MindCam: An Approach for Sketch Based Image Retrieval

Sini Thankachan[1]Smita C Thomas[2]
[1]Mount Zion College of Engineering, Pathanamthitta, India sinibinoy2013@gmail.com
[2]Mount Zion College of Engineering, Pathanamthitta, India smitabejoy@gmail.com

ABSTRACT

Image retrieval plays a major role in computer vision applications.FaceSketch Based Image Retrieval (FSBIR) draws attention with its wide application in criminal investigation by law enforcement agencies. Automatic retrieval of face images from police mug-shot databases guides investigators to locate or narrow down the search area. MindCam is an efficient image retrievalsystem based on face sketch queries. MindCam leverages Scale invariant Feature Transform (SIFT)to figure out the interest points and extract the features independent of scale.Binary hashing is used to compute binary hash codes from SIFT feature descriptors.This makes feature matching more efficient by computing the Hamming distance of the binary codes. Pool of matching candidates is identified based on Hamming distance and Fine Grained Matching is done on candidates in the pool. A user interactive approach is guaranteed by re-ranking of retrieved results. Reduced computational complexity makes MindCam suitable for realistic applications like mobile devices and wearable devices. System also provides audio description of results which makes the system user friendly.

Key words: Binary Hashing, Face Sketch Recognition, MindCam, Hamming Distance, SIFT , Sketch Based Image Retrieval.

1. INTRODUCTION

A computer system for browsing, searching and retrieving images from a large digital image database is termed as an Image Retrieval system. Sketch Based Image Retrieval(SBIR) is used where precise queries are not available to be used with the conventional approaches like Content based image retrieval and Text based Image retrieval. Face images play a crucial role in modern intelligent systems. Face image retrieval has invited substantial attraction for law enforcement, surveillance, authentication andsecurity control. Recognition of facial images using sketch has an extended application as criminal face sketching is still a routine part of investigation and law enforcement. The recollection of information about the criminal by the eyewitness will become vital in cases where criminals have left no traces behind. The verbal description given by eyewitness can be used by forensic artists to figure out the sketch of the culprit. This is the vital information used by law enforcement agencies. Existing systems of criminal investigation and law enforcement agencieslike police department identify the image from sketch by exposing it in public places or will be given to a dedicated forensic examiner who will do the manual comparison of sketch with gallery images. Mug shot galleries are too large that it will be slow and difficult to perform sketch to photo matching for criminal identification.Criminal investigators then go ahead with the top N retrieved results and this reduces the probability of finding the correct match. Automation of this process can potentially make the search easier and reduce the number of suspects.

Major Challenges with Face Sketch Recognition (FSR) are
(i) Cross domain mismatch between input and output.
(ii) Sketches are inaccurate depictions of the face i.e. accuracy of sketch depends onthe memory of the eyewitness.
(iii) Demands anexhaustive search compared to general SBIR. Feature extraction and matching have tobe performed for close features which will be difficult to match and distinguishaccurately.
(iv) Computational complexity involved in matching very deep features of largedatasets.

2. LITERATURE REVIEW

Various studies have attempted to address thechallenges of face sketch recognition. Basic idea of Canny edge detection is given by [1]. Histogram of Oriented Gradient approach for feature extraction is highlighted in [2]. [3] proposed an approach of SBIR by creating sketch tokens as reference to retrieve images based on sketch.

A Sketch Based Image Retrieval System based on Learned Key Shapes (LKS)[4] retrieves images based on set of mid-level patterns. LKS Histogram are built and are used for feature matching. Minkowski's functions or Hellinger distance is used to measure thedistance for comparison which makes the system complex.

Different approaches of face sketch recognition are given by 10] and [12].A content based image retrieval approach for face photo recognition is given in [9].SBIR for face sketch queries [6] using hand-drawn face sketches which are collected from five different artists. Mahalanobis distance is used as a classifier in this approach. However, the use of distance measures like Mahalanobis distance for

the large dataset makes the feature matching and retrieval complex.

An approach based ontransformation of sketch to photo is used in [8]. Face sketch is transformed to a pseudo-photo which will be used for the retrieval process. This approach may compromise in accuracy as it depends on the accuracy of synthesised photo from sketch.

