AN EFFICIENT PERSONAL CLOTHING RETRIEVAL BASED ON EXTREME LEARNING MACHINE

J.Jessy Goldina[1], M.Sharon Nisha[2]
[1]*PG scholar Department of CSE, Francis Xavier Engineering College, Tamilnadu,India.*
[2]*AP, Department of CSE, Francis Xavier Engineering College, Tamilnadu,India.*

***Abstract*- The main objective of this project is clothing retrieval. In clothing retrieval, first apply the face detection. After face detection apply segmentation for skin, clothing and non clothing. In segmentation this project using the graph cut algorithm. Extract the dominant color patches and build a visual codebook with the high-dimensional descriptors. Next extract the SIFT, GIST and PHOG features. Finally apply the ELM classification. ELM is one of the best classification technique.**

I. INTRODUCTION

Clothing retrieval is used an wide range applications from automatic tagging to e-commerce of clothing. It is an very active topic in computer vision and multimedia research. If a person was in different photos, then he/she can be identified by using the face attributes including the clothing features . To do this, the re-ranking concept arises to re-rank the retrieved images based on the text around the image, metadata of image and visual feature of image. Learning Attributes is used to describe how to learn from the retrieved features, while comparing the other photos. Nowadays the online clothing shopping is becoming an increasingly popular shopping model. In many online shopping websites such as Amazon.com, eBay.com, and shopstyle. com, customers can conveniently find their favourite clothing by typing some keyword. This shopping model can be integrated into various platforms, e.g. in mobile applications as shown in Figure One can take a photo of any fashionably dressed lady with mobile device, and our system can parse the photo and search for the clothing with similar styles from online shopping websites. In social network platform, one may occasionally browse a photo in a friend’s album and is interested in the clothing. With one click over the photo, similar clothing from online shops can be returned. The color matching is also called as color harmony which is achieved when colors create a pleasing effect as Red, yellow and blue always harmonize. These colors are bold and eye-catching and they never really go out of style. Whether you're putting together a palette for your wardrobe a painting or your dining room you can depend on primary colors to lend your project a cheerful and bright appearance. It provides a context-sensitive environment adapting to the user’s interactions. This helps to relieve the burden of users. During the time of interaction to user and the system the operations of the user actually reflect the intention of the user. We can guess this intention and re-rank the photos or clusters to be labeled. For example, when the user clicks one photo/cluster, it indicates his/her attention on this one.We re-rank the photos/clusters and arrange the similar ones close to the clicked photo/cluster. The user will find it easier to label them together. This feature boosts labeling productivity of users. Given training data stating how object/scene categories relate according to different attributes learn a ranking function per attribute. The learned ranking functions predict the relative strength of each property in novel images. Then construct a generative model over the joint space of attribute ranking outputs and propose a novel form of zero-shot learning in which the supervisor relates the unseen object category to previously seen objects via attributes .

II. RELATED WORK

Ming Yang [1]This paper proposes clothing recognition in real time video surveillance. Clothing recognition involves several critical sub-tasks including localization of human figures, clothing segmentation. In this paper developed an efficient color segmentation method and 3 types of texture features tailored for real-time recognition of clothing categories and alignment, and extraction of clothing representation. Finally compared this 3 types of recognition. Y. Song [2]This paper proposes to automatically cluster pictures according to person's identity by exploiting as much context information. A clothes recognition algorithm is uses color and texture information and is effective for different types of clothes (smooth or highly textured). The results of clothes recognition are integrated with face recognition and to provide similarity measurements for clustering. The Picture-taken-time is used when combining faces and clothes and the cases of faces or clothes missing are handled in a principle way. Xinmei Tian[3].This paper proposes the video search reranking using the Bayesian. Two new methods are proposed in this paper to measure the ranking distance based on the disagreement in terms of pair-wise orders. First one is hinge distance penalizes the pairs with reversed order according to the degree of the reverse then next one is preference strength distance further considers the preference degree. This two reranking methods are developed which are solved using quadratic programming and matrix computation respectively. W. H. Hsu [4]In this paper propose a novel and generic video/image reranking algorithm, IB reranking, which reorders results from text-only searches by discovering the salient visual patterns of relevant and irrelevant shots from the approximate relevance provided by text results. The IB reranking method is based on a rigorous Information Bottleneck (IB) principle to finds the optimal clustering of images that preserves the maximal mutual information between the search relevance and the high-dimensional low-level visual features of the images in the text search results.

