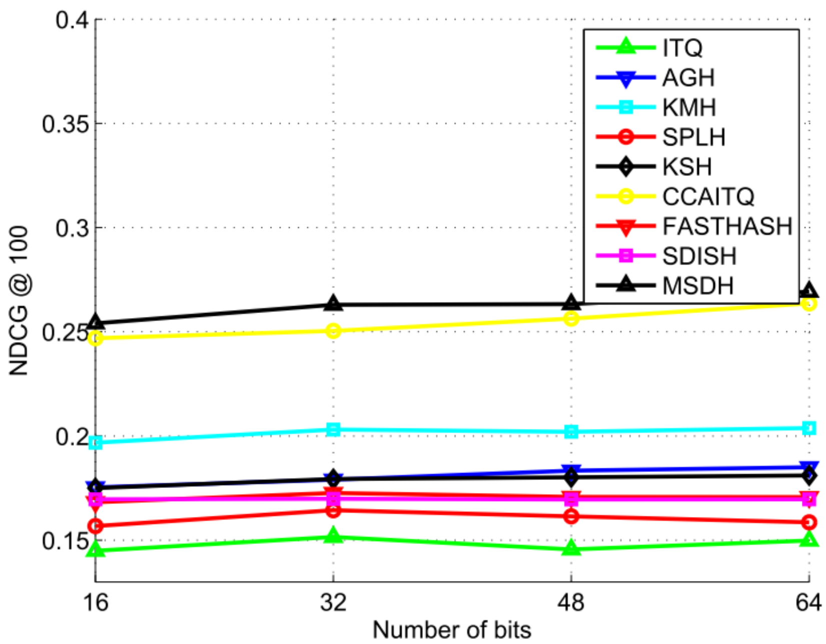
$$\Sigma_w = \sum_{l=1}^{L} \Sigma_w^{(l)}$$

$$\Sigma_b = \sum_{l=1}^{L} \Sigma_b^{(l)}$$

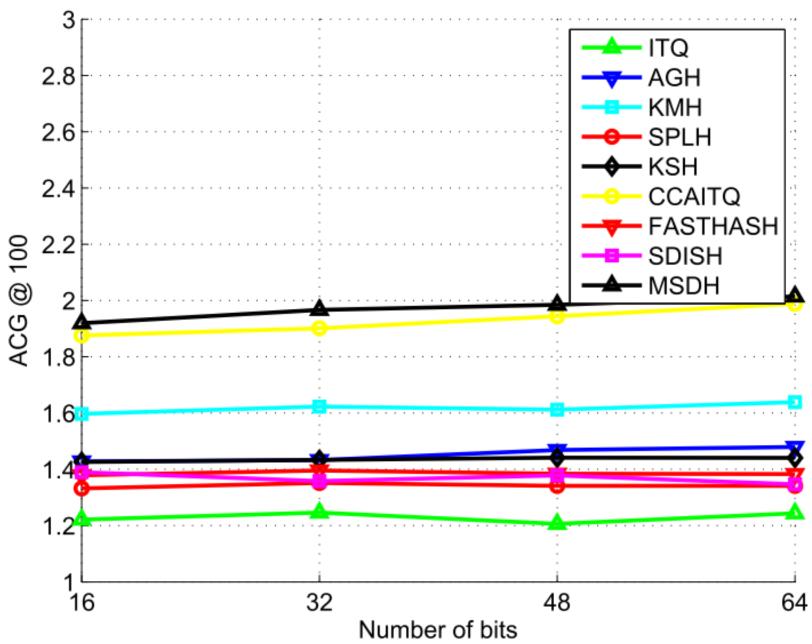
[9] **Jiwen Lu**, Venice Erin Liong, and Jie Zhou. Deep hashing for scalable image search, *IEEE Trans. on Image Processing*, vol. 26, no. 5, pp. 2352-2367, 2017.

