

Survey on Supervised Method for Face Image Retrieval Based on Euclidean Distance

Priyanka T
Vivekanandha college of
Engineering for Women
Tiruchengode, Tamilnadu
India

Kalaichelvi M
Vivekanandha college of
Engineering for Women
Tiruchengode, Tamilnadu
India

Uma Maheswari C
Vivekanandha college of
Engineering for Women
Tiruchengode, Tamilnadu
India

Abstract: Content-based image retrieval is a technique which uses visual contents to search images from large scale image databases according to users' interests. Given a query face image, content-based face image retrieval tries to find similar face images from a large image database. Initially face of the image is detected from the query image. After the removal of noise present in the image, it is separated as patches. For each patch, the Local binary pattern (LBP) is extracted which improves the detection performance. LBP is a type of feature used for classification in computer vision. The LBP operator assigns a label to every pixel of a gray level image. The label mapping to a pixel is affected by the relationship between this pixel and its eight neighbors. Support Vector Machine (SVM) is used then which will produce a model (based on the training data) that predicts the target values of the test data given only the test data attributes. When the feature values are provided to the SVM classifier, it will train about the feature. Finally it will classify about the result. SVM maps input vectors to a higher dimensional vector space where an optimal hyper plane is constructed. Among the available hyper planes, there is one hyper plane alone that maximizes the distance between itself and the nearest data vectors of each category. The Euclidean distance between the query image and database image is calculated and the index of the Euclidean distance is sorted. The indexing scheme used for this purpose provides an efficient way to search the image. Then the corresponding image from the database is retrieved based upon the index. This SVM classifier mainly improves the detection performance and the rate of accuracy.

Keywords: Content based face image retrieval, local binary pattern, support vector machine, Euclidean distance

1. INTRODUCTION

In typical content-based image retrieval systems, the visual contents in the images from the database are extracted and described by multi-dimensional feature values. The feature vectors in the images from the database form a feature database. Users provide the retrieval system with example images or sketched figures to retrieve images. The similarities /distances between the feature vectors of the query example or sketch and those of the images in the database are then calculated and retrieval is performed with the aid of an indexing scheme.

Recent retrieval systems have incorporated users' relevance feedback to modify the retrieval process in order to generate perceptually and semantically more meaningful retrieval results. Image content may include both visual and semantic features. Visual features are both general or domain specific. General visual content include color, texture, shape, etc. Domain specific visual features like the human faces, is dependent on application and may involve domain knowledge.

Semantic content is obtained either by textual annotation or by complex inference procedures based on visual content. It concentrates on general visual contents descriptions and domain specific and semantic features. A good visual feature descriptor should be invariant to the accidental variance introduced by the imaging process (e.g., the variation of the illuminating of the scene). However, there is a tradeoff between the invariance and the discriminative power of visual features, since a wide range of invariance loses the ability to discriminate between essential differences.

Invariant description has been largely investigated in computer vision (like object recognition), but it is new in image retrieval. A visual feature descriptor can be local or global descriptor. A global descriptor uses the visual features

from the whole image but a local descriptor uses the visual features of regions or objects to describe the image content.

Human attributes such as gender, race, hair style are high-level semantic descriptions about a person. The recent work shows automatic attribute detection has adequate quality (more than 80% accuracy) on many different human attributes. Using the described human attributes, many researchers has been achieved promising results in different applications such as verification of face, identification of face, keyword-based face image retrieval, and similar attribute search. These results indicate the power of the human attributes on face images.

Although human attributes have been shown useful on applications related to face images, it is non-trivial to apply it in content-based face image retrieval task due to several reasons. First, human attributes only contain limited dimensions. When there are too many people in the dataset, it loses discriminability because certain people might have similar attributes. Second, human attributes are represented as a vector of floating points. It does not work well with developing large-scale indexing methods, and therefore it suffers from slow response and scalability issue when the data size is huge.