In previous, the search is mainly based on the geometric patterns. Most existing image search engines, like Microsoft Live Image Search and Google Image Search, mainly rely on associated textual information for image retrieval such as surrounding text and human-submitted annotations. The Textual information have limited capability and it produces inaccurate result. The Clothing Retrieval techniques which uses the Shape Context (SC) solves garment matching, but it is very difficult. The least square posterior classifiers is used for classification which is time consume process. To overcome the above problems, we propose a novel approach to retrieve the clothing on photo collections. It uses the Fuzzy c-means clustering for clustering the same pixels in an efficient way. For classification, the Extreme Learning Machine (ELM) is used, which classifies the clothes in accurate manner. A re-ranking approach is then proposed to improve search quality by exploiting clothing attributes. This approach is robust to large variations of images taken in unconstrained environment. The rest of this paper discussed about the proposed method.

III. METHODOLOGY

3.1 System Architecture:

In this chapter discuss about the proposed method in detail. The below diagram shows the overall process of the proposed system.

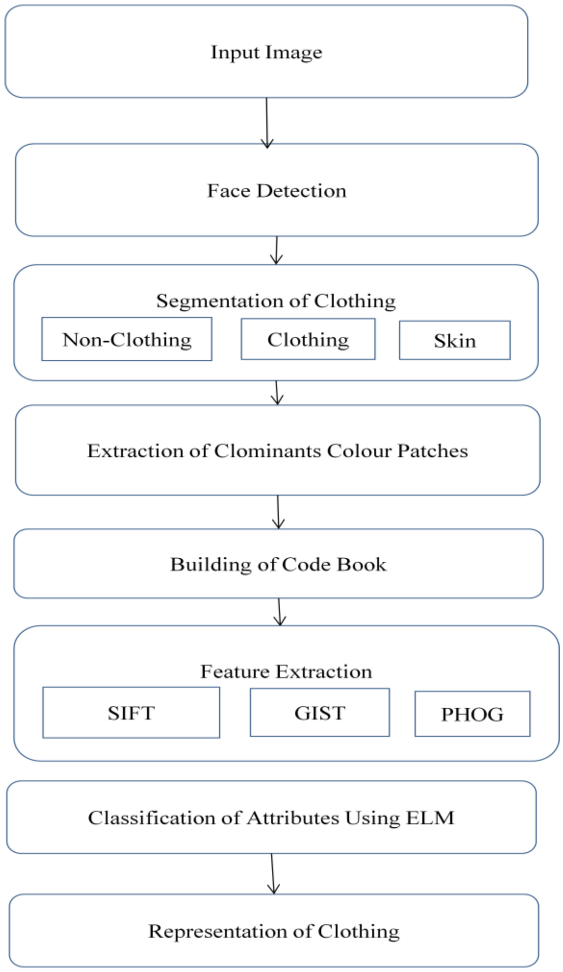

Fig:1 Overall diagram for proposed system

3.2 Modules:

3.2.1 Face Detection:

The input image is scanned across location and scale. At each location an independent decision is made regarding the presence of a face. Once faces are detected and located, To use the relative position and scale of each detected face to predict a bounding box below the head which is expected to contain the clothing of the upper-body.

3.2.2 Segmentation of Clothing:

The bounding box can be used as the initial estimation of the clothing location. Often, it also contains hair, skin, occlusion from another person or objects and cluttered background. Grabcut method is used to segment the clothing in each photo which improves the quality of the extracted clothing by removing the influences from the cluttered background and occlusions. Grabcut seeks the binary labeling over the superpixels that minimizes the energy using graph cuts. The energy function is defined with two terms as follows,

$$E(l) = \sum_{s \in \wp} D_s(l_s) + \sum_{s,t \in \aleph} V_{s,t}(l_s, l_t)$$

3.2.3 Extraction of dominants colour patches:

There are 5 steps to extract the dominant colour patches.

