EAI Endorsed Transactions
on Internet of Things Research Article EAI.EU

Perlustration on Image Processing under Free Hand Sketch Based Image Retrieval

S. Amarnadh[1,*], P.V.G.D. Prasad Reddy[2] and N.V.E.S. Murthy[3]

[1]Department of I.T, G.I.T, GITAM Deemed to be University, Visakhapatnam, India
[2]Department of CS&SE, Andhra University, Visakhapatnam, India
[3]Department of Maths, Andhra University, Visakhapatnam, India

Abstract

In general information retrieval has taken vast diversions in visualizing the content presentation for the users who generates the queries for the system, where it includes the concept of content based Image Retrieval to provide the results in a better way by adapting the approaches like Text Based Image Retrieval(TBIR) and Sketch Based Image Retrieval(SBIR). Nowadays the concept of search engines relies on deeper vision of content and perhaps interested in providing the results effectively by employing several algorithms from the areas like machine learning, neural networks, Fuzzy Logic and deep learning concept. As the digital world is spreading its dimensions, the supporting environments have been diversified from traditional computers to the mobile based environments. The idea of this paper is to present the survey on sketch based image retrieval adapting deep learning concept on the mobile platforms, by presenting various methodologies and techniques.

Keywords: IR, TBIR, SBIR, Deep Learning, Neural Networks, Machine Learning, Fuzzy Logic.

Received on 07 August 2018, accepted on 10 October 2018, published on 31 October 2018

doi: 10.4108/eai.21-12-2018.159334

* Corresponding author: amarnadh07@gmail.com

1. Introduction

The digital media over the internet is taking new diversions with respect to the user requirements on demand basis, hence lead to development of latest environments in devices such as tabs, mobile devices has come into existence. The generation of user interfaces over these devices has increased the flexibility of usage even for a naive user. In earlier days the user interfaces were designed with respect to the conventional use of traditional desktops or laptops. The key idea of digital technology is to simplify the design of user interface for flexible usage and minimize the design of the device for the portability, which leads to the concept of wherever you go access it (WUGA) systems. The concept of drawing based image retrieval has come into existence with respect to the changing requirements of the interface designs, which includes the techniques of content based image retrieval namely Sketch based technique which accepts input in the form of free hand sketches and performs retrieval process from the millions of image database using the techniques and methods like deep convolutional neural networks, SHELO and SHOG.

2. Related Work

Li Liu *et al.*[1], has projected Deep Sketch Hashing (DSH), technique which uses three convolutional neural networks [13] (CNN) to encode normal images, free-hand sketches, and the sketch-tokens which act as a bridges to provide simplicity in sketch-image geometric distortion. It includes: Semi heterogeneous Deep Architecture used to ease the geometric distortion among normal and sketch images. The “Sketch tokens” are referred as a set of edge structures identified from normal images, with the help of supervised information in the format of hand-drawn sketches. Initially an image is given and initial sketch-token is obtained, in which each pixel is initialized a score for likeliness of it acting as a contour point. They used the maximum score of 60% to mark as a threshold for each pixel and obtain the final sketch-tokens. The three forms of CNNs named Cross-weight Late-fusion Net, (C1-Net) and Shared-weight Sketch Net are considered as hash functions to train free-hand sketches, sketch-tokens and normal images and finally generate binary codes appropriately.

Jose M. Saavedra[2] discussed about the enhanced method for providing image descriptions in SBIR context named as Soft Histogram of Edge Local Orientations (S-HELO) which uses the advantages of HELO descriptor for efficient sketch descriptions and significant improvement in performance is noticed using a soft computation on local orientations considering the spatial information. HELO and HOG descriptors are used to compute orientation histograms which may be differed with respect to orientations. The HOG Mechanism applies Pixel orientation strategy and HELO used Cell- wise orientation. As Sketches implementation is sparse in nature HELO Strategy is applied. Soft-Histogram of Edge Local Orientations (SHELO) is developed from the base of HELO, in which it computes cell orientations in soft manner using tri-linear and bilinear interpolation according to a cell-based or block-based processing levels. This approach considers the spatial information by splitting image into blocks and obtains histogram for each block which follows the approach of Spatial Pyramid Matching.

Tseng, Yen-Liang Lin et.al [3] suggested the Distance Transform (DT) features utilization for the effective mapping of Query images and natural images. The features are then appended with extra compact hash bits. They analyzed that mobile sketch systems leads to less memory space utilization and effective search accuracy. The DT features are initially used to estimate the space information of images and then it uses very sparse random projection (VSRP) method to map DT features with hash bits. As it consumes less memory, the system can be controlled from the mobile platform, even if the system need to access images from server, mobile platform itself can control the operations of image retrieval which reduces operational cost of server systems and the features are compressed for efficient transmission.

