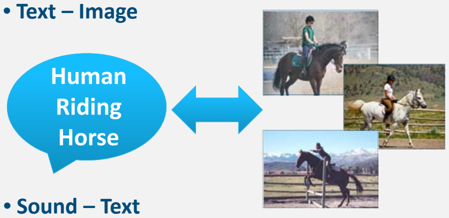
Predictable Dual-View Hashing

Mohammad Rastegari, Jonghyun Choi, Shobeir Fakhraei,
Hal Daumé III and Larry S. Davis
University of Maryland, College Park, MD, USA

Integrate different modalities

- Text – Image



- Sound – Text

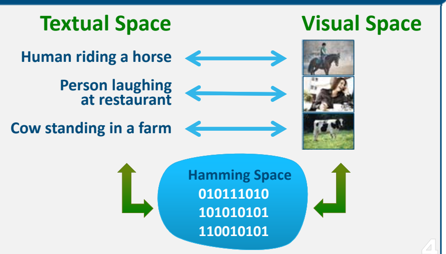
Challenges

- Modalities are not directly comparable
- High dimensional modalities
- Efficient data structure for search

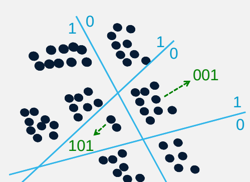
Previous approaches

- Domain specific [Farhadi et al. 2010]
- CCA-based [Gong et al. 2011, Sharma et al. 2012]

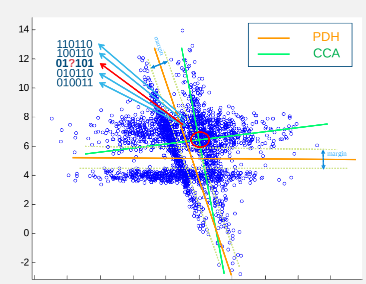
Dual-view hashing



Binary code assignment



Predictability



Each bit should be predictable based on the neighbors

Optimization

$$\min_{W_{\mathcal{V}},W_{\mathcal{T}}} \|\mathrm{sgn}(W_{\mathcal{V}}^T X_{\mathcal{V}}) - \mathrm{sgn}(W_{\mathcal{T}}^T X_{\mathcal{T}})\|_2^2$$

Trivial solution: both $W_{\mathcal{V}}, W_{\mathcal{T}}$ are zero

$$\min_{W_{\mathcal{V}},W_{\mathcal{T}}} \|W_{\mathcal{V}}^T X_{\mathcal{V}} - B_{\mathcal{T}}\|_2^2 + \|B_{\mathcal{T}} B_{\mathcal{T}}^T - I\|_2^2 + \|W_{\mathcal{T}}^T X_{\mathcal{T}} - B_{\mathcal{V}}\|_2^2 + \|B_{\mathcal{V}} B_{\mathcal{V}}^T - I\|_2^2$$

s.t.

$$B_{\mathcal{T}} = \mathrm{sgn}(W_{\mathcal{T}}^T X_{\mathcal{T}}),$$
$$B_{\mathcal{V}} = \mathrm{sgn}(W_{\mathcal{V}}^T X_{\mathcal{V}}).$$

Optimization is non-convex and combinatorial

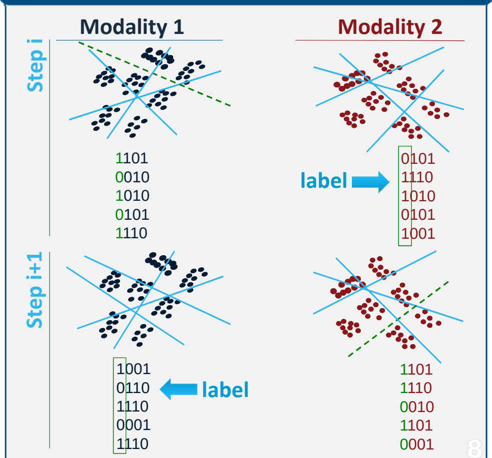
$$\min_{W_{\mathcal{V}},W_{\mathcal{T}},\xi_{\mathcal{V}},\xi_{\mathcal{T}}} \|B_{\mathcal{T}} B_{\mathcal{T}}^T - I\|_2^2 + \|B_{\mathcal{V}} B_{\mathcal{V}}^T - I\|_2^2 + \sum \|w_{\mathcal{V}i}\| + \sum \|w_{\mathcal{T}i}\| + C_1 \sum \xi_{\mathcal{V}} + C_2 \sum \xi_{\mathcal{T}}$$

s.t.

