COLUMBIA UNIVERSITY IN THE CITY OF NEW YORK

Google™

CVPR June 13-18, 2010 San Francisco, CA

Semi-Supervised Hashing for Scalable Image Retrieval

[1]Jun Wang, [2]Sanjiv Kumar and [1]Shih-Fu Chang

[1]*Department of Electrical Engineering, Columbia University, New York, USA*

[2]*Google Research, New York, USA*

Introduction and Motivation

- The emerging need of fast and accurate indexing for large scale content based image retrieval (CBIR)
  - ➢Exhaustive search is infeasible due to computational and storage cost
  - ➢Instead of exact nearest neighbor, approximate nearest neighbor (ANN)search is more practical for large databases
  - ➢Compared with tree based ANN methods, hashing approach is more scalable and efficient
- Main issues:
  - ➢ Existing hashing methods mostly rely on random or principal projections, which are not very compact or accurate
  - ➢ Simple metrics are usually not enough to express semantic similarity
- **Goal**: Given a large unlabeled set with a few pair-wise labeled points, learn data-dependent compact hash codes

Related Work

- Locality Sensitive Hashing (LSH) -- *Indyk et al., 98*

  $h(x) = sign(\mathbf{w}^\top \mathbf{x} + b)$

  $\mathbf{w}$ - - sampled from p distribution b - - random shift
- Spectral Hashing (SH) -- *Weiss et al. 08*

  $h(\mathbf{x}) = \text{sgn}\left[\sin(\frac{\pi}{2} + \frac{k\pi}{b-a}\mathbf{w}^\top \mathbf{x})\right]$

  k - - spatial frequency $b-a$ - - data range

  $\mathbf{w}$ - - principal projections of data

  Sinusoidal binarization over PCA projections
- Restricted Boltzmann Machines (RBMs) – *Hinton et al. 06*
  - ➢ Learns deep belief networks to obtain compact binary codes
  - ➢ Two training steps, unsupervised pre-training and supervised fine tuning
  - ➢ Estimating a large number of weights requires lots of labeled data and long training time

Method	Projection Dependency	Learning Paradigm
LSH	data-independent	unsupervised
SH	data-dependent	unsupervised
RBMs	–	unsupervised/supervised
SSH	data-dependent	semi-supervised

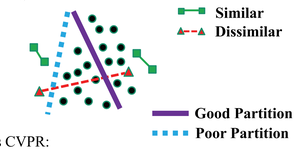
The conceptual comparison of the proposed SSH method with LSH, SH and RBMs.

Semi-Supervised Formulation

- Incorporate both metric similarity and semantic similarity for fast and accurate indexing
- Given pair wise labeled data $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M}$: neighbor pair; $(\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C}$ nonneighbor-pair
- Assume data zero centered and use mean threshold: $h_k(\mathbf{x}) = \text{sgn}(\mathbf{w}_k^\top \mathbf{x})$
- Empirical fitness: $J(\mathbf{H}) = \sum_k \left\{ \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{M}} h_k(\mathbf{x}_i)h_k(\mathbf{x}_j) - \sum_{(\mathbf{x}_i,\mathbf{x}_j)\in\mathcal{C}} h_k(\mathbf{x}_i)h_k(\mathbf{x}_j) \right\}$
- Regularization of maximum variance: $\sum_k E[(h_k(\mathbf{x}) - \mu_k)^2]$
- Final objective function

$$J(\mathbf{W}) = \frac{1}{2}\text{tr}\left[\mathbf{W}^\top \mathbf{X}_l \mathbf{S} \mathbf{X}_l^\top \mathbf{W}\right] + \frac{\eta}{2}\sum_k E[(\mathbf{w}_k^\top \mathbf{x})^2]$$

$$S_{ij} = \begin{cases} 1 & : (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{M} \\ -1 & : (\mathbf{x}_i, \mathbf{x}_j) \in \mathcal{C} \\ 0 & : otherwise. \end{cases}$$



AD The other semi-supervised hashing work in this year's CVPR: Y. Mu, et al., *Weakly-Supervised Hashing in Kernel Space*