1.Generate the 3D histogram in the RGB color space with 16 bins on each axis for each clothing. The colors in the same bin are assumed to be similar, and the centroid of each bin is selected as its quantized color.

2. Select the quantized colors of bins with total number of pixels and pre-defined Threshold

3. Divide the clothing image into 7 x7 non-overlapping patches; go through all patches, and select the patch as one candidate of dominant color patches.

4. Sort the candidate patches for each DC in the order of the number of pixels.

5. Perform illumination normalization on each DC patch.

3.2.4 Code Book Generation:

Step1: Input the given image of size m × m pixels.
Step2: Divide it into blocks of size 4×4 pixels.
Step3: Generate N training vectors
Step4: Fix the codebook size to M.
Step5: Compute p=N/M
Step6: Select every pth training vector as code vector to generate the codebook.

3.2.5 Feature Extraction:

In pattern recognition and in image processing, feature extraction is a special form of dimensional reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant (e.g. the same measurement in both feet and meters) then the input data will be transformed into a reduced representation set of features . Transforming the input data into the set of features is called feature extraction. If the features are extracted are carefully chosen it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.
The following features are extracted in this project. The algorithms are explained in below.

SIFT Features:
Scale-invariant feature transform (or SIFT) is an algorithm in computer vision to detect and describe local features in images. Applications are include object recognition, robotic mapping, navigation, image stitching, 3D modeling, gesture recognition, video tracking, individual identification of wildlife and match moving. SIFT keypoints of objects are first extracted from a set of reference images and stored in a database.

Algorithm:
1. Scale Space Extrema Detection

- This stage of the filtering attempts to identify those locations and scales that are identifiable from different views of the same object.
- To locating scale-space extrema $D(x, y, \sigma)$ by computing the difference between two

images. one with scale k times the other. D(x, y, σ) is then given by:

$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma)$

- To detect the local maxima and minima of D(x, y, σ) each point is compared with its 8 neighbours at the same scale, and its 9 neighbours up and down one scale.
- If this value is the minimum or maximum of all these points then this point is an extrema.

2. Key point Localization

- To improve Key points and throw out bad once

3. Orientation Assignment

- To remove the effects of rotation and scales

4. Create Descriptor

- Using the histograms for orientation

PHOG Descriptors:

The PHOG (Pyramid of Histograms of Orientation Gradients) descriptor consists of a histogram of orientation gradients over each image subregion at each resolution level.

Local shape

Local shape can be described by a histogram of edge orientations (quantized into K bins) within an image subregion. The edge orientations are quantized into K bins, each of which represents the number of edges which have a certain angular range orientations. The contribution of each edge is weighted by its magnitude.

Spatial layout

The PHOG image descriptor is a concatenation of all the HOG vectors, each of which is computed for each grid cell at each pyramid. Consequently, level 0 is represented by K-bin histogram, level 1 is represented by a 4K-bin histogram, etc, and the final PHOG descriptor of the entire image is a vector with dimensionality.For example, for levels up to L= 1 and K= 30 bins it will be a 150-vector. To prevent overfitting, we limit the number of levels to L= 3 in our implementation. More to the point, the PHOG is normalized to sum to unity to ensures that images with more edges are not weighted more strongly than others.

GIST Descriptors:

The GIST descriptor has proved to be powerful in scene categorization and retrieval. Each clothing image is convolved with Gabor filters at 8 scales and 8 orientations. The filter responses are averaged within each of 4x4 divisions of the image.

3.2.6 Classification using ELM:

A new learning algorithm called **e**xtreme **l**earning**m**achine (ELM) for **s**ingle-hidden **l**ayer **f**eedforward neural **n**etworks (SLFNs) which randomly chooses hidden nodes and analytically determines the output weights of SLFNs. This algorithm tends to provide good generalization performance at extremely fast learning speed.

Algorithm:

1. Given a training set, activation function , and the number of hidden nodes.
2. Assign randomly input weight vectors or centers and hidden node bias or impact factor.
3. Calculate the hidden layer output matrix .
4. Calculate the output weight.