Learning to Hashing for Visual Search

NDCG

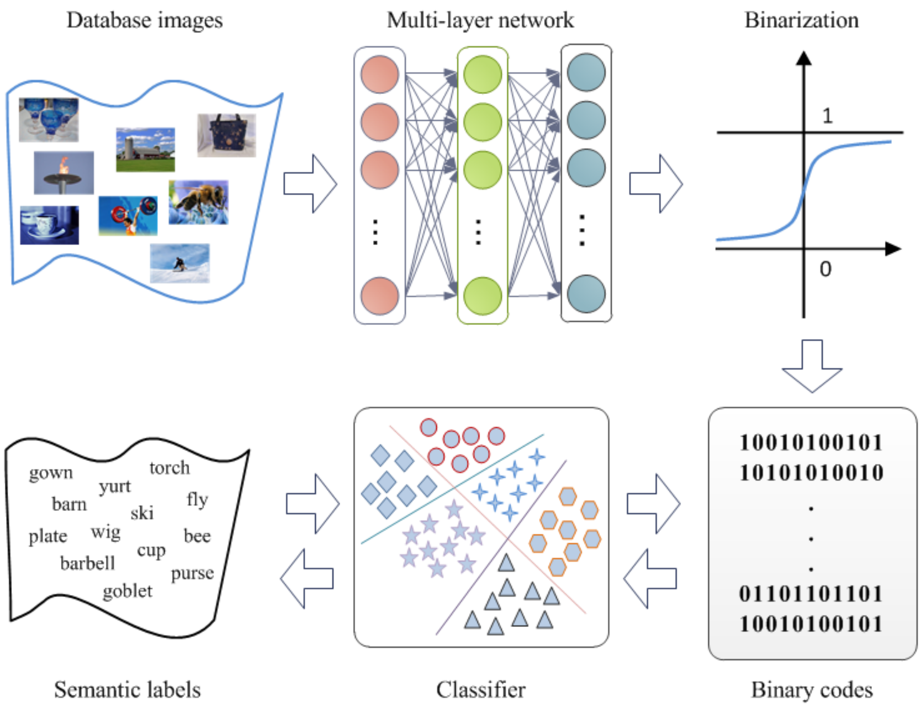


ACG



Learning to Hashing for Visual Search

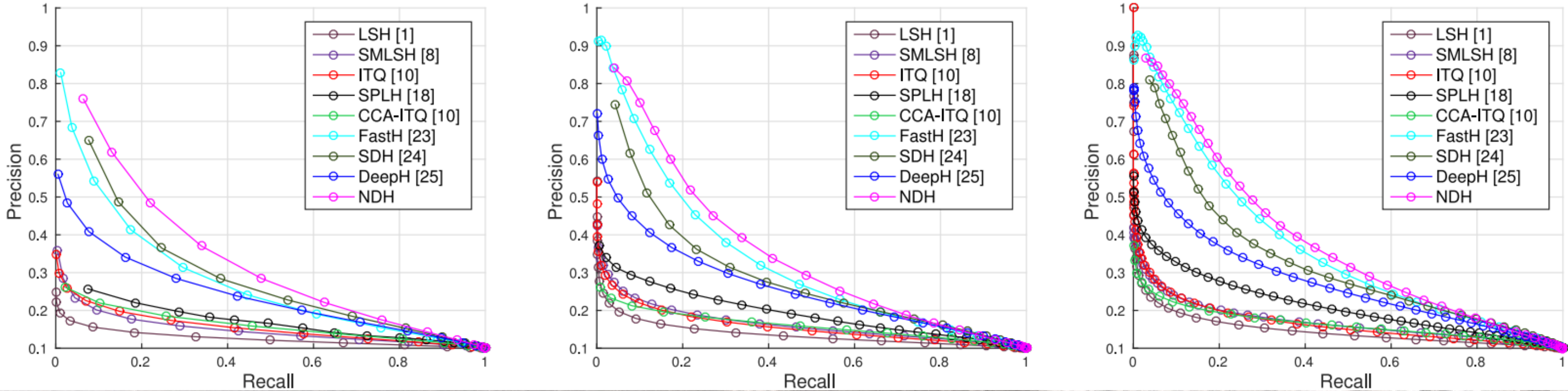
- Motivation
  - Exploit the nonlinear relationship of samples with nonlinear hashing functions
  - Solving the discrete optimization problem to eliminate the quantization error accumulation



[10] Zhixiang Chen, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou. Nonlinear discrete hashing, *IEEE Trans. on Multimedia,* vol. 19, no. 1, pp. 123-135, 2017.