Orthogonal Hash Codes

- Simplified objective function in matrix form

$$J(\mathbf{W}) = \frac{1}{2}\text{tr}\left[\mathbf{W}^\top \mathbf{X}_l \mathbf{S} \mathbf{X}_l^\top \mathbf{W}\right] + \frac{\eta}{2}\text{tr}\left[\mathbf{W}^\top \mathbf{X}\mathbf{X}^\top \mathbf{W}\right]$$

- By imposing orthogonality constraints, the objective function can be converted to a standard eigen problem

$$\arg\max_{\mathbf{W}} \frac{1}{2}\text{tr}\{\mathbf{W}^\top \mathbf{M} \mathbf{W}\} \quad \text{subject to:} \quad \mathbf{W}^\top \mathbf{W} = \mathbf{I}$$

"adjusted" covariance matrix

$$\mathbf{M} = \underbrace{\mathbf{X}_l \mathbf{S} \mathbf{X}_l^\top}_{\text{supervised}} + \underbrace{\eta \mathbf{X}\mathbf{X}^\top}_{\text{unsupervised}}$$

- Finding maximum variance projections over the data covariance matrix "adjusted" by incorporating the pair-wise labeled data
- **Issues**: Most real datasets have the variance concentrated on the top several few projections. orthogonality constrains force to progressively pick those low-variance directions, substantially reducing the quality of hamming embedding

Non-Orthogonal Hash Codes

- Relax orthogonality constraints and combine them to the objective function as penalty term

$$J(\mathbf{W}) = \frac{1}{2}\text{tr}\{\mathbf{W}^\top \mathbf{M} \mathbf{W}\} - \frac{\rho}{2}\|\mathbf{W}^\top \mathbf{W} - \mathbf{I}\|_{\mathcal{F}}^2$$

- Can be solved by a simple modification of the previous orthogonal solution

$$\mathbf{W} = [\mathbf{e}_1 \cdots \mathbf{e}_K] \quad \text{--- orthogonal solution}$$

principal projections of "adjusted" covariance matrix $\mathbf{M}$

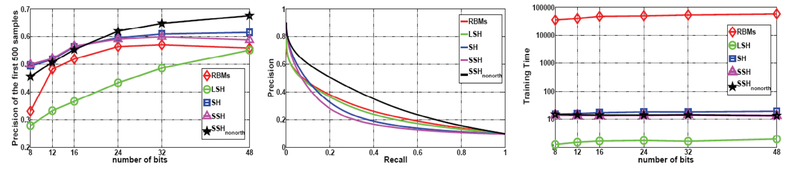
- Choose proper coefficient ρ to guarantee $\mathbf{Q}$ positive-definite

$$\mathbf{L}\mathbf{L}^\top = \mathbf{Q} = (\mathbf{I} + \frac{1}{\rho}\mathbf{M}) \quad \text{-- - Cholesky factorization}$$

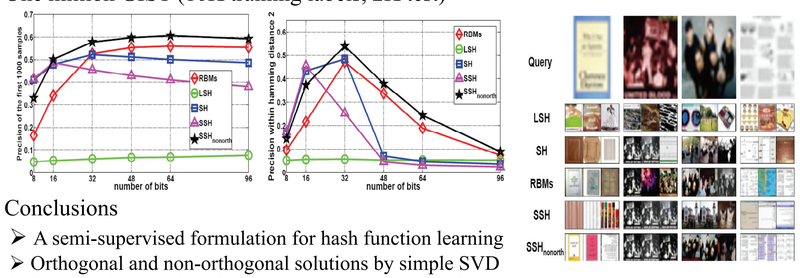
$$\mathbf{W}_{nonorth} = \mathbf{L} \cdot \mathbf{W} \quad \text{--- non-orthogonal solution}$$

Experiments and Conclusions

- MNIST (70K, 2K training labels, 1K test)



- One million GIST (10K training labels, 2K test)



- Conclusions
  - ➢ A semi-supervised formulation for hash function learning
  - ➢ Orthogonal and non-orthogonal solutions by simple SVD
  - ➢ Easy implementation, fast training, and superior performance