IJCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Special Issue-14, May 2019 E-ISSN: 2347-2693

"Click-n-Purchase: A Shopping guide with Image Retrieval based on Mobile Visual Search in Fashion Domain": A Survey

Nikhil S. Tengli[1*], Suvarna Nandyal[2]

[1,2] Department of Computer Science, PDA College of Engineering, Kalaburagi, Karnataka

[*]*Corresponding Author: nikhil.tengli039@gmail.com, Tel.: +91-9164499131*

DOI: https://doi.org/10.26438/ijcse/v7si14.8892 | Available online at: www.ijcseonline.org

***Abstract*—** In the recent years, the use of e-commerce based applications via Internet has grown rapidly, thus increasing the volume of data in the web. Therefore it necessary to have faster retrieval of required data from the web. This paper provides a comprehensive review of various image retrieval techniques with their problems. The survey presents various techniques used so far for the Image Retrieval from the Web based applications, in order to make more efficient way of retrieving the information by using image retrieval techniques. The survey describes which techniques are used for image retrieval and the problem faced during the retrieval process. Finally, based on the use of existing techniques and the demand from the real-time applications a shopping guide will be presented with enhanced features of image retrieval techniques named as Click-n-Purchase, where the input for this application is taken from the mobiles and the visual search of the related images can be extracted from web based fashion domain based applications, so that user can be able to search their favourite items in less amount of time.

***Keywords*—** Image Retrieval Techniques, Mobile Visual Search, Fashion domain, Click-n-Purchase.

I. INTRODUCTION

Recently, there has been an increase in the number of multimedia applications including text, images, video and audio data in every field with equal number of processing requirements. The coming up of huge sized image datasets calls for optimal techniques for storing, searching, indexing and retrieving desired data at a faster speed. Target object recognition is one of the most popular and widely researched topics in computer vision field that has gained a lot of momentum. The goal of this topic is to recognize the objects belonging to the target object class based on its features. Shape of an object is a visually distinguishable feature that can be easily utilized for recognition task but it has its own challenges like pose variations, occlusions in the scene, lighting conditions, resolution etc [1]. A number of methods are developed for achieving recognition of target based on shape cues. The basic means of object recognition involves matching the extracted features of objects in the image with a sample target object features.

This feature matching step is carried on all the objects in every test image to find potential target object present. Search technique is included in object recognition methodology where the feature set extracted will be put to comparison with the already extracted features of the target object for easy and accurate recognition. To ensure good recognition rate, the features extracted must be distinguishable well and relevant for the target object to be recognized. A variety of techniques are available for the purpose of target recognition. Some of them are briefly described here. The most important points to be considered in object recognition tasks are the extraction of features that are independent of rotation and translation [2]. The shape invariant feature based recognition of target objects involves following three steps-

- Data Pre-processing
- Feature extraction
- Classification

In the first step, the input image is filtered using noise removal methods and processed to improve its quality for recognition purpose. Image processing aims to enhance the image for obtaining good measures of its features [19]. In the next step, feature extraction from pre-processed image is performed where the number and type of features extracted depends on the end goal of the research. The extracted features of all input images in the dataset are stored in the database to be matched further for recognition of similar target. Thus, feature extraction plays an important role in the entire object recognition task that decides the performance of the system developed.

II. FEATURE EXTRACTION TECHNIQUES

Many image retrieval systems rely on features like shape, intensity, and texture etc. [2]. The shape of an object is a physical descriptor. Shape representation can be done in the form of interior region, border, moment, edges etc. These are

used to compare the target as well as input image objects by noticing the similarity in making measurements. Texture is also a structural descriptor which is a characteristic of object surfaces such as glass, grains, sand, and textile. The name texture is derived from the primitive texture elements organized in a pattern which is called Texel's. A Texel may itself have many pixels, with periodic or random arrangement depending on the object. Naturally occurring objects have random texture while synthetic objects have predefined ordered textures.

Texture may be of any category such as coarse, fine, regular, irregular, or linear. In image processing, texture has two broad classes- statistical and structural.

The Statistical category includes the textures usually random in nature. The structural textures are those which are certain and that recur based on few rules be it deterministic or random. Another method is a combination of the above two patterns which we call as mosaic models.