Learning to Hashing for Visual Search

Methods	Mean average precision(%)			Precision@500(%)			Precision@(radius==2)(%)		
	16	32	64	16	32	64	16	32	64
LSH [1]	12.63	13.70	14.62	15.32	17.23	19.36	16.67	6.35	0.1
SMLSH [8]	14.96	16.41	16.98	17.82	19.75	20.36	18.28	14.65	4.03
ITQ [10]	15.57	15.80	16.57	19.91	21.04	22.53	22.89	15.66	1.44
SPLH [18]	17.08	19.38	21.21	21.22	26.39	29.34	16.70	27.17	30.02
CCA-ITQ [10]	16.21	16.02	16.49	24.63	24.44	26.77	21.45	28.22	26.47
FastH [23]	27.94	33.09	36.55	37.74	43.13	46.84	37.76	34.42	11.64
SDH [24]	29.21	29.22	32.67	39.08	39.62	42.15	30.19	36.90	38.98
DeepH [25]	24.04	25.96	27.53	32.45	34.99	36.85	33.25	37.42	25.43
NDH	33.75	35.93	37.90	43.58	46.67	48.24	36.10	43.62	32.32



Learning to Hashing for Visual Search

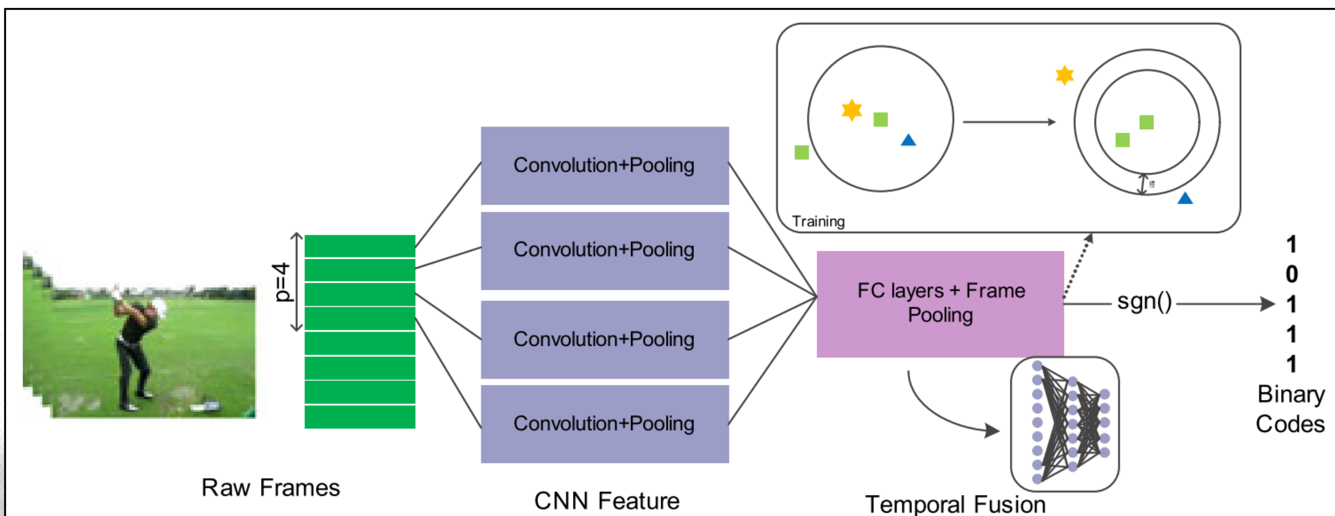
- Deep Video Hashing

Extract features for each frame

Image hashing techniques

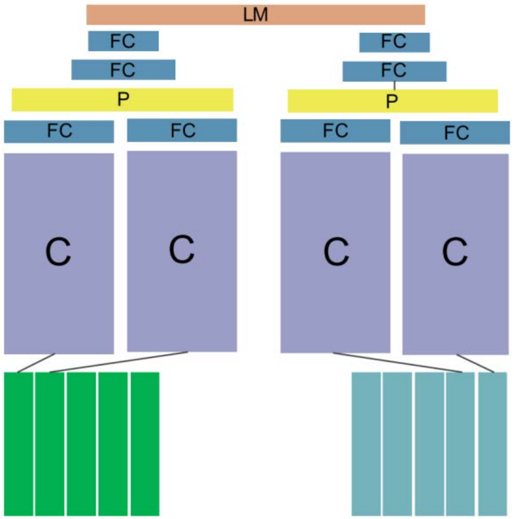
Handle entire video with a deep learning framework

