USING DEEP CROSS MODAL HASHING AND ERROR CORRECTING CODES FOR IMPROVING THE EFFICIENCY OF ATTRIBUTE GUIDED FACIAL IMAGE RETRIEVAL

Veeru Talreja, Fariborz Taherkhani, Matthew C. Valenti, Nasser M.Nasrabadi
Lane Department of Computer Science and Electrical Engineering
West Virginia University

WestVirginiaUniversity

Outline

Problem and Motivation

Introduction

Related Work

Error-Corrected Deep Cross-Modal Hashing

Implementation

Results

Conclusion

Outline

Problem and Motivation

- A **fast image** retrieval system is needed for **quickly** searching over a **large gallery of faces** based solely on a **soft-biometric query** (i.e., facial attributes).
- Such a system would **drastically down select** the number of suspects for **line-up** or **post processing** applications.
- The solution is a **multimodal** system, with one modality being **face** and the other being the **soft-biometric attributes**.
- **Issues** and design **goals**:
  - How to map the **soft biometrics** of subjects and their corresponding **facial photo** from the original space into a **common latent subspace**?
  - The **semantic information** across the two modalities in the original space needs to be **preserved** in the **latent subspace** for fast **cross-modal retrieval**.

Input
Search String

Output
Virtual Line-Up

WestVirginiaUniversity

Outline

Problem and Motivation

Introduction

Related Work

Error-Corrected Deep Cross-Modal Hashing

Implementation

Results

Conclusion

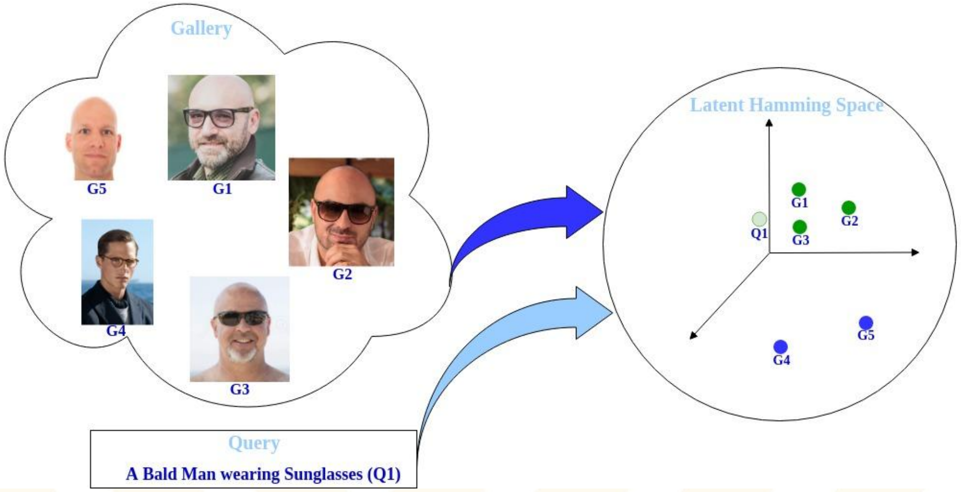
Introduction

- **Facial Attributes**
  - Invariant and **semantic visual properties** (i.e., visual properties that have names).
  - The **content** of a face image can be described by its **facial attributes** such as "a bald old man wearing glasses".
  - Have been used in a variety of computer vision applications such as **face search engine** and **face image retrieval**.
  - Have been **significantly exploited** by the biometric society to **improve performance** of object recognition, face verification and image search.
- **Hashing**
  - A **fast and an advantageous** solution for an approximate binary nearest neighbors (ANN) in **image retrieval**.
  - **Transform high-dimensional** media data into **similarity-preserving binary codes** for efficient image search.

https://brightside.me/inspiration-psychology/14-facial-features-and-personality-traits-that-everybody-loves-377660/

Introduction

- **Cross Modal Hashing**
  - Cross modal hashing returns **relevant results** for **one modality** in response to a **query** in **another modality**.
  - **Binary hash codes** in the same **latent hamming space** are generated for each individual modality.
  - Mostly commonly applied in **text-based image retrieval (TBIR)** and **image-based text retrieval (IBTR)**.