The mosaic models depict randomized geometrical processes. Visual Seek considers object layout as an image feature [2]. The visually distinguishing features like object shape, colour or edges are very good descriptors for image retrieval but they are vulnerable in case the query image is a grayscale image that too, a sketch of the object [2].
Since almost every object recognition system concentrates on obtaining rotation and translation invariant features, there are methods that classify these into the following: 1. Image alignment 2. Invariant features.

III. RELATED WORK

This section discusses the related work in various papers with respect to image retrieval in order to analyze the previous approaches and come up with a better system for designing a shopping guide with image retrieval using a mobile interface.

A. Nodari et.al [3] The authors have proposed a mobile app that works in sync with a Content-Based Image Retrieval system for online shopping of fashion products. The application is able to retrieve most matched products of a textile image given as an input. The proposed method works by manual selection of the product name by the end user framed by the camera image which is then sent to the server. Next, a search for image retrieval is initiated that automatically recognizes and maps the object similar to the requested one reducing the effort of a prolonged manual search. To measure the performance of the image retrieval system designed, the authors have validated against three datasets- First set involves products of clothing category from various online shopping websites and the other two datasets are the images as well as video frames of clothing obtaining from Internet users. The system is found to be feasible for real time application.

The application works on client-server architecture. The client's role is to present the end user with the graphical user interface asking query image dynamically. The client then sends the final query formed to the server. Initially, an XML file is downloaded that includes the different product names, their underlying category and the focus masks.

Jan Cychnerski, Adam Brzeski et.al [4] In this work, the authors have ddeveloped a computer vision system for efficient detection and classification of clothing images for e-shopping images. A combined architecture involving convolution neural networks that cover Residual networks, Squeeze Net and Single Shot MultiBox Detector (SSD) is considered. The training experiments for detecting the clothing images was done on Deep Fashion dataset containing cloth location annotated by rectangular box. The system was evaluated against various cloth images available online on e-commerce sites. Ground truth for the image dataset were obtained using the online shopping catalog containing metadata for five properties of color, style, sleeve-type, neck design and hemline. Automatic annotation collection showed up mean rate of 83% rate accuracy. In the analysis part, the authors have sought the effects of a number of improvements attempted like data augmentation under varying backgrounds, network size increase and ensemble usage on the classification rate. Also, the classification rate versus processing ease is computed in this work. The most optimal network organizations for accurately classifying the input clothing images are also presented.

The authors have carried out experiments by incrementing the network model size gradually to test and record the best possible classification results but the computational complexity is compromised. Apart from a single network, ensembles of 5 networks are considered. The popular pre-trained GoogLeNet classifier is used for fashion textile classification by eliminating the last fully-connected layer and pre-training on the Image Net dataset. Image Net dataset is a huge sized image collection containing around 1.2 million images with 1000 categories. Fine tuning the last fully-connected layer with the fashion image dataset collected is finally done. The system classifies a training image into one of the classes from a possible 24 fine grained classes.

Dibin Zhou, Baokun Hu et.al [5] In this work, the authors have carried out a study on mobile phone based image retrieval system for the design of shopping guide for online shopping website. This guide helps in various items to be retrieved from the online repository of products using mobile interface. The design is aimed at providing easy to use interactive methods and innovative layout for image retrieval

so that the drawbacks of existing image retrieval systems are resolved. The developed system has a decent practical value and simplicity.

Liu Shuguang et.al [6] the authors have applied the wavelet transform and frequency gain for developing a good clothing classification system based on the texture features in this work. According to them, the wavelet packet algorithm works very well in case of classification of clothing types efficiently for the following facts- The middle and low frequency regain respectively contain the core energy of the texture image and usual image. So, with an increase in the image resolution, the wavelet transform is driven towards low frequency range, while wavelet packet is focused on any available frequency range. The authors have exploited the wavelet packet with Back Propagation neural network to classify the fabric types based on the texture. The developed system showed 98% classification accuracy.

Tom Yeh1 et.al [7] Here, the authors have developed a mobile interface based image search system producing the relevant web pages on the Internet containing matching objects similar to the query image. Image matching produces very accurate results when matching whole image or the scene is carried out like landmarks etc. or also in cases where the object to be matched has a distinguishable outline. The authors have devised an interactive interface to obtain a segmented object boundary with shape matching algorithm to spot the desired objects of interest on Web pages.