IV. RESULTS

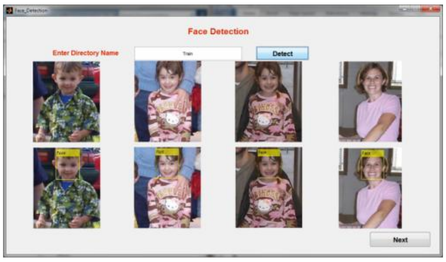

Fig: 4.1 Face Detection

Fig: 4.2 Clothing Estimation

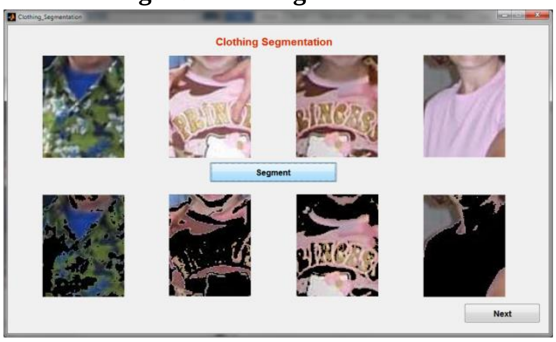

Fig: 4.3 Clothing Segmentation

Fig: 4.4 Skin Estimation

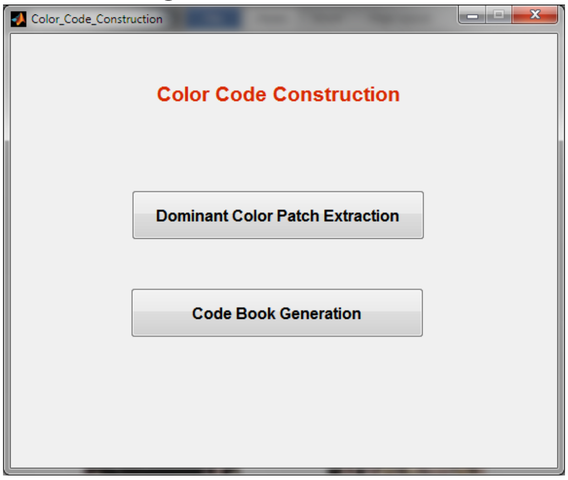

Fig: 4.5 Color code Construction

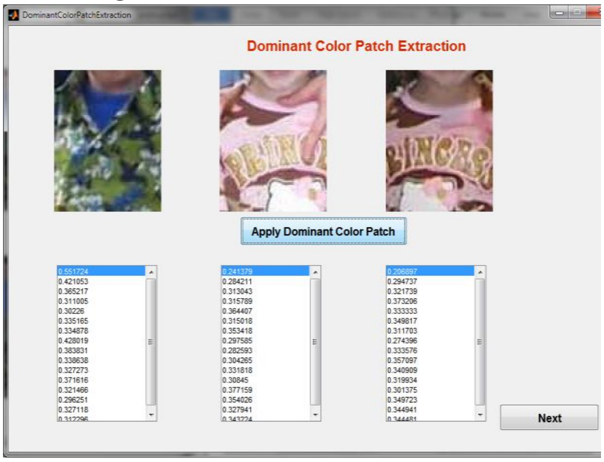

Fig: 4.6 Dominent Color Patches

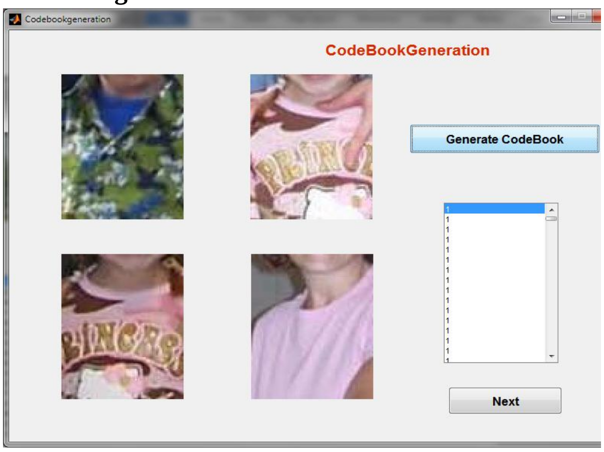

Fig: 4.7 Code book Generation

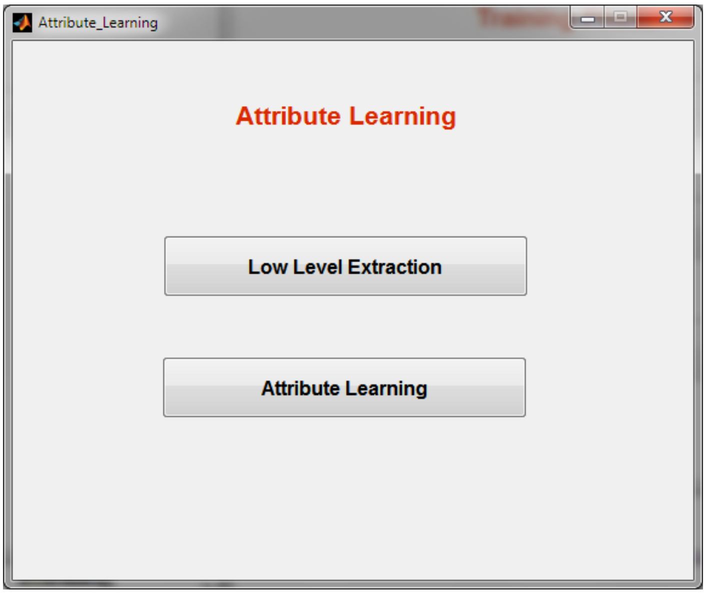

Fig: 4.8 Attribute Learning

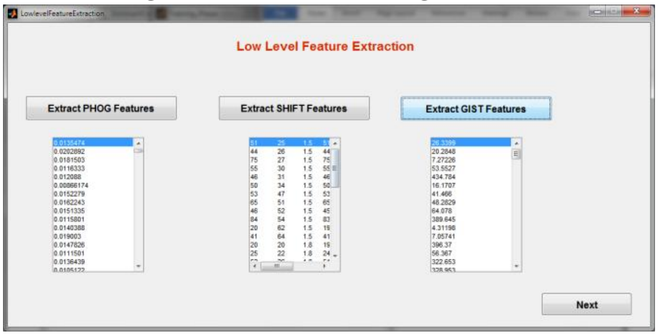

Fig: 4.9 Low level features

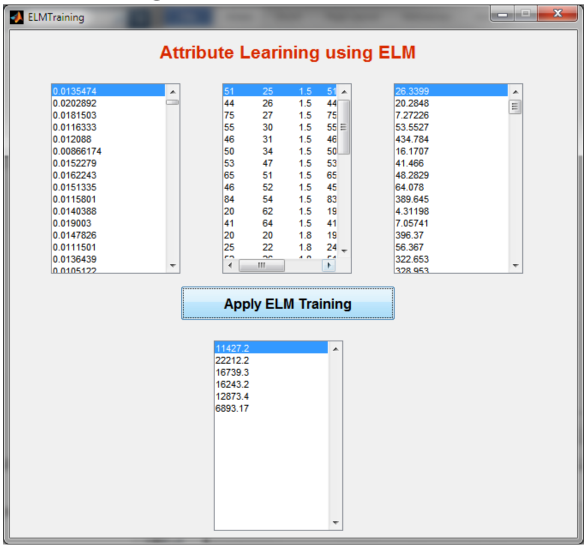

Fig: 4.10 ELM Training

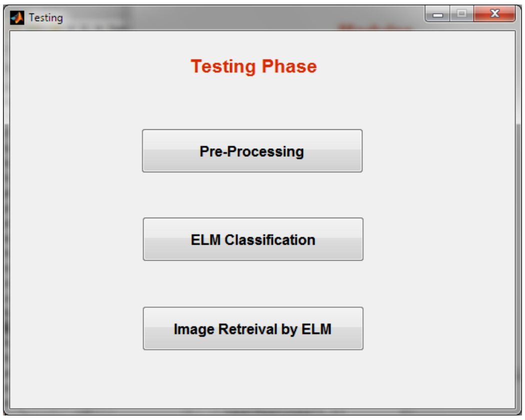

Fig: 4.11 Testing