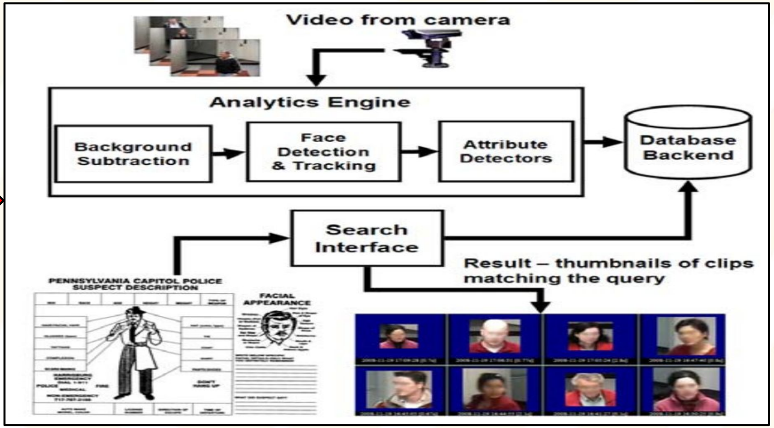


- **Deep Cross Modal Hashing**
  - Application of **deep learning techniques** for **cross modal hashing**.
  - **End-to-end learning** of binary codes in a common latent space for both modalities.
  - **Improved cross modal retrieval** performance when compared to **hand crafted feature models**.

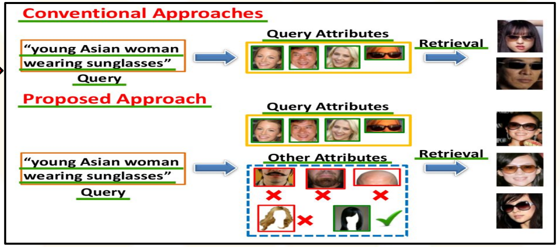
Outline

Related Work

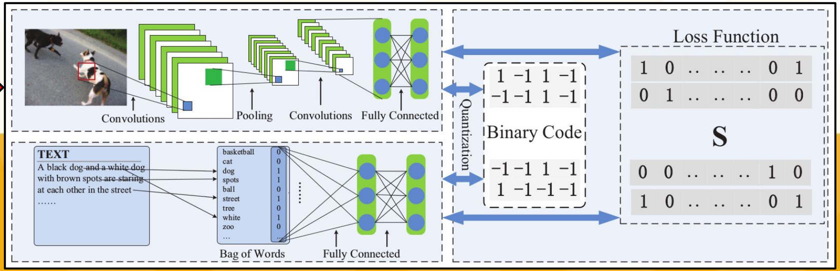
- D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk, "Attribute-based people search in surveillance environments," WACV, Dec 2009.
  - Uses face detection and tracking to search for people in surveillance systems based on a parsing of human parts and their attributes, including facial hair, eye-glasses, clothing color, etc.



- B. Siddiquie, R. S. Feris, and L. S. Davis, "Image ranking and retrieval based on multi-attribute queries," CVPR, June 2011.
  - Uses the concept of reverse learning and hand-crafted features for image retrieval/ranking.



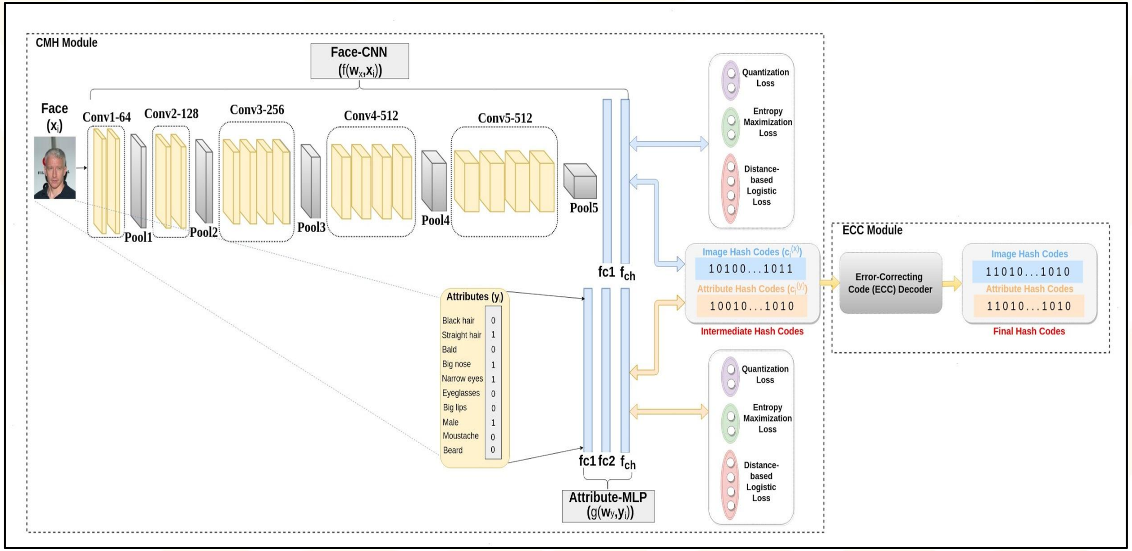
- Q.-Y. Jiang and W.-J. Li, "Deep cross-modal hashing," CVPR, June 2017.
  - Uses deep cross modal hashing for image retrieval using text query.