Exploit both the temporal and discriminative information

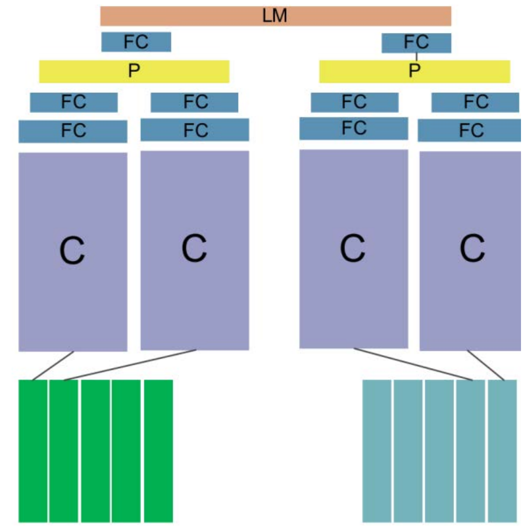


[11] Venice Erin Liong, **Jiwen Lu**, Yap-Peng Tan, and Jie Zhou. Deep video hashing, *IEEE Trans. on Multimedia*, vol. 19, no. 6, pp. 1234-1244, 2017.

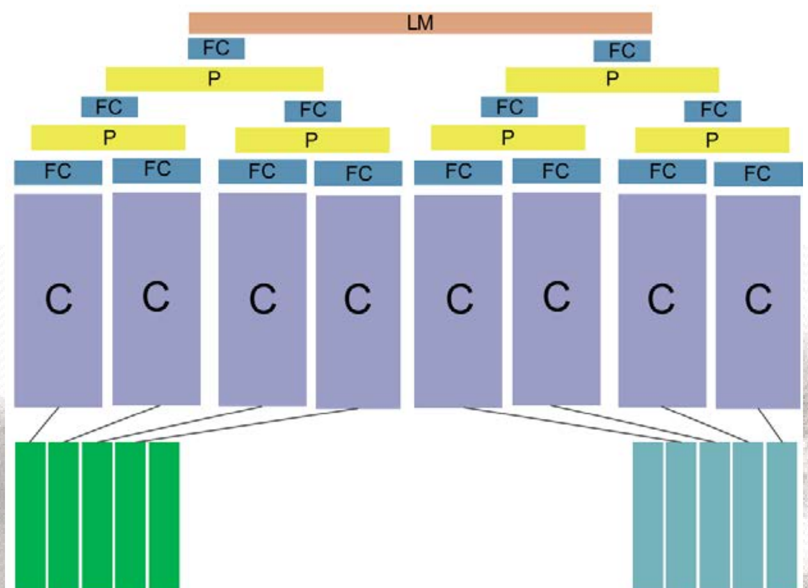
Learning to Hashing for Visual Search



Early fusion



Late fusion



Slow fusion

Learning to Hashing for Visual Search

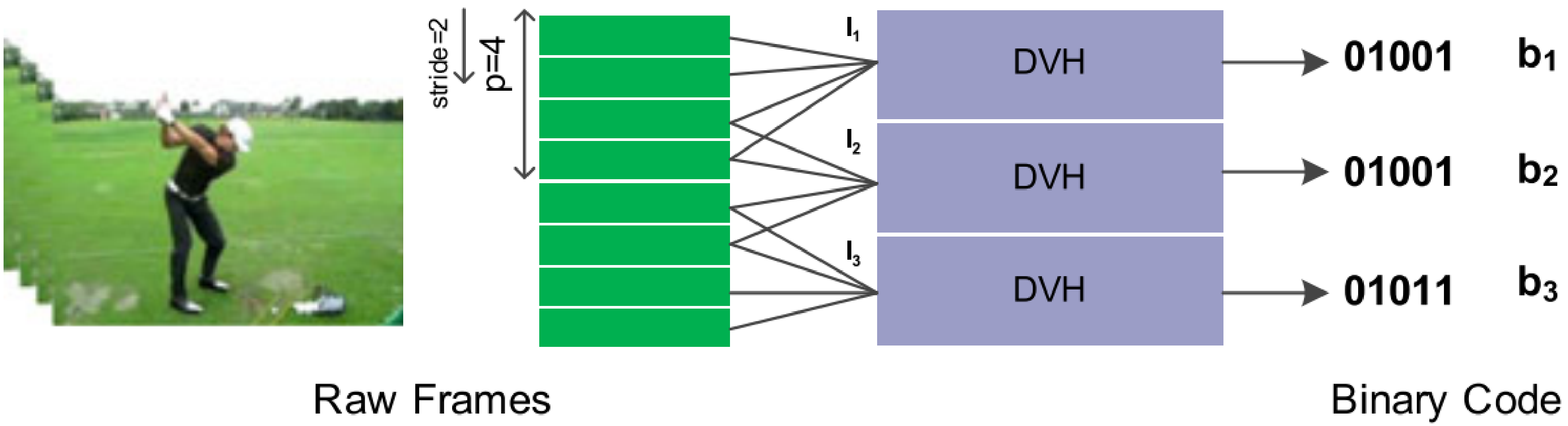
Tsinghua University

- Formulation
  - J1: discriminative learning. Minimize the intra-class variation and maximize the inter-class variation of the binary feature representation.
  - J2: efficient binary coding with minimizing the quantization loss.

$$
\begin{aligned}
\min_{\mathbf{b}_v,\mathbf{b}_v} J = \; & J_1 + \lambda J_2 \\
= \; & f(1 - \delta_{u,v}(\theta - d_{u,v}(\mathbf{b}_u, \mathbf{b}_v))) \\
& + \lambda(\|s(\mathbf{I}_u) - \mathbf{b}_u\|_F^2 + \|s(\mathbf{I}_v) - \mathbf{b}_v\|_F^2)
\end{aligned}
$$