Mathias Eitz *et al.* **[4]** presented a work on large scale sketch-based image retrieval, for searching a image over million images. They used descriptors to preprocess the sketches and images. Initially the image is analyzed for the gradient orientations, and the best matching images are segmented or clustered using the dominant color distributions for balancing the color-based decision while performing initial search. They adapted Sketch-based image descriptors namely Edge histogram Descriptor (EHD), Histogram of oriented gradients descriptor, angular radial partitioning , Tensor descriptor and Image Search and Clustering which are searched over huge dataset of images and clustered based on query type search techniques like K-means Tree search, Linear Search and Clustering- Lloyd clustering. The approach of retrieval process from the huge datasets using user-drawn sketches uses the methods like translation, scale, rotation and deformations. The retrieval is made fast by preprocessing the training images to extract the sequences of contour segments [11] which represent the shape information using variable length descriptors. Contour segments or chain similarities are achieved by a Dynamic Programming-based *approximate substring* matching algorithm which provides fast and partial matching of chains. The hierarchical K-Medoids based indexing is used for efficient and quick retrieval process from millions of images.

Steps:

1. In an offline process, the shape information of all the database images are represented in a similar-invariant way by extracting the long sequences of contour segments from each image and stored using variable length descriptors in a similarity preserving way.
2. For the effective matching of chains or contours between sketch and images, a DP-based approximate substring matching algorithm is used. Occlusions and non-object portions in the chains leads to challenges.

3. To facilitate the quick retrieval, the matching process is initiated using the chain descriptors along the tree which is referred as a hierarchical indexing tree structure and a geometric verification scheme is proposed for elimination of false matching which may be obtained due to partial shape similarity.

Yonggang Qi *et al.* [5], suggested a Convolutional Neural network (CNN) based on Siamese network. The main objective of the model is to extract the output feature vectors which are nearer for the input-sketch image pairs. They joined two CNN's linked with one loss function which aim at mapping input space to target space like Euclidean distance.

Ting Yao *et al.* **[6]** presented a Deep Semantic Preserving And Ranking Based Hashing(DSRH) which consists of three main components namely a deep CNN , binary mapping layer using hash streaming which divides the learned values into n no. of bags and encodes each bag into one hash bit and a classification stream.

Fei Huang *et al.* [7] presented deep cross-modal correlation learning novel scheme to provide an effective SBIR system from large scale annotated images which incorporates deep multimodal feature generation with deep cross modal correlation learning and provides optimized similarity search by mining all multimodal information in sketches and images as correlation distributions in sketches/sketch-like forms to implement pair-wise similarity with triplet loss.

Tu Bui and John Collomosse[8] implemented a scalable-system by extending the Gradient- Field framework(GF-HOG) with two forms of contributions for the effective SBIR systems. The initial contribution works towards color, shape retrieval and evaluate fusion approaches to enhance the accuracy and speed of sketch retrieval. Next contribution is an index representation for GF-HOG which is proposed to deliver the scalable search mechanism over millions of images with interactive query times. They used *Scalable color Sketch Retrieval* method for the effective approach, which adapts Classic GF-HOG for extracting the color in turn concentrates on perceptual uniformity and uses methods like feature sampling, Luminance Transfer function and early fusion techniques. They also adapted inverted indexing mechanism to design a codewords (BoVW) for the image representation it uses Hybrid inverted table and late fusion technique.

Xueming Qian *et al.* [9] proposed an effective SBIR approach using relevance feedback and re-ranking schemes. They included the semantic approach in query images and ranked the results retrieved. A Feedback technique is used to obtain relevant images for the given input query. They adopted edgel SBIR scheme to explain the approach. The edgel index structure is constructed in the offline process for each image using berkeley edge detector and SIFT descriptors[10][12] are recorded using the SIFT features extracted with the orientations and locations. In the next step, the contour similarity index is calculated for each image. In the foreground process the execution is performed in five stages, as a first step it includes SBIR which provides initial results, next stage involves grouping the results from stage-1 into relevant images from the top priority in the top N ranked results using Re-ranking by Visual Feature Verification and Contour Based Relevance Feedback. Third stage includes result verification using SIFT Descriptors matching. Fourth Stage include relevant images identification using contour based relevant feedback. The last stage involves performance enhancement using re-ranking of results.

3. Findings of the literature survey

The major findings of methods and techniques proposed for the purpose of image retrieval process using free form sketches are listed as follows:

- **S-HELO (Soft-Histogram of Edge Local Orientations):**

This method can be said as an improvement over HELO Strategy. S-HELO computation adopts a bilinear and tri-linear interpolation for identifying cell orientations with respect to block based and cell based processing levels and grabs the spatial information by dividing the image into blocks and a soft orientation histogram is constructed for each block and this division is referred as Spatial Pyramid matching.