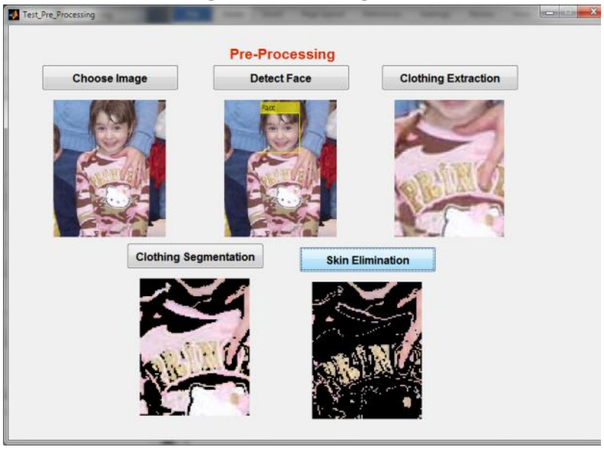

Fig: 4.12 Pre-Processing

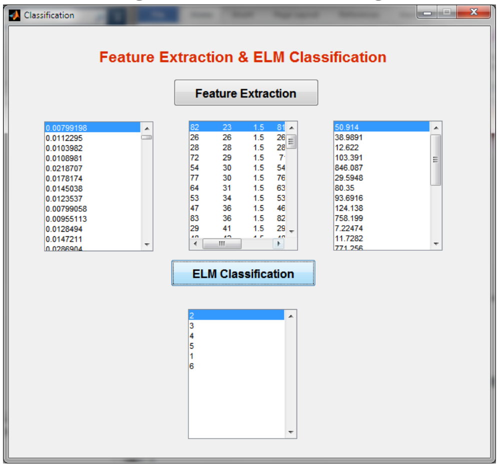

Fig: 4.13 ELM Classification

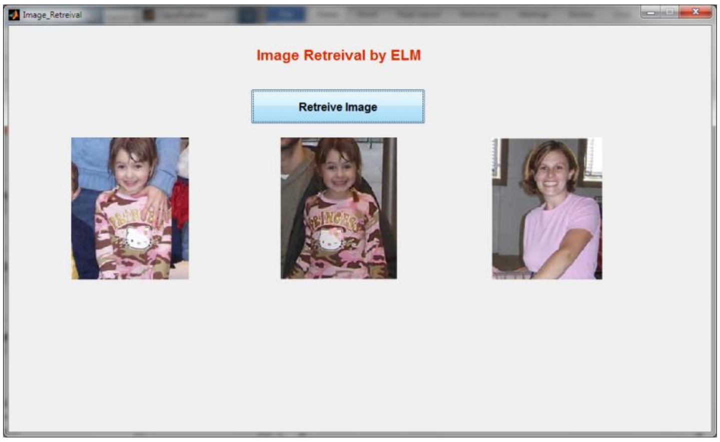

Fig: 4.14 Image Reterival

V. CONCLUSION

This project proposed the clothing retrieval using the ELM. First the face was detected. The skin, clothing and non clothing features are segmented using the graph cut algorithm. The dominant color patches are extracted. After extraction of color patches constructed the codebook. The SIFT,GIST and PHOG features are extracted and finally applied the ELM classification.

VI. FUTURE WORK

This project extracted the features such as SIFT, GIST and PHOG. In future the LBP (Local Binary Pattern), LVP(Local Vector Pattern) and LDP(Local Derivative Pattern) features extracted instead of above texture features.

REFERENCE

[1] Ming Yang, Kai Yu " REAL-TIME CLOTHING RECOGNITION IN SURVEILLANCE VIDEOS" international conference on image processing 2011.

[2] Y. Song andT.Leung, "Context-aided human recognition: Clustering," in Proc. ECCV, 2006, pp. 382–395.

[3] Xinmei Tian ,Linjun Yang,Jingdong Wang"Bayesian Video Search Reranking" MM'08, October 26–31, 2008, Vancouver, British Columbia, Canada.

[4] W. H. Hsu, L. S. Kennedy, and S.— F. Chang, "Video search reranking via information bottleneck principle," in Proc. ACM Int. Conf. Multimedia, 2006, pp. 35–44.