2. LITERARY SURVEY

2.1 Active Shape Models (ASM)

Active shape models (ASMs) are statistical models of the shape of objects which iteratively deform to fit to an example of the object in a new image. ASM relies upon each object or image structure being represented by a set of points. The points represents the boundary of the image and internal features, etc.. The sets of points are aligned automatically to minimize the variance in distance between equivalent points.

Points are placed in the same way on each of a training set of examples belonging to the object. By the examining of statistics of the positions of the labeled points a model is derived called "Point Distribution Model (PDM)". The model gives the average positions of the set of points, and it has a various number of parameters which control the main modes of variation found in the training set. The ASM algorithm aims to match the model to a new image.

The ASM performs by alternating the following methods:

- Generate a suggested shape by looking in the image around each point for a better position for that point. This is usually done using a model called a "profile model", which looks for strong edges to match a model template for the point.
- Conform the suggested shape to the point distribution model, commonly called a "shape model".

One of the main drawbacks of the approach is the amount of labeled training examples required to build a good model which are very time consuming to generate. They are also not appropriate for objects with widely varying shapes.

2.2 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a statistical procedure that uses orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables. Those variables are called the principal components. The number of principal variables is less than or equal to the number of original components. The transformation of the variables is defined in such a way that the first principal component has the largest possible variability, and each succeeding variable in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components.

The objectives of principal component analysis are to discover or to reduce the dimensionality of the database, to identify meaningful underlying variables. PCA is sensitive to the relative scaling of the original variables. PCA defines a new orthogonal coordinate system that optimally describes variance in a single dataset. The limitations include the results of PCA depend on the scaling of the variables and the applicability of PCA is limited by certain assumptions made in its derivation.

2.3 Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a classification method that is mathematically robust and often produces models whose accuracy is as good as more complex methods. LDA is used in pattern recognition and machine learning to find a linear combination of features which characterizes or separates two or more classes of attributes or events. The resulting combination is the linear classifier for dimensionality reduction before classification.

LDA is closely related to regression analysis, which attempts to express one dependent variable as a linear combination of other features or measurements and also to principal component analysis (PCA) and factor analysis in that they both look for linear combinations of variables which best define the data. LDA explicitly tries to model the difference between the classes of data. PCA does not take into account any difference in class. Discriminant analysis is also different from factor analysis in that it is not an interdependence technique, a distinction between independent variables and dependent variables must be made. LDA works when the

measurements made on independent variables for each observation are continuous quantities.

In face recognition, each face is represented by a large number of feature values. Linear discriminant analysis is commonly used to reduce the number of features to a more manageable number before the object classification. Every new dimension is a linear combination of pixel values, which form a template. The advantages of LDA include multiple dependent variables and reduced error rates.

Drawbacks of LDA

- Linear discriminant analysis is extremely sensitive to outliers.
- The dependent variables are not correlated to the linear combination of other variables is not perfect.
- The relation between the variables are assumed to be linear in all groups.

2.4 Locality-constrained Linear Coding (LLC)

LLC is a simple but effective coding scheme which utilizes the locality constraints to project each descriptor into its local-coordinate coding system, and the coordinates which are projected will be integrated by max pooling to generate the final representation. LLC is easy to compute and gives superior image classification performance than any other approaches. It is a fast approximated method that proposes by first performing a K-nearest-neighbor search and then solving a constrained least square fitting problem, bearing computational difficulty. So even with very large codebooks, the system processes multiple frames per second.

LLC incorporates locality constraint instead of the scarcity constraint which leads to several favorable properties like better reconstruction, local smooth sparsity, analytical solution. Jinjun Wang and Jianchao Yang presented a promising image representation method called Locality-constrained Linear Coding (LLC).

LLC applies locality constraint to select similar basis of local image descriptors of a codebook, and learns a combination weight of these basis to reconstruct each descriptor. LLC coding criteria is used for codebook training, to improve the performance.

