ISSN 2229-5518

Similarity based Query Optimization on Map Reduce using Euler Angle Oriented Approach

Dileep Kumar Kadali, R.N.V.Jagan Mohan and Y. Vamsidhar

Abstract— In recent years, large data sets are used not only for internet application but also used on normal applications. A beneficial research results in the area of similarity-based data retrieval techniques, a favorable approach. As a result, the inclusion of similarity-based retrieval techniques of image data into DBMS is currently an active research issue. In query information system the query-by-example is a method which allows the user to provide an image as a query and to search for similar images based on one-to-one or all comparisons of features of the query in large image database. While studying of this paper, we described Euler's angle oriented recognition on Map Reduce algorithm. In this algorithm, the data made into two clusters namely, Clock wise and Anti-clock wise rotations. The image is extracted using an angle oriented DCT that invokes certain normalization techniques. Furthermore, face matching is compared to the technique of similarity based query optimization operation using Tanimatto distance. It is observed from the results that use of reliable optimization technique gives a recognition rate is high.

Index Terms— Angle oriented; Discrete Cosine Transform; Map Reduce; Query Optimization; Similarity etc.

1 INTRODUCTION

Chellappa et al., 1995 [4] discussed the categories of face recognition, which falls into two categories feature based and Holistic approaches. Feature based face recognition relies on detection and characterization of the individual facial features like eyes, nose and mouth etc., and their geometrical relationships. On other hand the holistic approach, involves encoding of the entire facial image. Earlier works on these lines were proposed by various authors: Ziad M.Hafed and Martin Levine 2001[16] proposed a holistic approach to face recognition, an affine transformation was used to correct for scale, position and orientation changes in faces. Ting Shan et al., 2006 [15] developed a face model which can be used to interpret facial features and synthesize realistic frontal face images when a single face novel image given. Jagan Mohan R.N.V., and Subbarao R.,2011 [6] given a new computational approach that is converting the input image to database image using angle orientation technique and developed an algorithm for face recognition system using DCT and studied K-Means Cluster Reliability.

Jagan Mohan RNV et al., 2011 [6] discussed the kinds of angle oriented algorithm which falls into two categories, is to identifying the feature images of the faces even though they are angle oriented. If the input image angle is not 90^0, rotate the image to 90^0 and then apply normalization technique such as geometric and illumination technique. Recognition of an image by using rotational axis is easy to achieve or recognize the face. When the input image rotates from horizontal axis to vertical axis the face rotates anti-clock wise and the face appears in which it is the same as the database pose, then the object is recognized. Similarly, when the input image rotates from vertical axis to horizontal axis the face rotates clock wise and the face appears in which it is the same as the database pose, then the object is recognized. Therefore if input image is angle oriented, the pose is changed or angle is altered using rotational axis and then compared.

Different authors discussed the image retrieval application of the image database, the user takes an image is he or she are involved in and wants to find similar images from the entire database. Among these two-steps the first step is for each image in the database a feature vector describing some image properties is computed and stored in a feature database. The given a query image, its feature vector is computed and compared to the feature vectors in the feature database, and images most similar to the query image are returned to the user in the second category of to search the image database. The features and the similarity measure used to compare two feature vectors should be efficient enough to match similar images as well as being able to discriminate dissimilar ones.

In this paper we introduced cluster classification for Angle Oriented approaches in section1.The Section 2 deals with the Map Reduce. Query Optimization on image database discussed in Section 3. The basics algorithm for Euler angle oriented approach system is discussed in section 4. Section 5 deals with the experimental results of the proposed system are highlighted. The conclusion and future perceptive are mentioned in section 6

- *Dileep Kumar Kadali, is currently pursuing masters degree program in Computer Science & engineering in Swarnandra College of Engg & Tech, Narsapuram, A.P. email:dileepkumarkadali@gmail.com*
- *R.N.V.Jagan Mohan, is Assistant Professor in Swarnandra College of Engg & Tech, Narsapuram, A.P. email:mohanrnvj@gmail.com*
- *Y. Vamsidhar, is HOD & AssociateProfessor in Swarnandra College of Engg & Tech, Narsapuram, A.P. email: scetcsehod@gmail.com*