Learning to Hashing for Visual Search

- Extracting binary codes from one video



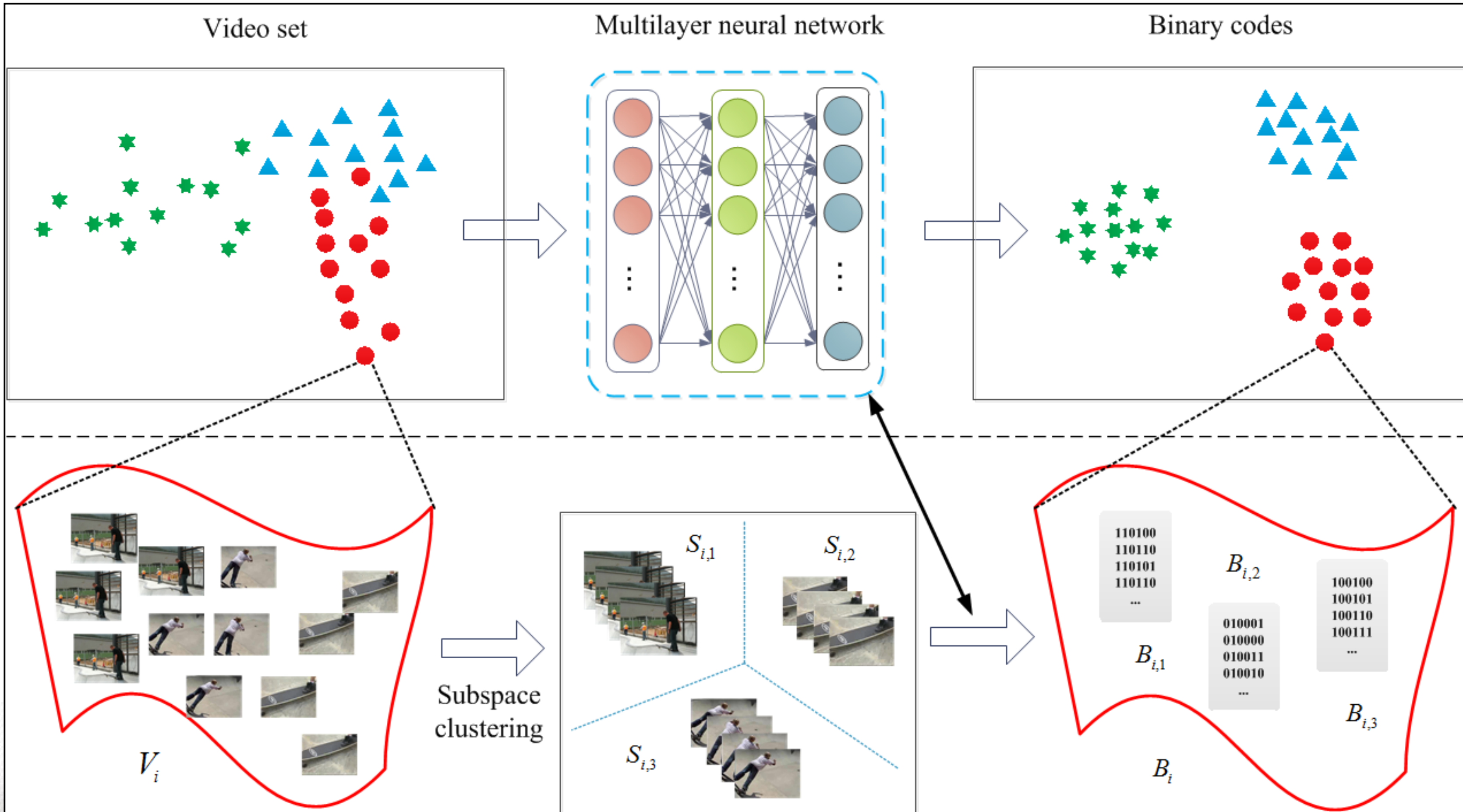
Learning to Hashing for Visual Search

清華大學 Tsinghua University

Method	Hamming ranking (mAP, %)			precision (%) @ N = 100			precision (%) @ r = 2	
	16	32	64	16	32	64	16	32
Single	32.62	34.23	35.02	37.37	38.86	38.50	23.38	5.63
Single+Temporal	33.60	35.60	37.27	38.05	39.74	40.93	30.76	16.44
Video-Level	29.45	30.79	29.19	34.44	35.98	34.05	22.48	11.11
Early Fusion	37.18	40.86	**41.54**	40.11	41.89	**42.41**	36.61	22.80
Late Fusion	**38.54**	**41.08**	41.51	**40.29**	**42.08**	42.23	**37.32**	**23.10**