Yixin Chen et.al [8] In this paper, a novel Multiple-Instance Learning (MIL) based technique for image classification is described. The authors have extended the usual MIL method by incorporating DD-SVM, where a bag is labelled based on a few instances satisfying stated rules. Diverse Density (DD) function is used for learning instance prototype categories in a DD-SVM. Since every instance prototype is representative of the class of instances to appear in bags with a fixed label than in the unlabeled bags, a nonlinear mapping to map each bag to a point in bag feature space is done. Support vector machines are used for training in the bag feature space.

P.F. Li, J. Wang et.al [9] In this work, the authors concentrate on the task of automatic woven fabric classification using a novel feature extraction technique. The local binary pattern and the gray level co-occurrence matrix features are initially extracted for the fabric images in the data collection. To reduce high dimensional feature information, principal component analysis is performed. Support vector machine classifier is employed for recognition of the fabric image. The method is tested against three different types of woven fabric categories namely plain, twill and satin weave fabrics.

The below table 1 describes comprehensive analysis of research work done so far in image retrieval techniques. Since the accuracy of the methods used in the existing work are more comparative and effective in detecting the object from the images, these techniques can be used for our research work in improving the accuracy for the detection of the images.

Table 1: Comprehensive Analysis of Image retrieval Techniques used in Existing Research Works

Existing work details	Methodology used in Existing work	Advantages	Limitations	Percentage of Accuracy
A mobile visual search application for content based image retrieval in the fashion domain By **A. Nodari et.al [3]**	In this paper, the authors describe a mobile app that works in sync with a Content-Based Image Retrieval system for online shopping of fashion products. The proposed method works by manual selection of the product name by the end user framed by the camera image which is then sent to the server.	Content based image retrieval system. Dynamic query image processing Automated search technique	No image based retrieval technique No visual words fusion for Image CLEF and image collections using deep learning	87.434
Clothes detection and classification using convolutional neural networks By **Jan Cychnerski, Adam Brzeski et.al.[4]**	In this paper, the authors have developed a computer vision system for efficient detection and classification of clothing images for e-shopping images. A combined architecture involving convolution neural networks that cover Residual networks, Squeeze Net and Single Shot MultiBox Detector (SSD) is considered.	Proposed Convolution neural networks that cover Residual network Data Augmentation Ease of processing speed.	The retrieval of the query image attributes should still be improved. Network size increased.	76
Design of Shopping Guide System with Image	In this paper, the authors have carried out a study on mobile phone based image retrieval system for the	The developed system has a decent practical value and simplicity.	The proposed CBIR is not extended with a more variety of image features.	Performance is measured in terms of time,

Retrieval Based on Mobile Platform By **Zhou, Dibin et. al.[5]**	design of shopping guide for online shopping website. This shopping guide helps in various items to be retrieved from the online repository of products using mobile interface.	Providing easy to use interactive methods and innovative layout for image retrieval.	The weight assignment process is not refined based on detailed analysis of feature vectors.	which retrieves images from 0.01 to 0.3 seconds.
Fabric Texture Classification Based on Wavelet Packet By **Liu Shuguang et.al [6]**	The authors have presented the wavelet transform and frequency gain for developing a good clothing classification system based on the texture features in this work.	Wavelet packet algorithm works very well in case of classification of clothing. Back Propagation neural.	Low frequency transformation.	99.1 for D17D55
A Picture is Worth a Thousand Keywords: Image-Based Object Search on a Mobile Platform By **Tom Yeh1 et. al.[7]**	The authors have developed a mobile interface based image search system producing the relevant web pages on the Internet containing matching objects similar to the query image. Image matching produces very accurate results when matching whole image or the scene is carried out like landmarks etc. or also in cases where the object to be matched has a distinguishable outline.	Search system producing the relevant web pages. Accurate object matching.	Problem with similar object detection.	Accuracy is measured based on the shape of object detection
MILES: Multiple-Instance Learning By **Yixin Chen et.al [8]**	This paper presents a novel Multiple-Instance Learning (MIL) based technique for image classification is described. The authors have extended the usual MIL method by incorporating DD-SVM, where a bag is labelled based on a few instances satisfying stated rules.	Diverse Density (DD) function is used for learning instance prototype categories in a DD-SVM. Multiple-Instance Learning (MIL) based technique	Nonlinear mapping is a big concern. Proposed MIL failed in classifying overall image retrieval.	97.17

IV. CONCLUSION

This paper presents a comprehensive review analysis of various existing image retrieval techniques with their problems. This survey provides the analysis of how the various image retrieval techniques can be applied over fashion domain and the problem faced during the existing retrieval process. Once the existing problems are analyzed, In future by using this work novel solution can be provided by implementing new image retrieval guide called "Click-n-Purchase: A shopping guide" , which provides an improved image retrieval based searching of the various interested items from the fashion domain. The accuracy of the new application will be measured by comparing with the existing techniques used and presented in this work.