$$B_{\mathcal{T}} = \mathrm{sgn}(W_{\mathcal{T}}^T X_{\mathcal{T}}),$$
$$B_{\mathcal{V}} = \mathrm{sgn}(W_{\mathcal{V}}^T X_{\mathcal{V}}),$$
$$B_{\mathcal{T}}^{ij}({w_{\mathcal{V}i}}^T X_{\mathcal{V}}^j) \geq 1 - \xi_{\mathcal{V}}^{ij} \quad \forall i,j,$$
$$B_{\mathcal{V}}^{ij}({w_{\mathcal{T}i}}^T X_{\mathcal{T}}^j) \geq 1 - \xi_{\mathcal{T}}^{ij} \quad \forall i,j.$$

Using a block coordinate descent algorithm

How we do it



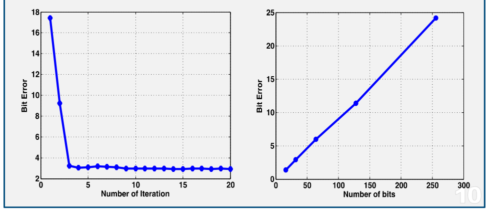
Algorithm

Algorithm 1 Predictable Dual-View Hashing

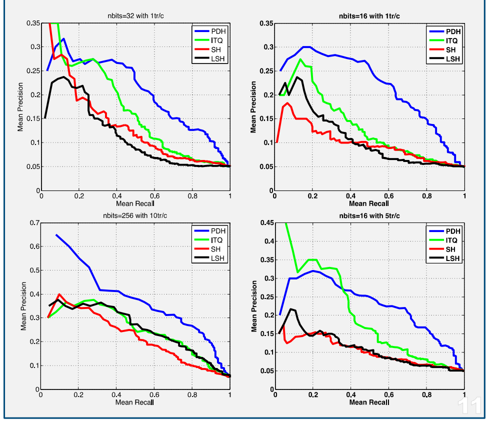
Input: $X_{\mathcal{V}},\ X_{\mathcal{T}} \in \mathbb{R}^{d_* \times n}$.
Output: $B_{\mathcal{V}},\ B_{\mathcal{T}} \in \mathbb{B}^{d_* \times k}$.

1: $W_{\mathcal{V}}, W_{\mathcal{T}} \in \mathbb{R}^{d_* \times k} \leftarrow CCA(X_{\mathcal{V}}, X_{\mathcal{T}}, k)$
2: $B_{\mathcal{V}} \leftarrow \mathrm{sgn}({W_{\mathcal{V}}}^T X_{\mathcal{V}})$
3: $B_{\mathcal{T}} \leftarrow \mathrm{sgn}({W_{\mathcal{T}}}^T X_{\mathcal{T}})$
4: **repeat**
5: $W_{\mathcal{V}} \leftarrow$ Weights of k linear SVMs (for i^{th} SVM: training features are columns of $X_{\mathcal{V}}$ and training labels are elements of i^{th} row of $B_{\mathcal{T}}$)
6: $B_{\mathcal{V}} \leftarrow \mathrm{sgn}({W_{\mathcal{V}}}^T X_{\mathcal{V}})$
7: Update $B_{\mathcal{V}}$ using Eq. (5)
8: $W_{\mathcal{T}} \leftarrow$ Weights of k linear SVMs (for i^{th} SVM: training features are columns of $X_{\mathcal{T}}$ and training labels are elements of i^{th} row of $B_{\mathcal{V}}$)
9: $B_{\mathcal{T}} \leftarrow \mathrm{sgn}({W_{\mathcal{T}}}^T X_{\mathcal{T}})$
10: Update $B_{\mathcal{T}}$ using Eq. (5)
11: **until** convergence
12: $B_{\mathcal{V}} \leftarrow \mathrm{sgn}({W_{\mathcal{V}}}^T X_{\mathcal{V}})$
13: $B_{\mathcal{T}} \leftarrow \mathrm{sgn}({W_{\mathcal{T}}}^T X_{\mathcal{T}})$

Optimization analysis

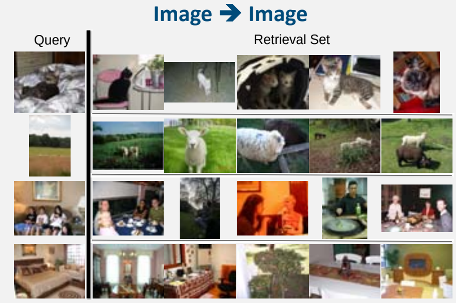


Experiments (category retrieval)



Qualitative results (w/ 32-bit codes)

Image ➔ Image



Text ➔ Image