Face Sketch Recognition (FSR) using conventional Eigen Face approach [11] is used to retrieve facial images for criminal detection agency applications. An Eigen sketch transformation algorithm is used to convert a photo into a sketch at first, and then classification is performed using Eigen sketch features. However, this approach is scale sensitive.

[5] and [7] give insight to recognition of images from sketch using convolutional neural networks. These approaches demand highcomputational cost and large number of images to train for a particular image. However, it is not always feasible to obtain large number of different images of a criminal to store in mug shot gallery. The increased computational and storage cost make these approaches unsuitable for realistic applications.

3. EXISTING SYSTEM

Basic steps in Sketch Based Image Retrieval are:
   i.   Image pre-processing
   ii.   Feature extraction
   iii. Feature matching
   iv.   Image retrieval

Image pre-processing involves extraction of edges from the natural images inorder to approximate the sketch of the image. Usually, Canny Edge Detection is used and it is shown in figure1. Major limitation of Canny edge detection is that the Gaussian smoothing blurs corners and junctions out, making them harder to detect. It needs to connect the resulting edges to extract the complete edges that are obvious for the human eye. Also, the corner pixels look in the wrong directions for their neighbours, leaving open ended edges, and missing junctions.

Feature extraction involves extraction of features to be matched with the features of input query. Various Face feature extraction techniques are being used. Some of feature descriptors are Eigen face, Histogram of oriented gradients etc. However, these approaches are sensitive to scale and orientation which may result in inaccurate output. Features are extracted from input sketch query and are matched with features of imagesin the database.

Usually, a K-Nearest Neighbour approach is used formatching of sketch features. Approaches used for distance measurement includes Euclidean distance, Hellinger distance, Minkowski distance etc. These distance measures when applied directly to huge number of features in a large dataset, increases the complexity of the system and this in turn slows down the system. This makes the system inappropriate for realistic applications like wearable and mobile devices.

4. PROPOSED SYSTEM

The proposed system, MindCam is an efficient Sketch Based Image Retrieval system especially for face images.MindCam uses bilinear interpolation to predict the data points in order to make the edges detected by Canny edge detection accurate.It uses Scale Invariant Feature Transform (SIFT) feature descriptor which is invariant to scale, orientation etc. To address the problem of complexity in distance calculations, MindCam uses hash functions to obtain binary codes. It encodes high dimensional features into low dimensional binary codes which enables the system to perform fast similarity searches using Hamming distances. Pool of matching candidates are identified and searching is done on candidates in the pool. This adds on to the efficiency of system.

4.1 Preprocessing

The Canny edge detection is a multistage edge detection algorithm. The image is first smoothed using Gaussian convolution. A two dimensional first derivative operator is applied to the smoothed image obtained from Gaussian convolution. This highlights the regions of the image with high first spatial derivatives. Edges result inridges in the gradient magnitude image. A non-maximal suppression is done next where the algorithm then tracks along the top of these ridges and sets all pixels that are not present on the top of the ridge to zero in order to produce a thin line in the output. The tracking process exhibits hysteresis controlled using two thresholds: T1 and T2, where T1 > T2. Tracking begins at a point on a ridge higher than T1. Tracking continues in both directions out from that point until the height ofthe ridge falls below T2. This hysteresis allows the noisy edges to be broken up into multiple edge fragments.

Figure 1: Edge Detection

To address the limitation of open ended edges detected using Canny approach, interpolation is done for the approximation resulted from Canny edge detection.Interpolation is the process by which values of a continuous function are estimated from discrete samples.Bilinear Interpolation computes the missing datapoint value from the weighted average of the four closest pixels. Bilinear interpolation is the combination of two linear interpolations. First one is performed in one direction and then another one is performed in the perpendicular direction.Bilinear interpolation is applied for edge pixelsto obtain the boundary of the image to be matched with the sketch.

4.2 Feature Extraction

A combined approach using Harris Corner detection and Scale Invariant Feature Transform (SIFT) is used to find the key points and extract the features. Harris corner detector efficiently locates the corners and is rotation invariant, but is not scale invariant. SIFT is an even more stable method to figure out key points that are scale invariant. Thus it is profitable to make use of these approaches together to gain from the advantage from both. Harris Corner detection is performed at first and SIFT is applied for the Harris detected points. This adds on to the reduction of computational complexity of the system. Major steps involved SIFT are as follows.