- **Sketch-Based Image Descriptors[10]** which includes basic descriptors to generate the structure of the image namely angular radial partitioning which includes orientation of textures with respect to pixel direction, Edge histogram Descriptor(EHD) includes the histogram of the edges with respect to the texture.

- **Deep Convolutional Network Using Siamese CNN Architecture:**

The Deep convolutional network includes fusing two convolutional neural networks(CNN)'s consisting of a single loss function and the method includes **Siamese CNN architecture** which aim towards mapping the input space with target space so that there is a semantic distance in input space with respect to Euclidean distance in target space.

- **Gradient Field HoG (GF-HoG):**

It includes a bilateral filter and Binary dilation approach which provides a soft and noise free images, by including **color extraction** process into the GF-HOG pipeling using CIELab Color space supporting perceptual uniformity with the help of Feature sampling and Luminance transfer function and Early fusion techniques and applies an **Inverted Index technique** for retrieval process hashing a document into a set of unordered tokens or 'codewords'. The vector quantization (BoVW) method is used for representing images and a Hybrid inverted table and late fusion technique is also used.

4. Proposed Architecture

In our proposed method, we accept the free hand sketch as an input and generate the feature maps in multiple layers using Convolutional neural network to classify the features and finally compared with the set of training images with their distinct features. In our approach, the image generated using canvas can be in any size say 500 X 500 pixels image, but the Region of Interest (ROI) we are trying to search for be classified with their distinct features should be normalized to a specific sub-window size, say 250X250 or at least 3/4th size of the original image, so that classification time can be reduced instead of parsing the spatial orientation of the image where it can be simply white in intensity which may add unwanted pixels. Feature maps are generated at layer1 where these features are again fed to layer 2 so that some features may get classified, these feature maps are generated basing on the layers we are designing for the classification. The feature maps are generated at each layer consists the specific features like contour information, gradient information, edge histogram descriptors with texture representations. In the output layer, the identification of similar images with respect to the features classified are displayed. The implementation flow of the proposed system can be represented as follow:

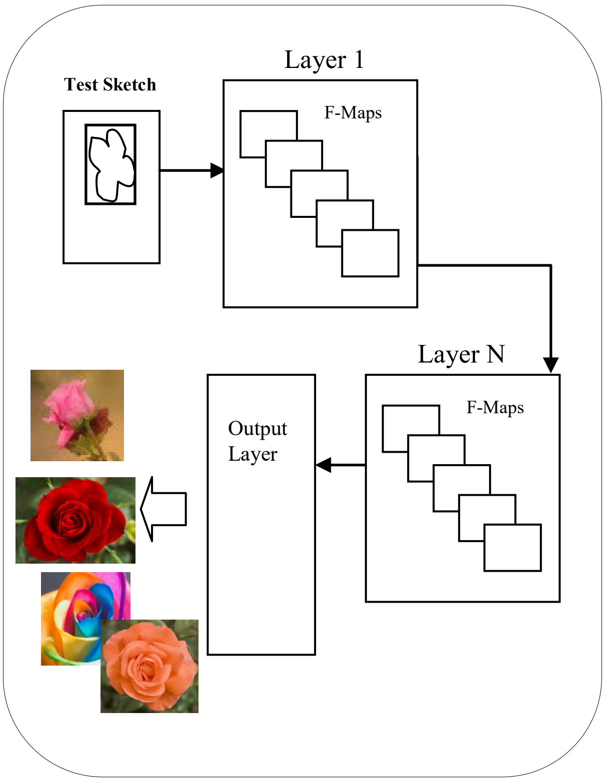

Figure 1. Proposed Architecture

4.1 Experimental Results

The general architecture with 20% dropout rate achieved a test accuracy of 65.6%. It exhibited better results than the classifier proposed by Eitz et al. which utilized SVM classifier trained on features extracted using SIFT. In model proposed by Yesilbek et al. achieved 71.30% cross-validation accuracy using an SVM classifier by training traditional image features. The proposed method is expected to provide the higher accuracy, obtaining 75.4% and 77.7% cross-validation accuracy with their convolutional Deep Sketch and DeepSketch 2 models as shown in table 1.

Table 1. Test accuracy comparison of our model versus other methods.

Model	Accuracy
SIFT+ SVM	0.560
Basic, 20% dropout	0.656
IDM + SVM	0.713
DeepSketch	0.754
DeepSketch2	0.777

While performing experimentation on "seagull" image using free hand sketch, the examination reveals that the examples are misclassified as a pigeon, standing bird, flying bird, and once as a syringe, duck, and canoe. This demonstrates the problems which arise from interclass overlap because for many people, there is no difference between drawing these varieties of birds.

5. Comparison Study

The comparative studies of several SBIR methods are discussed in table1.

