Graph Regularised Hashing

Sean Moran and Victor Lavrenko

Institute of Language, Cognition and Computation
School of Informatics
University of Edinburgh

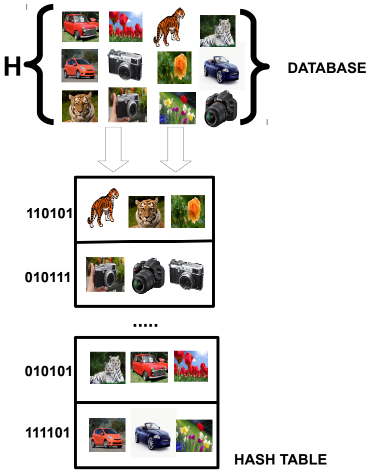
ECIR'15 Vienna, March 2015

Graph Regularised Hashing (GRH)

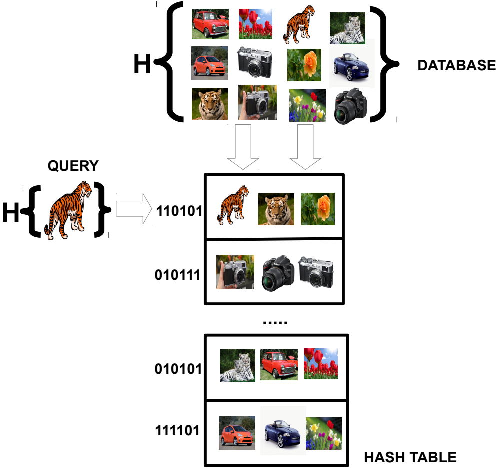
Graph Regularised Hashing (GRH)

Locality Sensitive Hashing

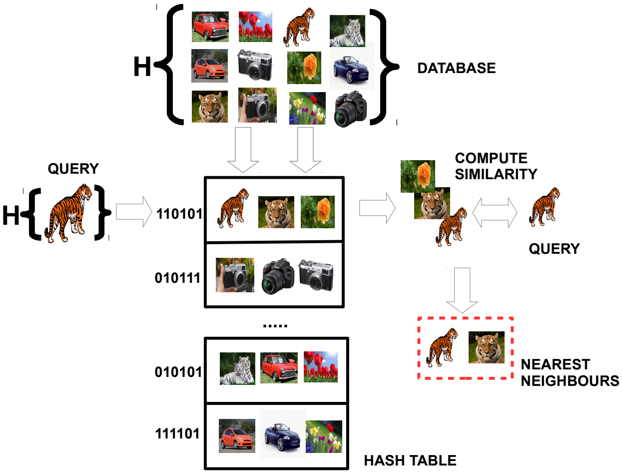
Locality Sensitive Hashing



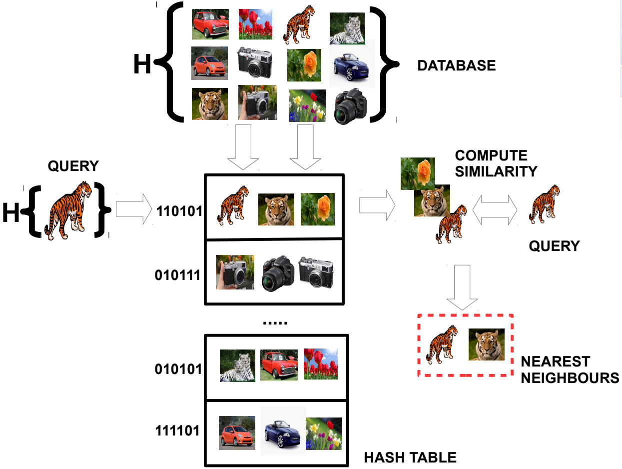
Locality Sensitive Hashing



Locality Sensitive Hashing



Locality Sensitive Hashing



Content Based IR

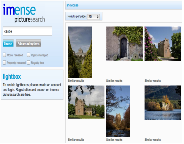


Image: Imense Ltd

Near duplicate detection

Image: Xu et al.

Location Recognition

Image: Doersch et al.

Previous work

- **Data-independent:** Locality Sensitive Hashing (LSH) [Indyk. 98]
- **Data-dependent (unsupervised):** Anchor Graph Hashing (AGH) [Liu et al. '11], Spectral Hashing (SH) [Weiss '08]
- **Data-dependent (supervised):** Self Taught Hashing (STH) [Zhang '10], Supervised Hashing with Kernels (KSH) [Liu et al. '12], ITQ + CCA [Gong and Lazebnik '11], Binary Reconstructive Embedding (BRE) [Kulis and Darrell. '09]

Previous work

Method	Data-Dependent	Supervised	Scalable	Effectiveness
LSH			✓	Low
SH	✓			Low
STH	✓	✓		Medium
BRE	✓	✓		Medium
ITQ+CCA	✓	✓		Medium
KSH	✓	✓		High
GRH	✓	✓	✓	**High**

Graph Regularised Hashing (GRH)

Graph Regularised Hashing (GRH)