ACKNOWLEDGMENT

The author would like to acknowledge the esteemed guidance given by Dr. Suvarna Nandyal, Professor and Head, PDA College of Engineering, Kalaburagi, Karnataka to complete this work.

REFERENCES

[1] Mehmood, Zahid and Abbas, Fakhar and Mahmood, Toqeer and Javid, Muhammad Arshad and Rehman, Amjad and Nawaz, Tabassam, Content-Based Image Retrieval Based on Visual Words Fusion Versus Features Fusion of Local and Global Features, Arabian Journal for Science and Engineering, 2018, pp. 1-20.

[2] Katrien Laenen, Susana Zoghbi, and Marie-Francine Moens, Web Search of Fashion Items with Multimodal Querying, Eleventh ACM International Conference on Web Search and Data Mining (WSDM '18), 2018, pp. 342-350.

[3] Angelo Nodari, Matteo Ghiringhelli, Alessandro Zamberletti, Marco Vanetti, Simone Albertini, Ignazio Gallo, "A mobile visual search application for content based image retrieval in the fashion domain", 10th International Workshop on Content-Based Multimedia Indexing, 2012.

[4] J. Cychnerski, A. Brzeski, A. Boguszewski, M. Marmolowski and M. Trojanowicz, "Clothes detection and classification using convolutional neural networks," *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, Limassol, 2017, pp. 1-8. doi: 10.1109/ETFA.2017.8247638

[5] Zhou, Dibin & Hu, Baokun & Wang, Qihui & Hu, Bin & Jia, Leiping & Wu, Yingfei & Xie, Lijun. "Design of Shopping Guide System with Image Retrieval Based on Mobile Platform". 10.2991/3ca-13.2013.37, 2013.

[6] Liu Shuguang, Qu Pingge "Fabric Texture Classification Based on Wavelet Packet", The Eighth International Conference on Electronic Measurement and Instruments,2017.