Outline

Error-Corrected Deep Cross-Modal Hashing (CMH-ECC)

- **Error-corrected deep cross modal hashing (CMH-ECC)**
  - Utilize **deep cross modal hashing** and **error-correcting codes** for **face image retrieval** in response to an **attribute query.**
  - Has **not** been done **previously**.
- **Scalable cross-modal hash**
  - Requires **neither pairs nor triplets** of training inputs.
  - CMH-ECC performs facial image retrieval using **point-wise data**.
  - This characteristic makes it **scalable** to large scale datasets
- **Contains two modules**
  - **Cross-modal hashing** module
  - **Error-Correcting code** module

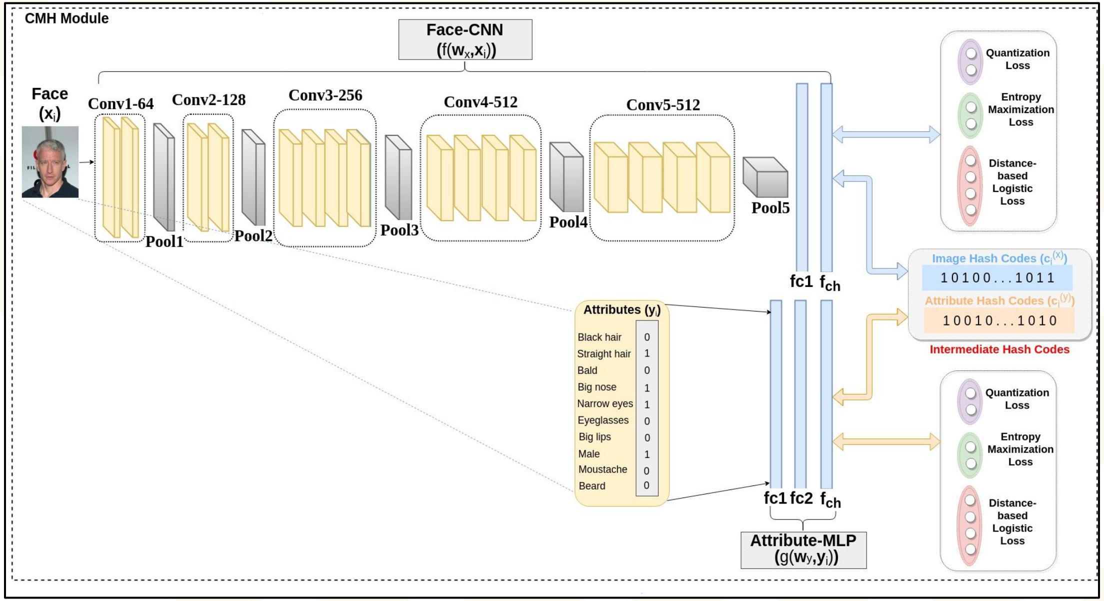


Error-Corrected Deep Cross-Modal Hashing (CMH-ECC)

Cross-Modal Hashing Module

- The proposed architecture integrates **two Deep NNs – one for the soft-biometric modality** and the other one for the **face modality**, which are **coupled together** through their common hash-codes.
- The two DNNs are **simultaneously trained** on a database of cross-modal pairs and are **coupled together** by forcing their binary hash-codes to be close to preserve the **cross-modal similarities** of each pair.
- The system performs better when using **distance-based logistic loss**, which has not been used previously for cross-modal hashing.
- **Intermediate hash codes** are generated using the CMH module.