- Two step *iterative* hashing model:
  - **Step A:** Graph Regularisation

$$\mathbf{L}_m \leftarrow \operatorname{sgn}\left(\alpha\ \mathbf{SD}^{-1}\mathbf{L}_{m-1} + (1-\alpha)\mathbf{L}_0\right)$$

  - **Step B:** Data-Space Partitioning

$$\begin{aligned} \text{for } k = 1 \ldots K: \quad & \min \ ||\mathbf{h}_k||^2 + C \textstyle\sum_{i=1}^{N} \xi_{ik} \\ \text{s.t.} \quad & L_{ik}(\mathbf{h}_k{}^{\top}\mathbf{x}_i + b_k) \geq 1 - \xi_{ik} \qquad \text{for } i = 1 \ldots N \end{aligned}$$

- Repeat for a set number of iterations (M)

Graph Regularised Hashing (GRH)

- **Step A:** Graph Regularisation [Diaz '07][1]

$$\mathbf{L}_m \leftarrow \operatorname{sgn}\left(\alpha\ \mathbf{S}\mathbf{D}^{-1}\mathbf{L}_{m-1} + (1-\alpha)\mathbf{L}_0\right)$$

- **S**: Affinity (adjacency) matrix
- **D**: Diagonal degree matrix
- **L**: Binary bits at specified iteration
- α: Interpolation parameter ($0 \leq \alpha \leq 1$)

[1] Diaz, F.: Regularizing query-based retrieval scores. In: IR (2007)

Graph Regularised Hashing (GRH)

- **Step A:** Graph Regularisation [Diaz '07]

$$\mathbf{L}_m \leftarrow \mathrm{sgn}\left(\alpha\, \mathbf{S}\mathbf{D}^{-1}\mathbf{L}_{m-1} + (1-\alpha)\mathbf{L}_0\right)$$

- **S**: Affinity (adjacency) matrix
- **D**: Diagonal degree matrix
- **L**: Binary bits at specified iteration
- α: Interpolation parameter $(0 \leq \alpha \leq 1)$

Graph Regularised Hashing (GRH)

- **Step A:** Graph Regularisation [Diaz '07]

$$\mathbf{L}_m \leftarrow \mathrm{sgn}\left(\alpha\ \mathbf{S}\mathbf{D}^{-1}\mathbf{L}_{m-1} + (1-\alpha)\mathbf{L}_0\right)$$

- **S**: Affinity (adjacency) matrix
- **D**: Diagonal degree matrix
- **L**: Binary bits at specified iteration
- α: Interpolation parameter ($0 \leq \alpha \leq 1$)

Graph Regularised Hashing (GRH)

- **Step A:** Graph Regularisation [Diaz '07]

$$\mathbf{L}_m \leftarrow \operatorname{sgn}\left(\alpha\ \mathbf{SD}^{-1}\mathbf{L}_{m-1} + (1-\alpha)\mathbf{L}_0\right)$$

- **S**: Affinity (adjacency) matrix
- **D**: Diagonal degree matrix
- **L**: Binary bits at specified iteration
- α: Interpolation parameter ($0 \leq \alpha \leq 1$)

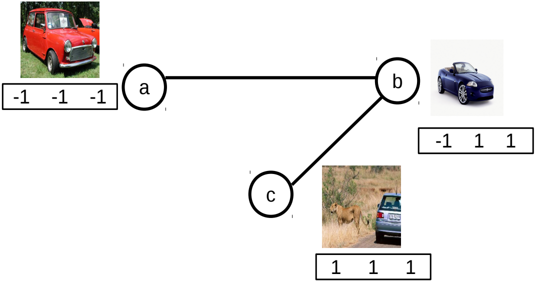
Graph Regularised Hashing (GRH)

- **Step A:** Graph Regularisation [Diaz '07]

$$\mathbf{L}_m \leftarrow \operatorname{sgn}\left(\alpha\ \mathbf{S}\mathbf{D}^{-1}\mathbf{L}_{m-1} + (1-\alpha)\mathbf{L}_0\right)$$

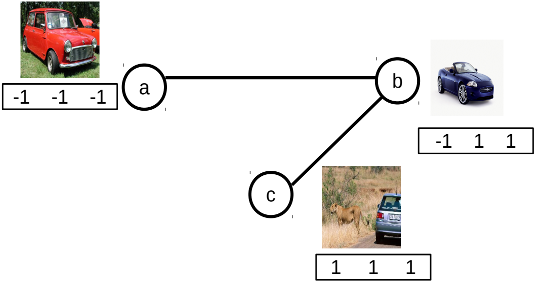
- **S**: Affinity (adjacency) matrix
- **D**: Diagonal degree matrix
- **L**: Binary bits at specified iteration
- α: Interpolation parameter ($0 \leq \alpha \leq 1$)

Graph Regularised Hashing (GRH)