Slow Fusion	38.27	40.80	41.41	39.95	41.88	42.34	36.55	23.06

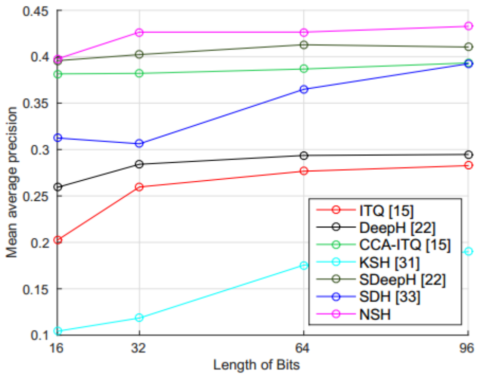
Method	Hamming ranking (mAP, %)			precision (%) @ N = 100			precision (%) @ r = 2	
	16	32	64	16	32	64	16	32
PCAH [28]	20.83	21.45	19.37	25.80	26.50	25.51	3.03	0
PCA-ITQ [6]	22.49	24.13	24.42	27.71	28.99	29.61	13.43	0
AGH [38]	14.91	15.22	11.24	20.52	23.37	20.16	13.43	1.58
KSH [31]	32.43	34.34	35.40	36.27	38.33	38.75	18.27	7.64
CCA-ITQ [6]	36.58	38.18	38.32	39.13	40.41	40.51	16.15	7.17
FastHash [37]	34.72	38.37	38.47	38.83	40.85	41.37	12.73	5.36
DVH	**38.54**	**41.08**	**41.51**	**40.29**	**42.08**	**42.23**	**37.32**	**23.10**