[7] Tom Yeh1, Kristen Grauman1, Konrad Tollmar2, Trevor Darrell, "A Picture is Worth a Thousand Keywords: Image-Based Object Search on a Mobile Platform", CHI 2005, April 2-7, 2005. Portland, Oregon, USA.
[8] Yixin Chen, Member, IEEE, Jinbo Bi, Member, IEEE, and James Z. Wang, Senior Member, IEEE, "MILES: Multiple-Instance Learning", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 12, DECEMBER 2006
[9] P. F. Li, J. Wang, H. H. Zhang and J. F. Jing, "Automatic woven fabric classification based on support vector machine," International Conference on Automatic Control and Artificial Intelligence (ACAI 2012), Xiamen, 2012, pp. 581-584.
[10] Min, Weiqing and Jiang, Shuqiang and Wang, Shuhui and Xu, Ruihan and Cao, Yushan and Herranz, Luis and He, Zhiqiang,"A survey on context-aware mobile visual recognition, Multimedia Systems, 23(6), 2017, pp. 647-665.
[11] Weiqing Min, Shuqiang Jiang, Shuhui Wang, Ruihan Xu, Yushan Cao, Luis Herranz, and Zhiqiang He, "A survey on context-aware mobile visual recognition". Multimedia Systems, 2017, pp. 647-665.
[12] Mitul Kumar Ahirwal, Anil Kumar, and Girish Kumar Singh, "An Approach to Design Self Assisted CBIR System", International Conference on Graphics and Signal Processing (ICGSP'17),pp. 21-25.
[13] Xin Ji, Wei Wang, Meihui Zhang, and Yang Yang, "Cross-Domain Image Retrieval with Attention Modelling", ACM on Multimedia Conference(MM'17),2017,pp.1654-1662.
[14] C. Huang, S. Zhang, X. Lin, X. Liu and R. Ji, "Deep-based fisher vector for mobile visual search", IEEE International Conference on Image Processing (ICIP), 2017, pp. 3430-3434.
[15] Y. H. Kuo and W. H. Hsu, "Dehashing: Server-Side Context-Aware Feature Reconstruction for Mobile Visual Search", IEEE Transactions on Circuits and Systems for Video Technology, 27(1), 2017, pp. 139-148.
[16] A. Rahman, E. Winarko and M. E. Wibowo, "Mobile content based image retrieval architectures," 4th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2017, pp.1-4.
[17] C. Corbière, H. Ben-Younes, A. Ramé and C. Ollion, "Leveraging Weakly Annotated Data for Fashion Image Retrieval and Label Prediction," IEEE International Conference on Computer Vision Workshops (ICCVW), 2017, pp. 2268-2274.
[18] J. Sivic and A. Zisserman, "Video Google: A text retrieval approach to object matching in videos," in IEEE Conference on Computer Vision and Pattern Recognition, 2003, pp. 1470–1477.
[19] J. Philbin, O. Chum, M. Isard, J. Sivic and A. Zisserman, Object retrieval with large vocabularies and fast spatial matching, 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, 2007, pp. 1-8.
[20] W. Zhou, Y. Lu, H. Li, Y. Song, and Q. Tian, "Spatial coding for large scale partial-duplicate web image search," in ACM International Conference on Multimedia, 2010, pp. 511–520.
[21] O. Chum, J. Philbin, J. Sivic, M. Isard, and A. Zisserman, "Total recall: Automatic query expansion with a generative feature model for object retrieval," in International Conference on Computer Vision, 2007, pp. 1–8.
[22] D. Nister and H. Stewenius, "Scalable recognition with a vocabulary tree," in IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2006, pp. 2161–2168.
[23] Z. Wu, Q. Ke, M. Isard, and J. Sun, "Bundling features for large scale partial-duplicate web image search," in IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 25–32.
[24] X. Wang, M. Yang, T. Cour, S. Zhu, K. Yu, and T. X. Han, "Contextual weighting for vocabulary tree based image retrieval," in International Conference on Computer Vision, 2011, pp. 209–216.
[25] L. Zheng, S. Wang, and Q. Tian, "Coupled binary embedding for large-scale image retrieval," IEEE Transactions on Image Processing (TIP), vol. 23, no. 8, pp. 3368–3380, 2014.
[26] Y. Cao, C. Wang, L. Zhang, and L. Zhang, "Edgel index for largescale sketch-based image search," in IEEE Conference on C Vision and Pattern Recognition (CVPR), 2011, pp. 761–768.
[27] J.-P. Heo, Y. Lee, J. He, S.-F. Chang, and S.-E. Yoon, "Spherical hashing," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2012, pp. 2957–2964.
[28] J. Tang, Z. Li, M. Wang, and R. Zhao, "Neighborhood discriminant hashing for large-scale image retrieval," IEEE Transactions on Image Processing (TPI), vol. 24, no. 9, pp. 2827–2840, 2015.
[29] L. Wu, K. Zhao, H. Lu, Z. Wei, and B. Lu, "Distance preserving marginal hashing for image retrieval," in IEEE International Conference on Multimedia and Expo (ICME), 2015, pp. 1–6.
[30] K. Jiang, Q. Que, and B. Kulis, "Revisiting kernelized localitysensitive hashing for improved large-scale image retrieval," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4933–4941.

Authors Profile

Mr. Nikhil. S. Tengli received his B. E. Degree in Computer Science from VTU, Belagavi, M.Tech. degree in Computer Science and Engineering from VTU, Belagavi. He is pursuing Ph.D. under VTU, Belagavi in the area of Image Processing and Machine Learning. He has experience of 5 years in teaching in Computer Science and Engineering and 1 year in Industry. His research interests are: Image processing, Cloud computing, Machine learning and Big data. He has published 8 papers in UGC indexed journals.

Dr. Suvarna Nandyal has completed BE and M.Tech from PDA College of Engg. Gulbarga and completed Ph.D. from JNT University, Hyderabad. She has taken charge of HOD computer science and engg. department from 8th may 2015. Her area of research is Digital Image Processing. She has published more than 50 research papers in International and National Journals. She has attended many national and international conferences. She has organized workshops on current topics like Digital Image Processing, Cloud computing and many more. She has presented a paper at International conference on Digital Image Processing at Singapore. She has worked as P.G. Coordinator. She has organized workshops on Digital Image Processing and Cloud Computing.