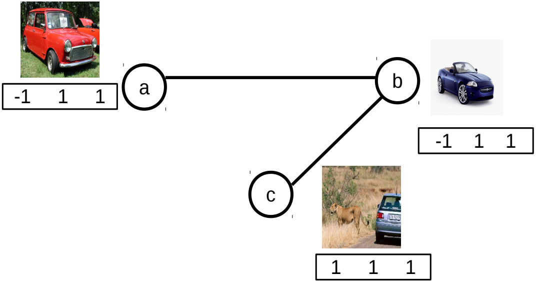
$$
\begin{array}{c c}
\mathbf{S} & \begin{array}{ccc} a & b & c \end{array} \\
\begin{array}{c} a \\ b \\ c \end{array} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}
\end{array}
\qquad
\begin{array}{c c}
\mathbf{D}^{-1} & \begin{array}{ccc} a & b & c \end{array} \\
\begin{array}{c} a \\ b \\ c \end{array} & \begin{pmatrix} 0.5 & 0 & 0 \\ 0 & 0.33 & 0 \\ 0 & 0 & 0.5 \end{pmatrix}
\end{array}
\qquad
\begin{array}{c c}
\mathbf{L}_0 & \begin{array}{ccc} b_1 & b_2 & b_3 \end{array} \\
\begin{array}{c} a \\ b \\ c \end{array} & \begin{pmatrix} -1 & -1 & -1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}
\end{array}
$$

Graph Regularised Hashing (GRH)



$$\mathbf{L}_1 = sgn\left\{ \begin{pmatrix} -1 & 0 & 0 \\ -0.33 & 0.33 & 0.33 \\ 0 & 1 & 1 \end{pmatrix} \right\}$$

Graph Regularised Hashing (GRH)



$$\mathbf{L}_1 = \begin{array}{c c} & \begin{array}{ccc} b_1 & b_2 & b_3 \end{array} \\ \begin{array}{c} a \\ b \\ c \end{array} & \begin{pmatrix} -1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

Graph Regularised Hashing (GRH)

- **Step B:** Data-Space Partitioning

$$\text{for } k = 1 \ldots K : \quad \min \; ||\mathbf{h}_k||^2 + C \textstyle\sum_{i=1}^{N} \xi_{ik}$$
$$\text{s.t.} \quad L_{ik}(\mathbf{h}_k{}^\intercal \mathbf{x}_i + b_k) \geq 1 - \xi_{ik} \qquad \text{for } i = 1 \ldots N$$

- $\mathbf{h}_k$: Hyperplane k
- b_k: bias of hyperplane k
- $\mathbf{x}_i$: data-point i
- L_{ik}: bit k of data-point i

ξ_{ik}: slack variable ij

K: # bits

N: # data-points

Graph Regularised Hashing (GRH)

- **Step B:** Data-Space Partitioning

$$\text{for } k = 1 \ldots K: \quad \min \; ||\mathbf{h}_k||^2 + C \sum_{i=1}^{N} \xi_{ik}$$
$$\text{s.t.} \quad L_{ik}(\mathbf{h}_k{}^{\intercal}\mathbf{x}_i + b_k) \geq 1 - \xi_{ik} \qquad \text{for } i = 1 \ldots N$$

- $\mathbf{h}_k$: Hyperplane k
- b_k: bias of hyperplane k
- $\mathbf{x}_i$: data-point i
- L_{ik}: bit k of data-point i

ξ_{ik}: slack variable ij

K: # bits

N: # data-points

Graph Regularised Hashing (GRH)

- **Step B:** Data-Space Partitioning

$$
\text{for } k = 1 \ldots K: \quad \min \; ||\mathbf{h}_k||^2 + C \textstyle\sum_{i=1}^{N} \xi_{ik}
$$
$$
\text{s.t.} \quad L_{ik}(\mathbf{h}_k{}^\intercal \mathbf{x}_i + b_k) \geq 1 - \xi_{ik} \qquad \text{for } i = 1 \ldots N
$$

- $\mathbf{h}_k$: Hyperplane k
- b_k: bias of hyperplane k
- $\mathbf{x}_i$: data-point i
- L_{ik}: bit k of data-point i

ξ_{ik}: slack variable ij

K: # bits

N: # data-points

Graph Regularised Hashing (GRH)

- **Step B:** Data-Space Partitioning

$$\text{for } k = 1 \ldots K: \quad \min \; ||\mathbf{h}_k||^2 + C\sum_{i=1}^{N} \xi_{ik}$$
$$\text{s.t.} \quad L_{ik}(\mathbf{h}_k{}^\intercal \mathbf{x}_i + b_k) \geq 1 - \xi_{ik} \qquad \text{for } i = 1 \ldots N$$

- $\mathbf{h}_k$: Hyperplane k
- b_k: bias of hyperplane k
- $\mathbf{x}_i$: data-point i
- L_{ik}: bit k of data-point i

ξ_{ik}: slack variable ij

K: # bits

N: # data-points

Graph Regularised Hashing (GRH)

- **Step B:** Data-Space Partitioning

$$
\begin{aligned}
\text{for } k = 1 \ldots K: \quad & \min \ ||\mathbf{h}_k||^2 + C \textstyle\sum_{i=1}^{N} \xi_{ik} \\
\text{s.t.} \quad & L_{ik}(\mathbf{h}_k{}^\intercal \mathbf{x}_i + b_k) \geq 1 - \xi_{ik} \qquad \text{for } i = 1 \ldots N
\end{aligned}
$$