CMH module Cost Function (Three Objectives: Entropy Maximization, Minimizing Quantization Error, Distance-based Logistic Loss)



Error-Corrected Deep Cross-Modal Hashing (CMH-ECC)

Cross-Modal Hashing Module

- **Convolutional Neural Network (CNN)**
  - Used to extract features for **image modality.**
  - Initialized with **VGG 19 network** with **same** filter size, convolutional layers and pooling layers.
  - Number of nodes in the **last layer** is equal to required **hash code length**
  - Fine-tuned using **CASIA-WebFace**
- **Multilayer Perceptron (MLP)**
  - Used to extract features for **attribute modality**.
  - Contains only **3 fully connected layers** with **4096 nodes** in the first **two nodes** and the number of nodes in the **last layer** is equal to the hash code length.
  - Activation function for first 2 layers is **rectified linear unit (ReLU)** and the activation for last layer is an **identity function**.
  - Input to the MLP is a **bit map** indicating the presence or absence of corresponding **facial attribute**.
- **Optimization for training the CMH**

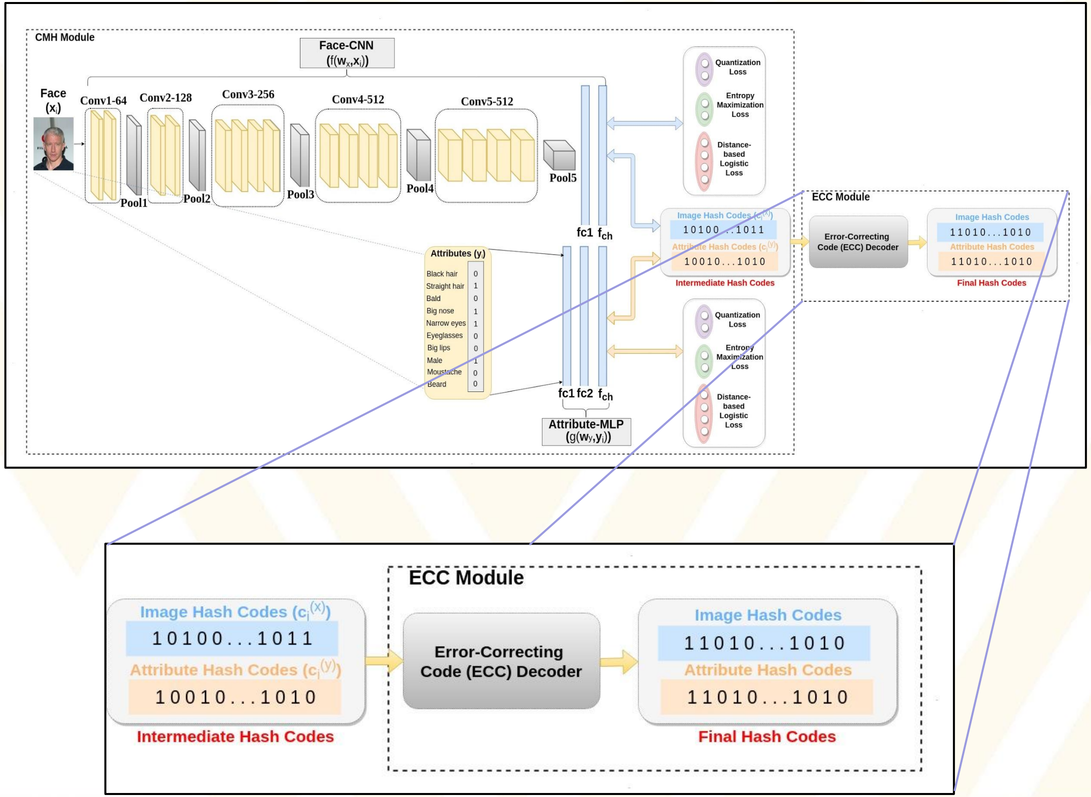
$$\min_{\mathbf{C}_{x,y},\mathbf{w}_x,\mathbf{w}_y} \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{n} \ell_c(p(\mathbf{F}_{*i},\mathbf{G}_{*j}),S_{ij})}_{\text{distance- based logistic loss}} +\alpha \underbrace{\left(||\mathbf{F}-\mathbf{C}_x||_F^2 + ||\mathbf{G}-\mathbf{C}_y||_F^2\right)}_{\text{quantization loss}} +\beta \underbrace{\left(||\mathbf{F1}||_F^2 + ||\mathbf{G1}||_F^2\right)}_{\text{entropy maximization}}$$

$$\text{s.t. } \mathbf{C}_{x,y} \in \{+1,-1\}^{c\times n}$$

Error-Corrected Deep Cross-Modal Hashing (CMH-ECC)