A. Scale Shape Extreme Detection
To obtain invariance over scale and orientation, at first different scales and orientations of pixels are computed and searches over all scales and image locations. This is implemented using a difference-of-Gaussian function. Potential interest points that are invariant to scale and orientation are identified.

B. KeyPoint Localization
At each interest point identified, a detailed model is fit to obtain location and scale. Based on measures stability or invariance, key points are selected.

C. Orientation Assignment
Once key points are identified, different orientations are assigned to each keypoint location based on gradient direction.Future operations are done on image data that is transformed relative to the assigned scale, orientation andlocation for each feature. This ensures invariance of features to these transformations.

D. KeyPoint Descriptor
At the selected scalein the region around each key point, local image gradients are measured. These are converted to representations which are used as the feature vectors.

4.3 Binary Hashing

Matching real valued vectors in a big database is slow and tiring. HenceMindCam leverages the approach of Binary Hashing for feature matching to reduce the overall complexity of system. Hashing is the process of converting data values of arbitrary length to values of fixed size. MindCam uses hashing to project high dimensional features to low dimensional space in order to generate compact binary codes corresponding to features. Given an image I, MindCam obtains the edge approximation and extractsthe features. SIFT feature descriptor has 128 values corresponding to each key point. The output of SIFTprocess is denoted by S(i). The binary codes are obtained by converting the feature values using threshold. For each feature value,

$$H_i=1, \text{ if } S(i) \geq T(1)$$
$$0, \text{ otherwise}$$

where T is the threshold value for binary hashing of SIFT features. For a given key-point, median of the SIFT features at that point is taken as threshold. Let $D = \{I_1, I_2, . . . , I_n\}$ denotes the dataset of images in the database which is used for retrieval. Binary codes are generated from the key point descriptors for each key point in each image I_i.

4.4 Feature Matching

For the user input sketch query, a coarse level matching is done at first using Hamming distance of binary codes. Given a query sketch S_q,Features are identified using Harris and SIFT approaches and binary code HS_q is obtained for each feature. Hamming distance is calculated for the obtained binary codes and the binary codes of the SIFT descriptors of images in database. A predetermined threshold valueD is used for comparing the Hamming distance between each pair. During Pooling, candidates whose hamming distance falls within the threshold get accumulated to the same pool. Fine grained search is done on the candidates in the pool.

5. SYSTEM OVERVIEW

MindCam consists of two main phases. Training phase and Testing Phase.

5.1 Training Phase

During the training phase, administrator adds new images to the database. Image undergoes a pre-processing stage involving Canny edge detection and bilinear interpolation. The edge approximations obtained after pre-processing of the image are fed into the feature extraction module. A combined approach using Harris corner detection and SIFT identifies the interest points and extracts the features. These feature vectors are used in binary hashing process to

obtain binary codes to be matched during feature matching. A brief overview of the flow of the system during training phase is given in Figure 2.
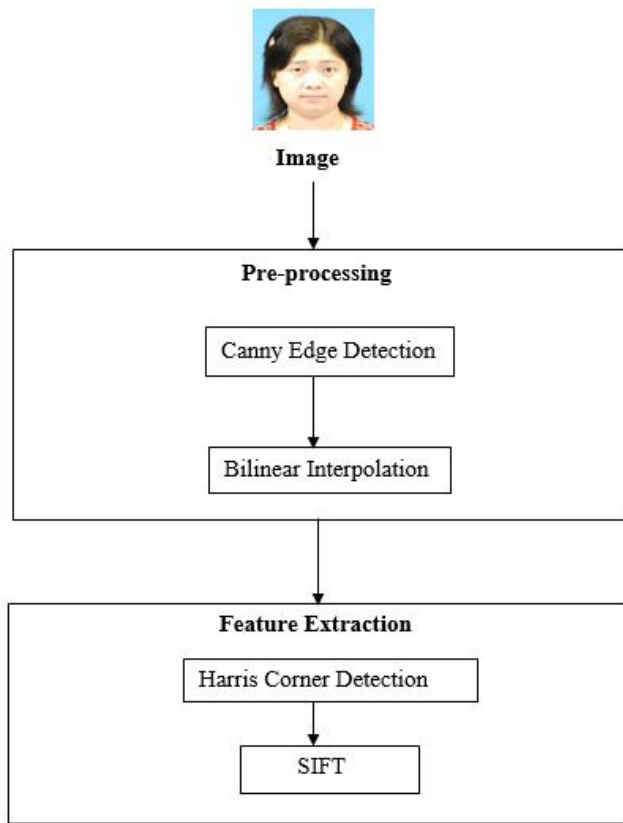