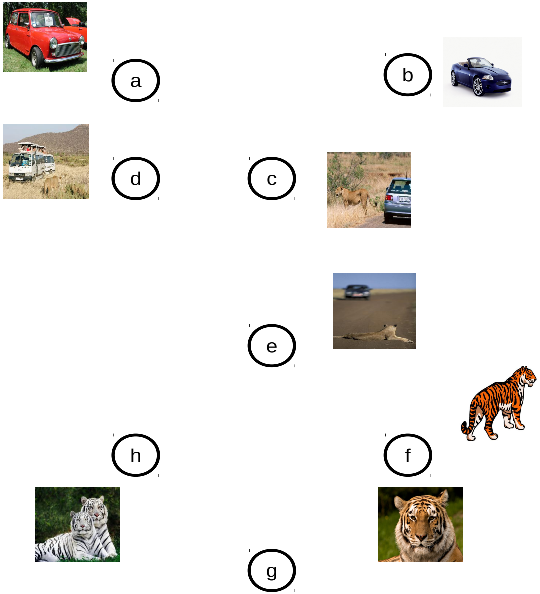
- $\mathbf{h}_k$: Hyperplane k
- b_k: bias of hyperplane k
- $\mathbf{x}_i$: data-point i
- L_{ik}: bit k of data-point i

ξ_{ik}: slack variable ij

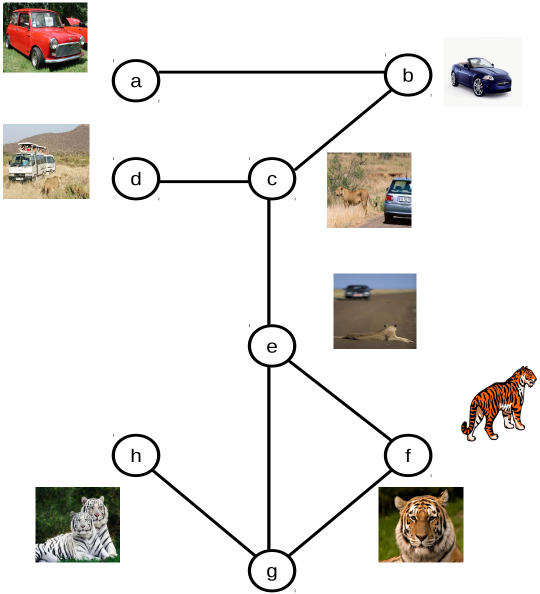
K: $\#$ bits

N: $\#$ data-points

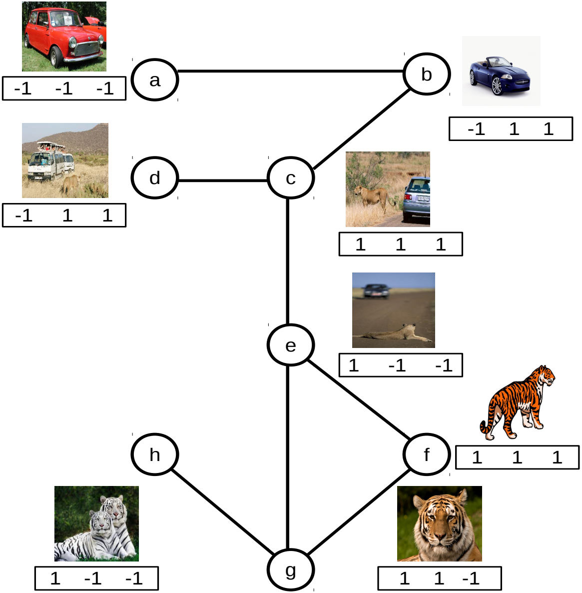
Graph Regularised Hashing (GRH)



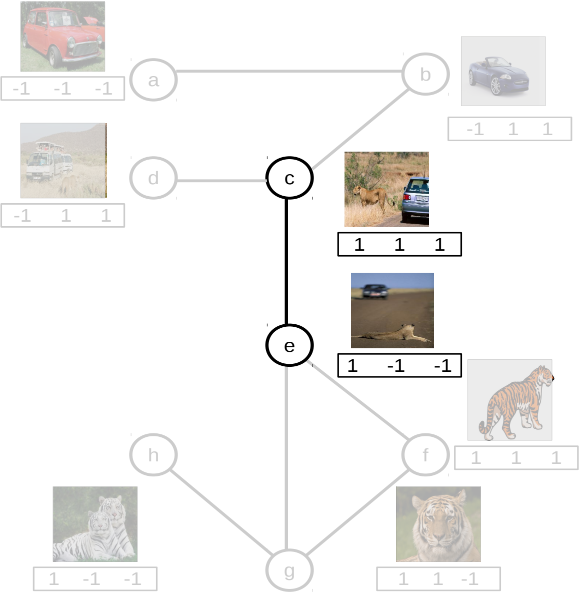
Graph Regularised Hashing (GRH)