Error-Correcting Code Module

- The **intermediate hash code** generated by the CMH module is a binary vector that is within a **certain distance** from a **codeword** of an **error-correcting code (ECC)**.
- The intermediate hash code is passed through an **appropriate ECC decoder**, the **closest codeword** is found and this closest codeword is used as a **final hash code** for the retrieval process.
- Benefits of ECC
    - The **attribute** hash and **image** hash of the same subject are (usually) mapped to the **same** codeword, thereby reducing the distance of the corresponding hash codes.
    - Brings more **relevant facial images** from the gallery **closer** to the **attribute query**, which leads to **improved** retrieval performance.



WestVirginiaUniversity

Outline

- Problem and Motivation
- Introduction
- Related Work
- Error-Corrected Deep Cross-Modal Hashing (CMH-ECC)
- Implementation
- Results
- Conclusion

Implementation and Datasets

- **Implementation Details**
    - **Alternative minimization** used to train the parameters.
    - Used **Adam optimizer** with default hyper-parameter values.
    - **Batch Size** used is 128.
    - Implemented in **Tensorflow** with Python API.
    - Implemented using **two NVIDIA GeForce GTX TITAN X 12GB GPU**s.
    - Used **Reed-Solomon codes** in the **ECC** module with **code rate=0.5**

- **Datasets**
    - First dataset used was **Labeled Faces in the Wild (LFW)**.
    - LFW contains more than **13,000 images of faces** collected from the internet for face recognition
      as well as attribute classification .
    - Second dataset used was **Face Tracer dataset.**
    - FaceTracer conatins **15000 real-world face images**, collected from the internet.
    - Both LFW and FaceTracer are annotated with **attribute specifications**.

Outline

Problem and Motivation

Introduction

Related Work

Error-Corrected Deep Cross-Modal Hashing

Implementation Details and Dataset

Results

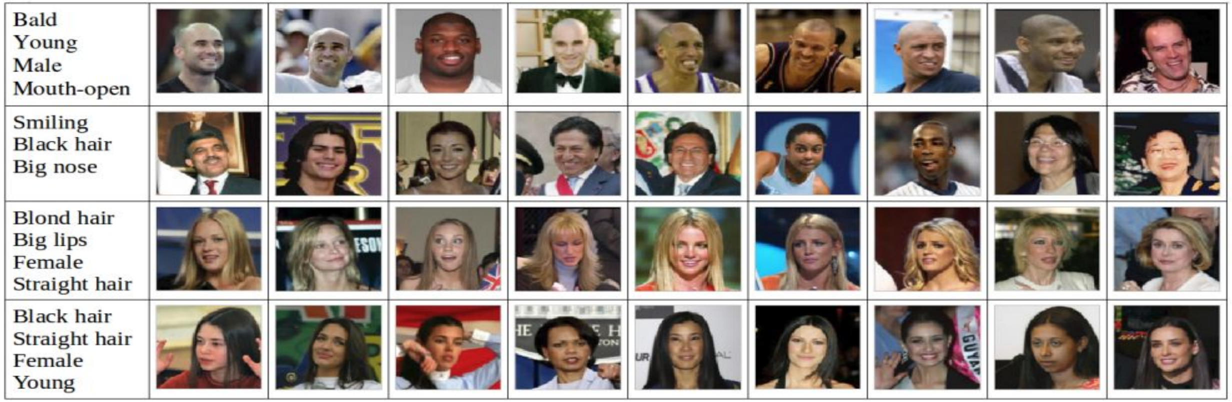
Conclusion

WestVirginiaUniversity.