Figure 2: Training Phase

5.2 Testing Phase

User query in the form of sketch will be processed in the testing phase. Flow of the system during the testing phase is shown in Figure 3.Features are extracted by identifying the interest points and obtaining the features using Harris corner detection and SIFT approaches. Binary hashing is applied to the features to obtain the binary codes. Hamming distance is computed for the obtained binary codes and the binary codes of the features of the stored images in the database. Computed Hamming distance is compared with a predefined threshold. Feature matrices having Hamming distance with in the threshold will be populated to a pool. Thus pooling narrows down the potential search area and this considerably reduces the computational complexity in case of large data set. Only the candidates belonging to the pool are considered for fine grained matching.

During fine grained matching, Euclidean distance is used as the distance measure to obtain the matches. This enables to locate the exact matches from the pool. The retrieved results will be displayed to user.

MindCam is a user interactive approach where the user is given facility to re-rank the results. The intermediate results matched from the pool will be displayed to user. In case of criminal image identification from mug shot gallery, results will be displayed to the eyewitness. The eyewitness can re-rank the results according to his memory. The input given by eyewitness will used by MindCam to refine the results and final desired results will be displayed to user.
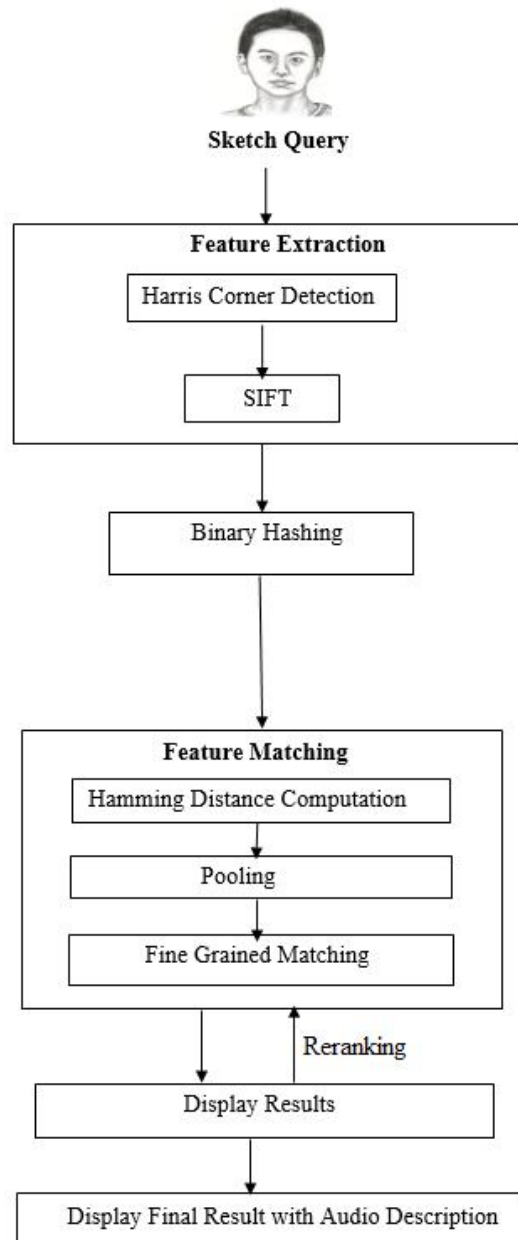

Figure 3: Testing Phase

MindCam also provide audio description of the details of the image, which makes the system more user friendly. It enables crime investigation officials to get the quick idea of the details of criminals.