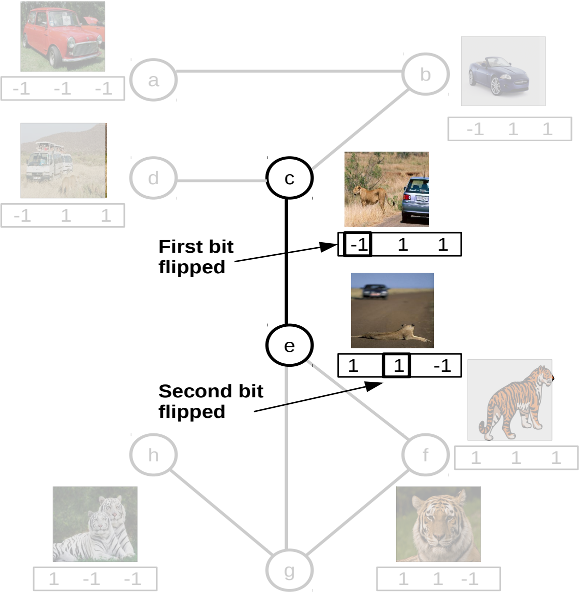
Graph Regularised Hashing (GRH)



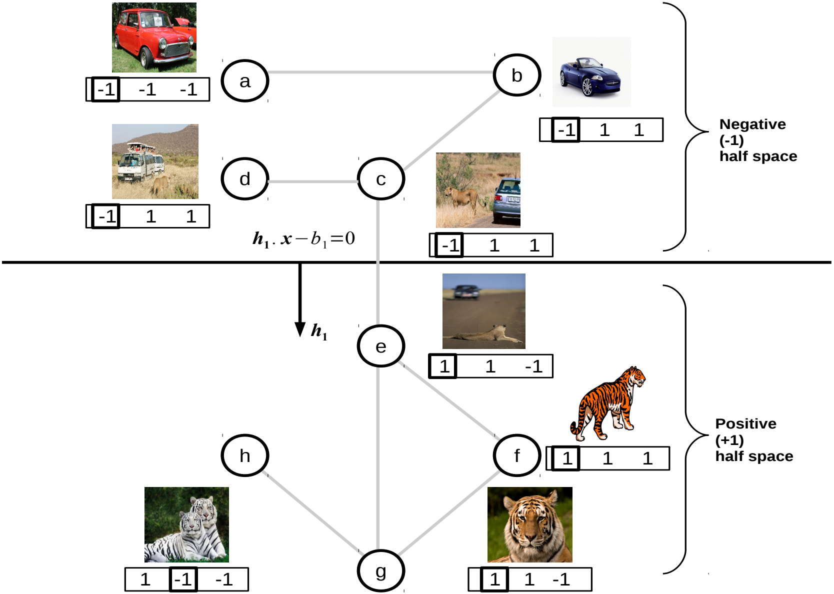
Graph Regularised Hashing (GRH)



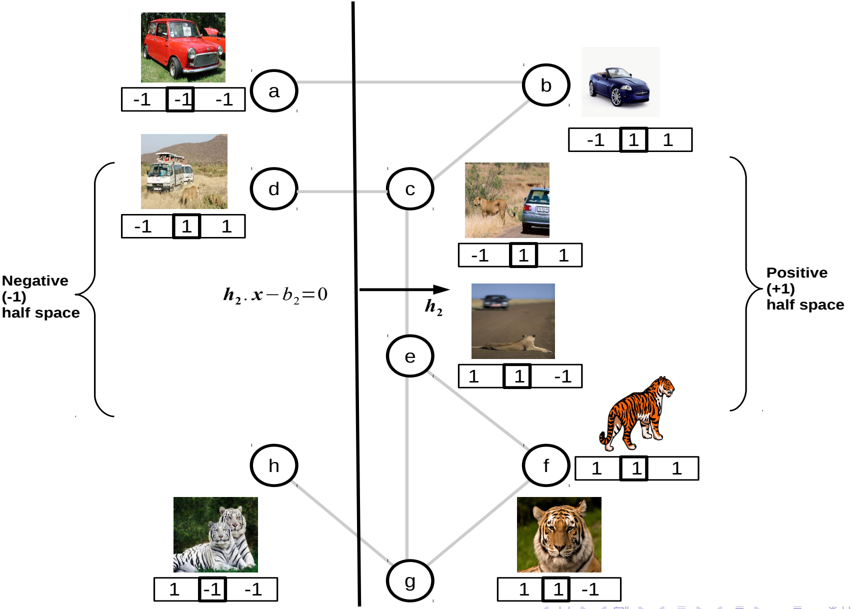
Graph Regularised Hashing (GRH)



Graph Regularised Hashing (GRH)



Graph Regularised Hashing (GRH)



Evaluation

Datasets/Features

- Standard evaluation datasets [Liu et al. '12], [Gong and Lazebnik '11]:
  - **CIFAR-10**: 60K images, GIST descriptors, 10 classes[1]
  - **MNIST**: 70K images, grayscale pixels, 10 classes[2]
  - **NUSWIDE**: 270K images, GIST descriptors, 21 classes[3]
- True NNs: images that share at least one class in common [Liu et al. '12]