2.5 Bag-of-Words (BoW)

Automatic image annotation is the process of employing computer programs to automatically assign an unlabeled image a set of keywords or tags, each of which represents certain semantic object/concept. Bag of visual words is a sparse vector of occurrence counts of a vocabulary of local image features. To represent an image using BoW model, an image can be treated as a document. It usually requires three steps: feature, description of feature and the generation of codebook. After detecting the feature, every image is extracted by many image patches. Feature representation technique concentrates with how to represent the patches as numerical values. These values are called feature descriptors. A descriptor should be able to handle intensity variation, rotation changes and variation of affine to some level. One of the most famous descriptors is Scale-invariant feature transform (SIFT). SIFT converts each patch to some dimensional vector. After completing this step, each image will be a collection of vectors of the same dimension. The final step for the BoW model is to convert vector represented patches to "codewords", which also produces a "codebook". Each patch in an image is mapped to a certain codeword

through the clustering process and the image can be represented by the histogram of the codewords.

The disadvantages of BoW include that it ignores the spatial relationships among the patches, which is very important in image representation and the BoW model for object segmentation and localization is not well understood.

Lei Wu and Steven C. H. Hoi proposed a Semantics-Preserving Bag-of-Words (SPBoW) model, for the measurement of the semantic gap, it takes the distance between the semantically identical features and tries to train a codebook by reducing the semantic gap. They formulate the codebook generation task as a distance metric learning problem, which can be formalized as semidefinite programming (SDP). This approach overcomes the limitation of semantics lost in BoW models by bridging the semantic gap via distance metric learning method. An efficient algorithm is implemented to solve the codebook learning task and the optimization problem. The SPBoW can automatically decide the size of the codebook for each category. There are two important issues for SPC “Semantics-Preserving Codebook”, such as assignment of codebook size, and generation of visual word.

2.6 Local Binary Pattern (LBP)

Local Binary Patterns is one of the best local feature describer operator which is nonparametric and computationally simple that is used to describe the local spatial structure of an image. LBP is a simple yet very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number.

The most important property of the LBP operator in real-world applications is its robustness to monotonic gray-scale changes occurred such as variation of illumination. One more important property is its simplicity regarding computational that makes the analysis of images in challenging real-time settings.

The LBP pattern encodes the textures of the facial regions while the whole shape of the face is recovered by the construction of the face feature histogram. This histogram effectively has a description of the face on three different levels of locality: the LBP labels for the histogram contain information about the patterns on the pixel-basis, the labels are concatenated over a small region to produce information on a regional level and the regional histograms are concatenated to build a global description of the face.

The idea behind using the LBP features is that the face images can be seen as composition of micro-patterns which are invariant with respect to monotonic transformations into grey scale image. Grouping these patterns, a description of the face image is obtained.

Timo Ahonen and Matti Pietikainen proposed a novel approach to face recognition for the purpose of representing the facial images which considers both shape and texture information. In this approach initially the face area is divided into very small patches. From these patches, Local Binary Pattern (LBP) histograms are extracted and concatenated into a single, improved feature histogram which uniquely represents the face image. The recognition will be performed using an attribute classifier called nearest neighbor classifier in the computed feature space. It increases the robustness of the systems against different factors.

Due to its discriminative power and simplicity regarding simplicity, LBP operator is widely used approach in various applications. LBP recognizes the facial expression and also used in many other applications of biometrics like localization

of eye and recognition of iris and fingerprint, palmprint recognition, and facial age classification.

2.7 Support Vector Machines (SVMs)

Support Vector Machines (SVMs) are supervised learning models with associated learning algorithms that analyse data and pattern recognition. SVM is primarily used for the analysis of classification and regression. The SVM considers an input data to predict that which of two classes forms the output. SVM is a binary classifier. Given a training data, marked as belonging to one of two classes, an SVM algorithm builds a template which assigns the example data into to the belonging class. SVM model is a representation of the example data, mapped so that the examples of the separate classes are segmented by a gap that is as big as possible. The example data are then mapped into that same data and then predicted to belong to a class based on gap formed. SVM can perform linear classification as well as non-linear classification by implicitly mapping their inputs into high-dimensional feature spaces. There are three possible extensions to SVM which are listed below:

- Multiclass SVM
- Transductive SVM and
- Structured SVM

3. CONCLUSION

In this survey overviewed various image classifiers needed for the purpose of efficient face image retrieval. Among all of them the Support Vector Machine (SVM) is the better classifier. SVMs are helpful in text and hypertext categorization as they can significantly reduce the need for labelled training instances in both the standard inductive and transductive settings. SVMs are also useful in medical science to classify proteins with up to 90% of the compounds classified correctly. Hand-written characters can be recognized using SVM. Thus the content based face image retrieval system retrieves similar images from the large image database when a query image is provided. The retrieval is followed by initially detecting the face from the image and performing some of the preprocessing steps to remove certain noise which will be present in the image. The extracted LBP feature assigns labels to every pixel of the gray converted image. Then SVM classifier classifies and predicts about the image. This will calculate the Euclidean distance of the both the query and the dataset image. This system greatly reduces the quantization error and can achieve salient gains in face retrieval from the database. And also avoids misclassification of the images with the help of the SVM classifier.

4. ACKNOWLEDGMENTS

We thank the reviewers and others those who have helped us in preparing this.

5. REFERENCES

- [1] Belhumeur.P, Boulton.T, Kumar.N, and Scheirer.W, (2012) “Multi-attribute spaces: Calibration for attribute fusion and similarity search” in Proc. IEEE Conf. Computer Vision and Pattern Recognit.
- [2] Cheng.W.H, Lin.H.T, Kuo.Y.H and Yang.Y.H, (2011) “Unsupervised auxiliary visual words discovery for large-scale image object retrieval.” in Proc. IEEE Conf. Computer Vision and Pattern Recognit.

- [3] Coates.A and Ng.A.Y, (2011) “The importance of encoding versus training with sparse coding and vector quantization,” in *Proc. ICML*.
- [4] Boulton.T.E, Kumar.N, Ricanek.K and Scheirer.W, (2011) “Fusing with context: A Bayesian approach to combining descriptive attributes,” in *Proc. Int. Joint Conf. Biometrics*.
- [5] Davis.L.S, Feris.R.S, and Siddiquie.B, (2011) “Image ranking and retrieval based on multi-attribute queries,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognit.*
- [6] Jain.A.K and Park.U, (2010) “Face matching and retrieval using soft biometrics,” *IEEE Trans. Inf. Forensics Security*, vol. 5, no. 3, pp. 406–415.
- [7] Ho.S.C.H, Wu.L, and Yu.N, (2010) “Semantics-preserving bag-of-words models and applications,” *IEEE Trans. Image Process.*, vol. 19, no. 7, pp. 1908–1920.
- [8] Cao.Z, Tang.X, Sun.J, and Yin.Q, (2010) “Face recognition with learning based descriptor,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognit.*
- [9] Ke.Q, Shum.H.Y and Sun.J, (2010) “Scalable face image retrieval with identity-based quantization and multi reference re-ranking,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognit.*
- [10] Belhumeur.P.N, Bergand.A.C, Kumar.N and Nayar.S.K, (2009) “Attribute and simile classifiers for face verification,” in *Proc. Int. Conf. Computer Vision*.
- [11] Ganesh.A, Sastry.S, Wright.J and Yang.A, (2009) “Robust face recognition via sparse representation,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227.
- [12] S. Milborrow and F. Nicolls, (2008) “Locating facial features with an extended active shape model,” in *Proc. Eur. Conf. Computer Vision*.
- [13] Lazebnik.S, Ponce.J and Schmid.C, (2006) “Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories,” in *Proc. IEEE Conf. Computer Vision and Pattern Recognit.*
- [14] T. Ahonen, A. Hadid, and M. Pietikainen, (2004) “Face recognition with local binary patterns,” in *Proc. Eur. Conf. Computer Vision*.
- [15] Lowe.D, (2003) “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vision*.
- [16] Gionis.A, Indyk.P, and Motwani.R, (1999) “Similarity search in high dimensions via hashing,” in *Proc. VLDB*.