Results

- ## Evaluation Metric
  - We use **normalized discounted cumulative gain (NDCG)** to compare ECC-CMH performance with other methods.
  - NDCG is a **standard single-number** measure of **ranking** quality that allows **non-binary** relevance judgments.
  - NDCG is given as : $\frac{1}{Z}\sum_{i=1}^{k}\frac{2^{rel(i)}-1}{log(i+1)}$

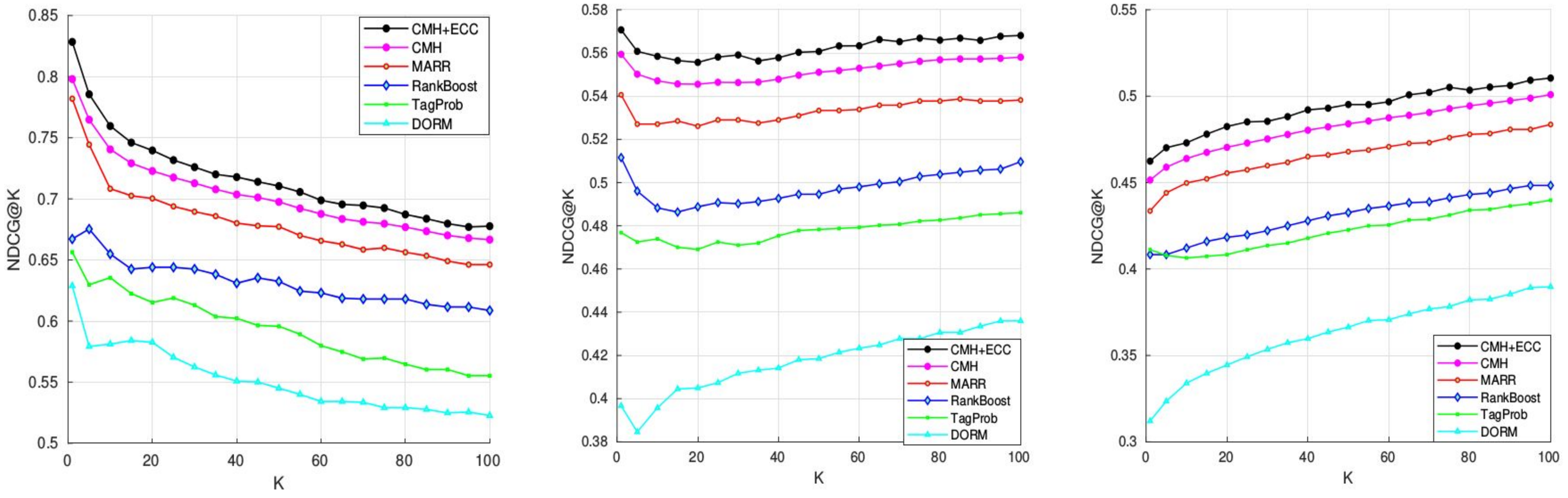
    where rel(i) is the relevance of the ith ranked image and Z is a normalization constant to ensure that the correct ranking results in an NDCG score of 1.
  - Compared our retrieval and ranking results with some of the other state-of-the-art ranking approaches including **Multi Attribute Retrieval & Ranking (MARR), rankBoost, Direct Optimization of Ranking Measures (DORM), TagProp**.
- ## Qualitative Results