[1] http://www.cs.toronto.edu/~kriz/cifar.html

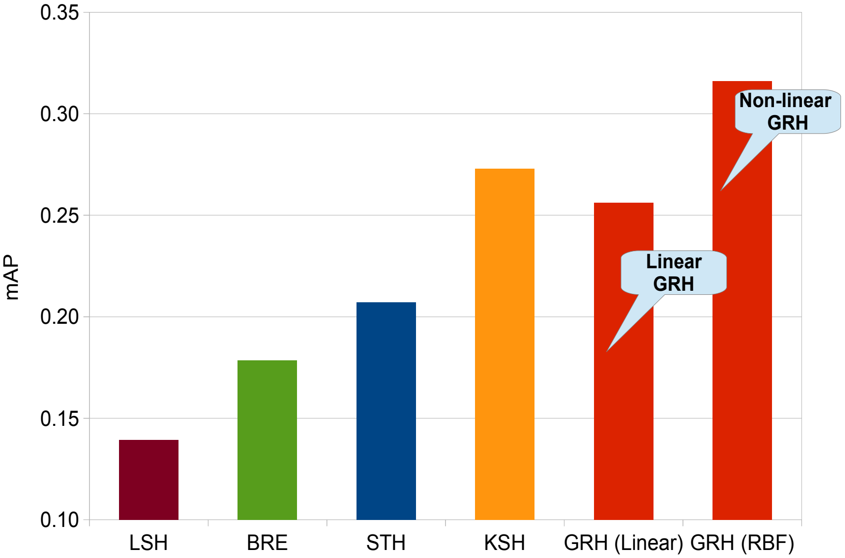
[2] http://yann.lecun.com/exdb/mnist/

[3] http://lms.comp.nus.edu.sg/research/NUS-WIDE.htm

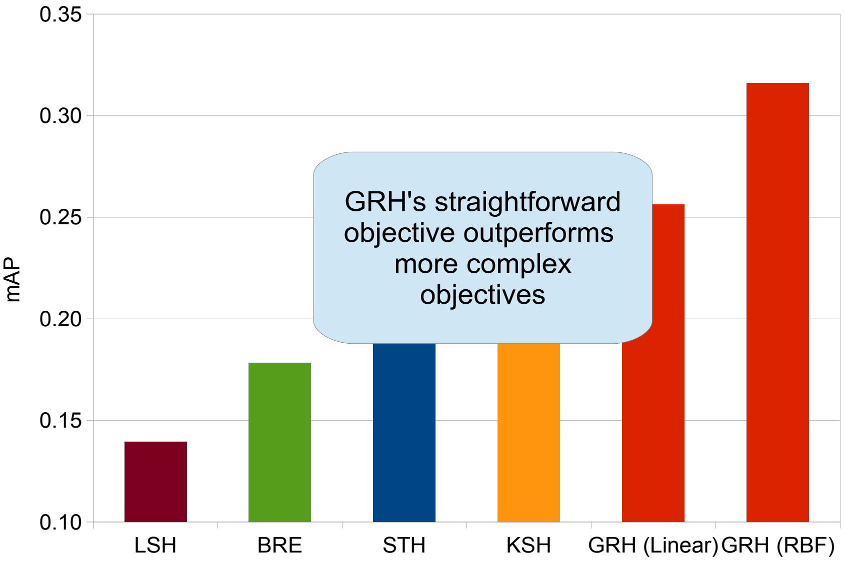
Evaluation Metrics

- Hamming ranking evaluation paradigm [Liu et al. '12], [Gong and Lazebnik '11]
- Standard evaluation metrics [Liu et al. '12], [Gong and Lazebnik '11]:
  - Mean average precison (mAP)
  - Precision at Hamming radius 2 (P@R2)

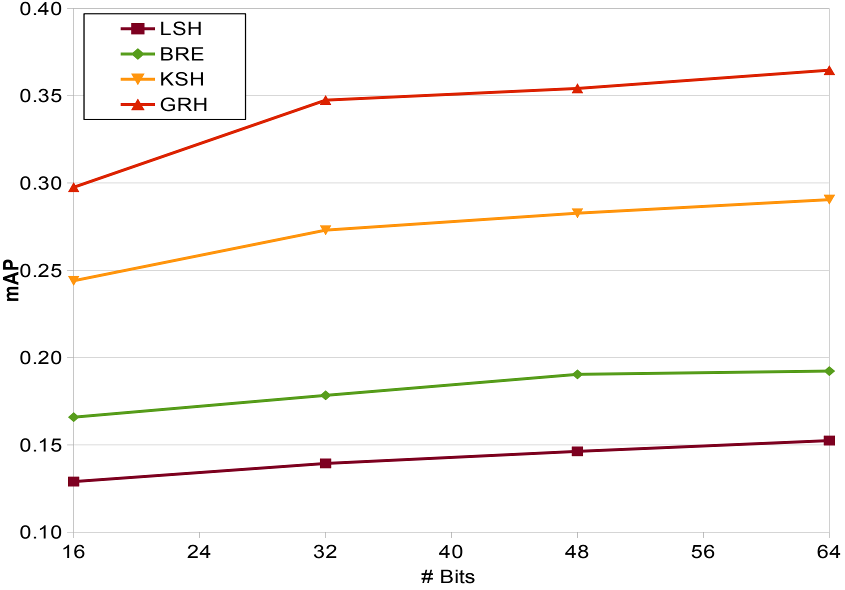
GRH vs Literature (CIFAR-10 @ 32 bits)