Table 2. Comparison studies of several SBIR methods

Methods	Feature/ Methodology	Advantages	Disadvantage
PCA -	Linear transformation is adopted for Mapping a original image with its similar sketch image	Low dimensional subspace representation with simplicity of use	A low level preprocessing is required as it is sensitive to scale.
LLE	Nonlinear process of face sketch synthesis	Association among face photos and sketches can be simply estimated.	Need More training samples
Multi dictionary sparse representati on framework	Using LLE method and multidictionar y sparse representation model.	The quality of the initial image is enhanced for better Sketch-photo recognition rate	Multiple Sketches cannot be retrieved
GF-Hog	Uses Inverted Index technique for the retrieval	Estimates and retrieves the results through the means of BOVW for each image.	Need to estimate the color and shape of the image for better results.
Siamese CNN	Features of Edgemaps and input sketch are evaluated for the similarity	Semantic embedding is learned where retrieval can be easily performed	May not support 3D models of sketches.

6. Conclusion

This paper presents, the several methods and approaches used for the Sketch Based Image Retrieval to provide efficient results, the images are composed with the set of bag of words(BOW) or Bag of Visual Words(BOVW) for the better identification in the methods like S-HELO and GF-HoG. For the enhance performance along with efficiency, the method of Siamese CNN is carried out where the features of the input image are mapped using n layered neural network approach with distinct features at each layer which uniquely identifies the specific features of the image like gradients, Textures and edge histograms at each layer to identify the images uniquely. As each layer is meant to identify the features uniquely, the mapping of image with respect to training set is performed well and identified accurately, and the time complexity to perform this operation is minimal.

References

[1] Li Liu, Fumin Shen, Yuming Shen, Xianglong Liu, and Ling Shao, "Deep Sketch Hashing: Fast Free-hand Sketch-Based Image Retrieval", Computer Vision and Pattern Recognition,16 Mar 2017.

[2] Jose M. Saavedra , "Sketch Based Image Retrieval Using a soft computation of the Histogram of Edge Local Orientations(S-HELO)", IEEE International Conference on Image Processing (ICIP) ICIP 2014, IEEE society, 27-30 Oct. 2014.

[3] Tseng,Yen-Liang Lin, Yu-Hsiu Chen, Kai-Yu, "Sketch-based Image Retrieval on Mobile Devices Using Compact Hash Bits", Proceedings of the 20th ACM international conference on Multimedia, ACM Transactions, 2012.

[4] Mathias Eitz, Kristian Hildebrand, Tamy Boubekeur and Marc Alexa, "An evaluation of descriptors for large-scale image retrieval from sketched feature lines", Conference on Computers and Graphics, 19th Oct, 2010.

[5] Yonggang Qi,Yi-Zhe Song, Honggang Zhang, Jun Liu, " Sketch based image retrieval via Siamese Convolutional Neural Network" , IEEE International Conference on Image Processing (ICIP) ,ICIP 2016,IEEE society.

[6] Tingyao, fuchenlong, taomei, yongrui, "Deep Semantic-Preserving And Ranking-Based Hashing For Image Retrieval", International Joint Conference on Artificial Intelligence(IJCAI), 9-15 July 2016.

[7] Fei Huang, Cheng Jin, Yuejie Zhang, Tao Zhang, " Towards sketch-Based Image Retrieval with Deep Cross-modal correlation learning", Proceedings of the IEEE International Conference on Multimedia and Expo (ICME) 2017 10-14 Jul 2017

[8] Tu Bui and John Collomosse, "Scalable Sketch-based Image Retrieval using Color Gradient Features", IEEE International Conference on Computer Vision Workshop (ICCVW), 7-13 Dec. 2015.

[9] Xueming Qian, Xianglong Tan, Yuting Zhang, Richang Hong, and Meng Wang, Enhancing "Sketch-Based Image Retrieval by Re-Ranking and Relevance Feedback", IEEE Transactions on Image Processing, January 2016.

[10] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. "Sketch-based image retrieval: Benchmark and bag-of features descriptors". IEEE Trans. Visualization and Computer Graphics, Oct 2011.

[11] X. Qian, Y. Zhang, and X. Tan, "Sketch-based image retrieval using contour segments", IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), 21 Oct. 2015.

[12] S. Zhang, Q. Tian, K. Lu, Q. Huang, and W. Gao, "Edge-SIFT: Discriminative binary descriptor for scalable partial-duplicate mobile search", IEEE Trans. Image Process., Jul, 2013.

[13] S. Bell and K. Bala. "Learning visual similarity for product design with convolutional neural networks". In ACM Transactions on Graphics (TOG), 2015.

[14] Y. Cao, C. Wang, L. Zhang, and L. Zhang. "Edgel index for large-scale sketch-based image search". CVPR, 2011.

[15] Y. Cao, H. Wang, C. Wang, Z. Li, L. Zhang, and L. Zhang. "Mindfinder: interactive sketch-based image search on millions of images". In International Conference on Multimedia, 2010.