WestVirginiaUniversity

Results

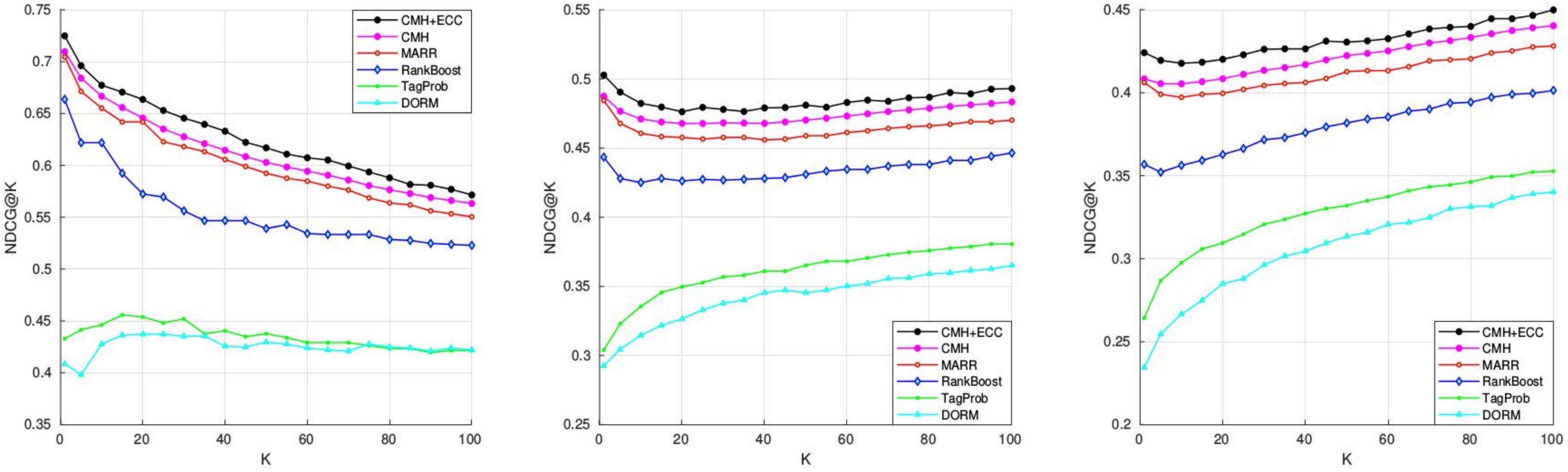
Ranking performance on the LFW dataset using normalized discounted cumulative gain (NDCG)



(a) Single Attribute Queries (b) Double Attribute Queries (c) Triple Attribute Queries

Results

Ranking performance on the Face Tracer dataset using NDCG



(a) Single Attribute Queries (b) Double Attribute Queries (c) Triple Attribute Queries

Outline

Problem and Motivation

Introduction

Related Work

Error-Corrected Deep Cross-Modal Hashing

Implementation

Results

Conclusion

Conclusion

- This proposed work **exploits the relationship** between **soft** and **hard** biometric modalities for **fast retrieval** from a large scale multi-modal biometric database.
- An algorithm combining **deep hashing** and **error-correcting codes** has been presented with application in **cross-modal retrieval**.
  - Using only **deep hashing** for face image retrieval gives **improved ranking** performance when compared to state-of-the-art methods.
  - The performance is **further improved** by using **error-correcting codes** in combination with deep hashing.
- The use of error-correcting codes for **improving** the **performance** of cross-modal retrieval is **novel**.
- The experimental results on two popular public datasets shows that our method **outperforms** the **current face image retrieval** approaches in the literature

References

- H. Liu, R. Wang, S. Shan, and X. Chen, “Deep supervised hashing for fast image retrieval,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2016, pp. 2064–2072.
- Zhangjie Cao, Mingsheng Long, Jianmin Wang, and Philip S Yu, “Hashnet: Deep learning to hash by continuation,” arXiv preprint arXiv:1702.00758, 2017.
- Erkun Yang, Cheng Deng, Wei Liu, Xianglong Liu, Dacheng Tao, and Xinbo Gao, “Pairwise relationship guided deep hashing for cross-modal retrieval,” in Proc. AAAI, Jan. 2017.
- Qing-Yuan Jiang and Wu-Jun Li, “Deep cross-modal hashing,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2017, pp.3232–3240.
- Neeraj Kumar, Peter Belhumeur, and Shree Nayar, “Facetracer: A search engine for large collections of images with faces,” in Proc. European Conference on Computer Vision (ECCV), Springer, 2008, pp. 340–353.
- D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur, and M. Turk, “Attribute-based people search in surveillance environments,” in Proc. Workshop on Applications of Computer Vision (WACV), Dec. 2009, pp.1–8.
- B. Siddiquie, R. S. Feris, and L. S. Davis, “Image ranking and retrieval based on multi-attribute queries,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2011, pp. 801–808.
- Fariborz Taherkhani, Nasser M Nasrabadi, and Jeremy Dawson, “A deep face identification network enhanced by facial attributes prediction,” in Proc. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), June 2018.
- V. Talreja, M. C. Valenti, and N. M. Nasrabadi, “Multibiometric secure system based on deep learning,” in Proc. IEEE Global Conference on Signal and Information Processing (GlobalSIP), Nov. 2017, pp. 298–302