Learning to Hashing for Visual Search

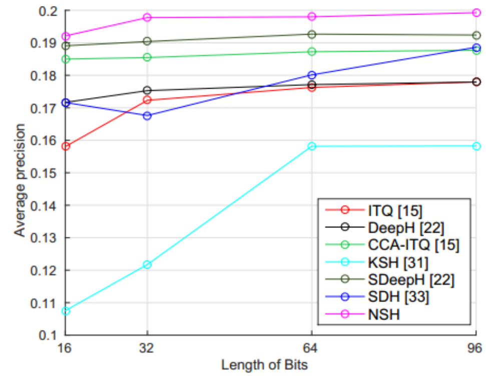


[12] Zhixiang Chen, **Jiwen Lu**, Jianjiang Feng, and Jie Zhou. Nonlinear structural hashing for scalable video search, *IEEE Transactions on Circuits and Systems for Video Technology*, 2017, accepted.

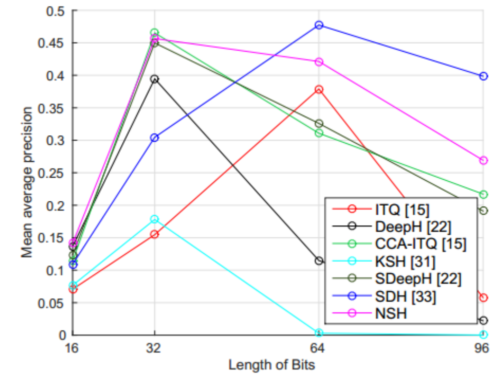
Learning to Hashing for Visual Search



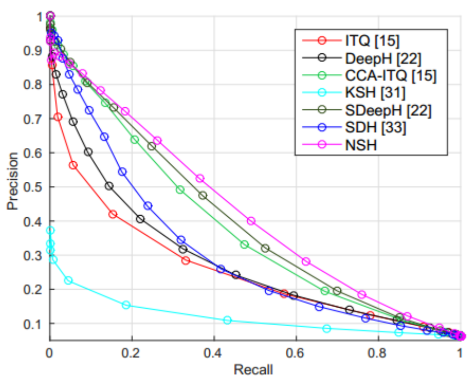
(a) mean Average Precision



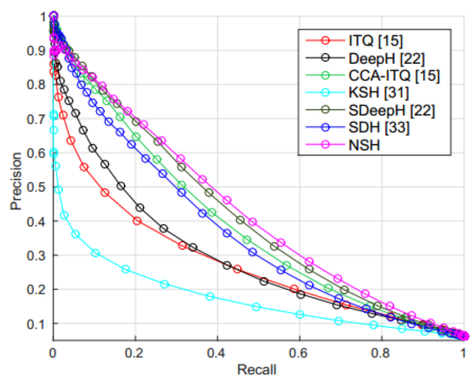
(b) Precision @500



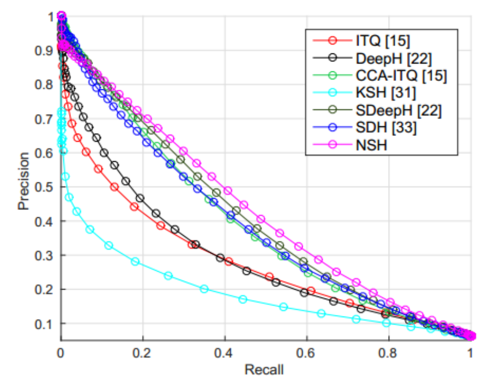
(c) Precision within Hamming radius 2



(a) 32 bits



(b) 64 bits



(c) 96 bits

Summary and Future Work

Tsinghua University

- Learning to hashing is very effective for many visual analysis tasks including visual recognition and search tasks.
- More efforts are desirable to further improve its real applications, especially on unsupervised hashing and structural hashing.
- New criterions are also required to better evaluate the performance of different hashing methods.