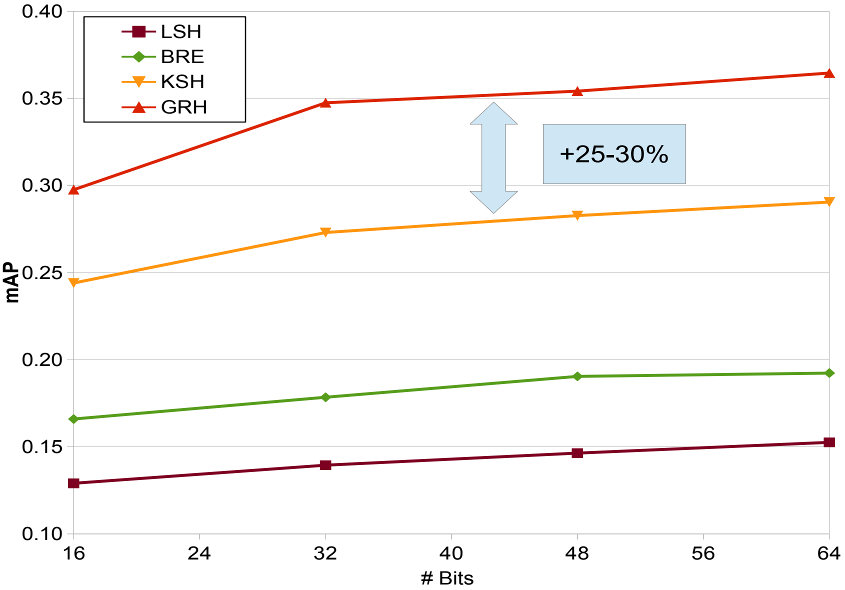
GRH vs Literature (CIFAR-10 @ 32 bits)



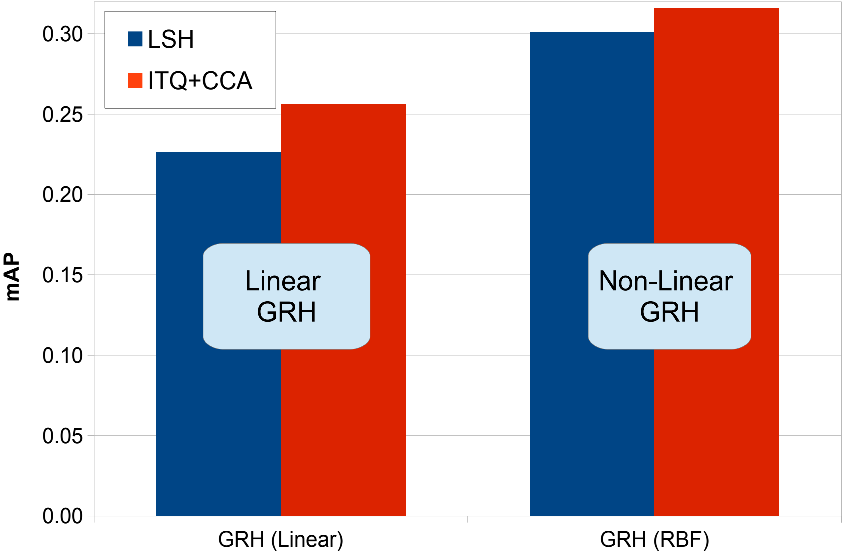
GRH vs Literature (CIFAR-10)



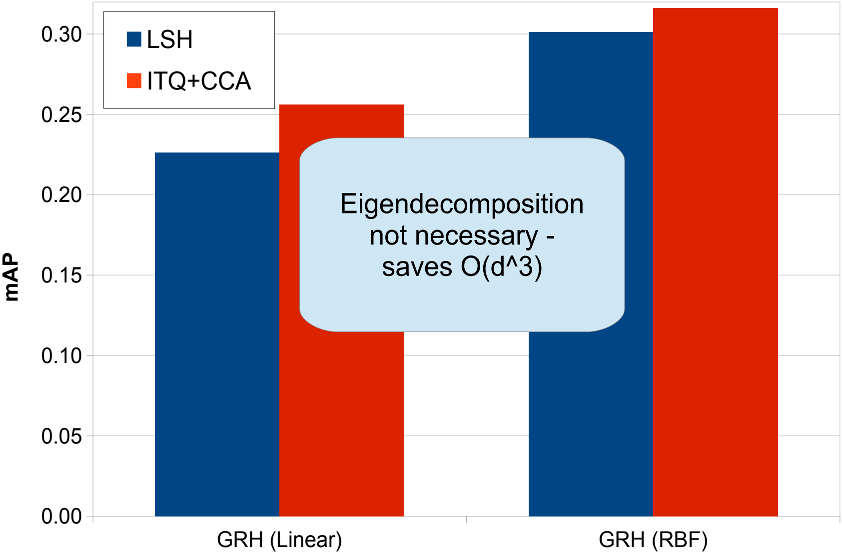
GRH vs Literature (CIFAR-10)