6. CONCLUSION

MindCam is an efficient SBIR system, designed especially for the retrieval of face images based on face sketch queries. Use of interpolation with edge detection effectively bridgesthe domain gap between sketches and images. Use of SIFT approach exerts the advantage of scale invariance to the system which makes the retrieval more accurate. MindCam effectively reduces computational complexity of existing system by using binary hashing approach. Benefitting from learned binary codes, similaritymatch is performed using Hamming Distance which reduces the computationalcomplexity.Pooling narrows down the potential search area and this results in reduced computation in case of large data sets.This in turn makes the system suitable for realistic applications such as mobile devices and wearable devices. MindCam assures a user interactive approach of imageretrieval by providing re-ranking facility to users. Audiodescriptions along with the details of the retrieved image makes the system friendlierto user.

7. ACKNOWLEDGEMENT

REFERENCES

[1] J. Canny. **A computational approach to edge detection**, *IEEE Transactions onPattern Analysis and Machine Intelligence*, Nov 1986, Vol.8, No.6, 679-698.
https://doi.org/10.1109/TPAMI.1986.4767851

[2] N. Dalal and B. Triggs. **Histogramsof oriented gradients for human Detection**, *IEEE Conference on Computer Vision and Pattern Recognition, SanDiego, USA*, 2005, 886-893.

[3] J. J. Lim, C. L. Zitnick and P. Dollar. **Sketch tokens:A learned mid-level representation for contour and object detection**, *IEEE Conference on Computer Vision and Pattern Recognition, Portland, USA*, 2013, 3158-3165.
https://doi.org/10.1109/CVPR.2013.406

[4] J. M. Saavedra, J. M. Barrios, and S. Orand. **Sketch based imageretrieval using learned keyshapes (LKS)**, *British Machine Vision Conference,Swansea, USA*,2015, 164.1-164.11.

https://doi.org/10.5244/C.29.164

[5] Christian Galea and Reuben A. Farrugia. **Forensic Face Photo-Sketch Recognition Using a Deep Learning-Based Architecture**, *IEEE Signal Processing Letters*, 2017, Vol. 24, No. 11, 1586-1590.
https://doi.org/10.1109/LSP.2017.2749266

[6] Y. Zhang, C. McCullough, J. R. Sullins, and C. R Ross.**Handdrawn face sketch recognition by humans and a pca-based algorithm for forensic applications**,*IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 2010, vol. 40, no. 3, pp. 475– 485
https://doi.org/10.1109/TSMCA.2010.2041654

[7] Li Liu, FuminShen, YumingShen, Xianglong Liu and Ling Shao. **DeepSketch Hashing: Fast Free-hand Sketch-Based Image Retrieval**, *IEEE Conference onComputer Vision and Pattern Recognition, Honolulu, USA*, 2017, 2298-2307
https://doi.org/10.1109/CVPR.2017.247

[8] Shu Zhan, Jingjing Zhao, Yucheng Tang and ZhenzhuXie.**Face imageretrieval: super-resolution based on sketch-photo transformation**, *Issue on SoftComputing*, 2018, Vol 22, Issue 4, 1351 – 1360.34
https://doi.org/10.1007/s00500-016-2427-0

[9] Salah EddineLahlali, AbdelalimSadiq and Samir Mbarki. "**Face SketchRecognition System- A Content Based Image Retrieval Approach**, *IEEEInternational Colloquium on Information Science and Technology, Tangier, Morocco,*2016, 428-433.
https://doi.org/10.1109/CIST.2016.7805085

[10] Pong C. Yuen and C. H. Man. **Human Face Image Searching System UsingSketches**, *IEEE Transactions on Systems, Man, and Cybernetics— Part A: Systemsand Humans*, Jul 2007, Vol. 37, No. 4, 493-504.
https://doi.org/10.1109/TSMCA.2007.897588

[11] Xiaoou Tang andXiaogang Wang. **Face Photo Recognition Using Sketch,**International Conference on Image Processing, Rochester, USA*, 2002, 257-260.

[12] Kokila R, Sannidhan MS and AbhirBhandary. **A Novel approach for Matching Composite Sketches to Mug shot Photos using the Fusion of SIFT andSURF Feature Descriptor**, *International Conference on Advances in Computing, Communications and Informatics, Udupi, India*, 2017, 1458-1464.
https://doi.org/10.1109/ICACCI.2017.8126046