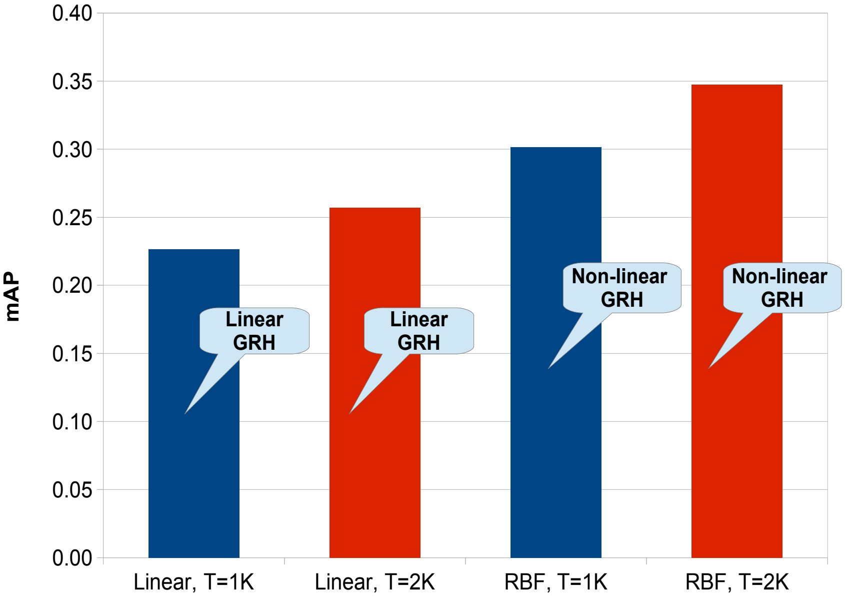
GRH vs. Initialisation Strategy (CIFAR-10 @ 32 bits)



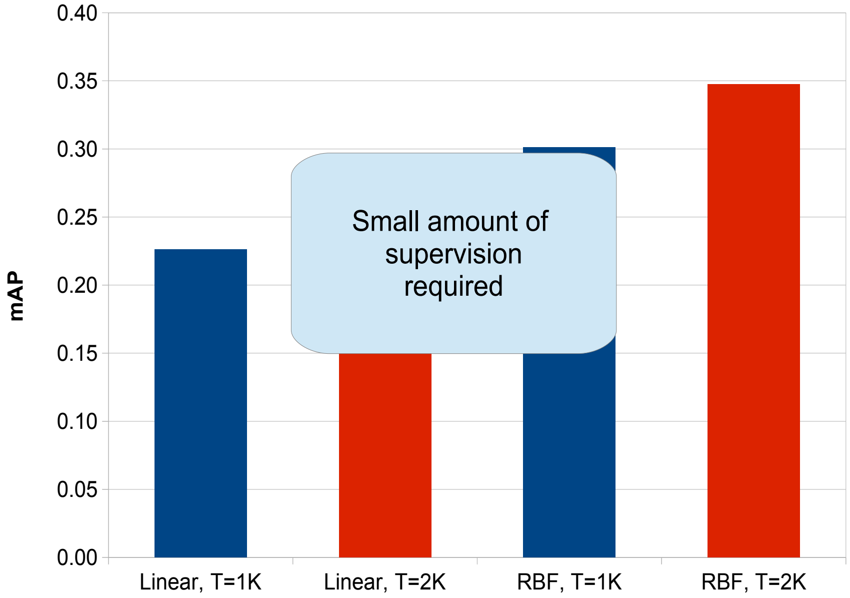
GRH vs. Initialisation Strategy (CIFAR-10 @ 32 bits)



GRH vs # Supervisory Data-Points (CIFAR-10)



GRH vs # Supervisory Data-Points (CIFAR-10)



GRH Timing (CIFAR-10 @ 32 bits)

Timings (s)			
Method	**Train**	**Test**	**Total**
GRH	42.68	0.613	**43.29**
KSH [1]	81.17	0.103	82.27
BRE [2]	231.1	0.370	231.4

[1] Liu, W.: Supervised Hashing with Kernels. In: CVPR (2012)
[2] Kulis, B.: Binary Reconstructive Embedding. In: NIPS (2009)

Conclusion

Conclusions and Future Work

- *Supervised* hashing model that is both *accurate* and easily *scalable*
- Take-home messages:
  - Regularising bits over a graph is effective (and efficient) for hashcode learning
  - An intermediate eigendecomposition step is not necessary
  - Hyperplanes (linear hypersurfaces) can achieve a very good retrieval accuracy
- Future work: extend to the cross-modal hashing scenario (e.g. Image $\leftrightarrow$ Text, English $\leftrightarrow$ Spanish)

Thank you for your attention

Sean Moran

Code and datasets available at:

sean.moran@ed.ac.uk
www.seanjmoran.com