2 TAXONOMY OF CLUSTERS

The Task of classification is the assigning objects to one of several pre-defined categories. Classification is a technique is a systematic approach for building classification models from an input dataset. Many authors discussed several approaches in classification and decision-tree based methods. Decision-tree is fast and easy-to-use approach for rule generation and classification problems. Kun-Che Lu and Don-Lin Yang proposed a

general framework based on the decision-tree for mining and processing image data [9]. Association rule and Decision-tree based methods for fuzzy rule generation were offered by Ferenc Peter Pach and Janos Abonyi [5]. Many studies have considered positively the induction and analysis of decision trees [11], [7], [8] and [17].

In this paper, we introduced a cluster classification tree for clustering database images. Entire database of the images represented as C which is the root node, divided into two groups of major similar objects called clusters C_1 and C_2 representing internal nodes. The similar image object groups being rotated in the clock-wise direction belong to the internal node, cluster C_1. On the other hand, C_2 which has a group of similar image objects rotate in anti-clock wise direction. Again, the cluster C_1 is grouped into three terminal nodes (segments), C_{11}, C_{12} and C_{13}, called nested clusters. Likewise C_{21}, C_{22} and C_{23} are defined as nested clusters, the three terminal nodes of cluster C_2. If θ is the angle of rotation in each cluster, the nested cluster C_{11}consists of images with angle 0 to 30^0, C_{12} comprises the images in between 31^0 to 60^0 and C_{13} bears 61^0 to 90^0. The same type of segmentation followed in internal node of C_2 in which the images rotate with an interval of 30^0 starts with 0 and ends at 90^0 in anti-clock wise direction. The cluster tree diagram is given in figure 1.1.

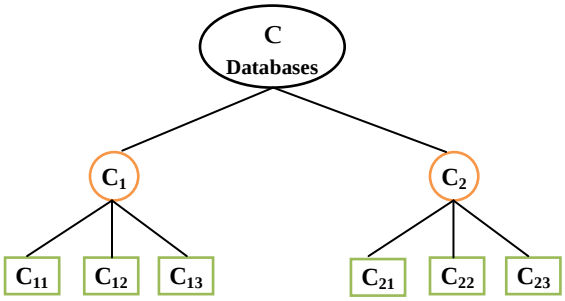


Figure 1.1: Cluster tree diagram

3 MapReduce Query Processing

MapReduce is a well-known technology used for query processing. It is a process and large amount of data set by dividing the various clusters. It is used for processing and generating the various clusters. MapReduce is simplifying the distribution of application system by providing two interfaces: Map and Reduce. To build applications on MapReduce, users must transform and code them as customized map and reduce functions. One can be identify the drawback of MapReduce is its lack of elevated and declarative languages. Hive [1,2] and Pig [3] suggested in much comparison, SQL, which is supported by most DBMSs, hides implementation details for example access method, plan optimization, there by simplifying application programming. Recently, some high-level languages have been proposed for MapReduce.

In this regard, we discuss the how to process query in the MapReduce framework. And then, we propose our query optimization problem. MapReduce was introduced by Google as a new processing technique for handling large-scale data analysis. It works on distributed and parallel file systems (DPFS). To simplifying the parallel processing, data in DPFS are separated into equivalent portions. MapReduce splits the development of parallel programs falls into two stages: map and reduce. In the map phase, each mapper loads a data portions from DPFS and transforms it into a list of key-value pairs. The key-value pairs are buffered as r local files, where r is the number of reducers. All key-value files are sorted by keys. In the reduce phase begin, When mappers finish their processing. The key-value files are shuffled to the reducers, where files from different mappers are combined together. For values with the same key, the user defined processing logic is applied by the reducer and a new key-value pair is generated as the result. Finally, the results are written back to DPFS.

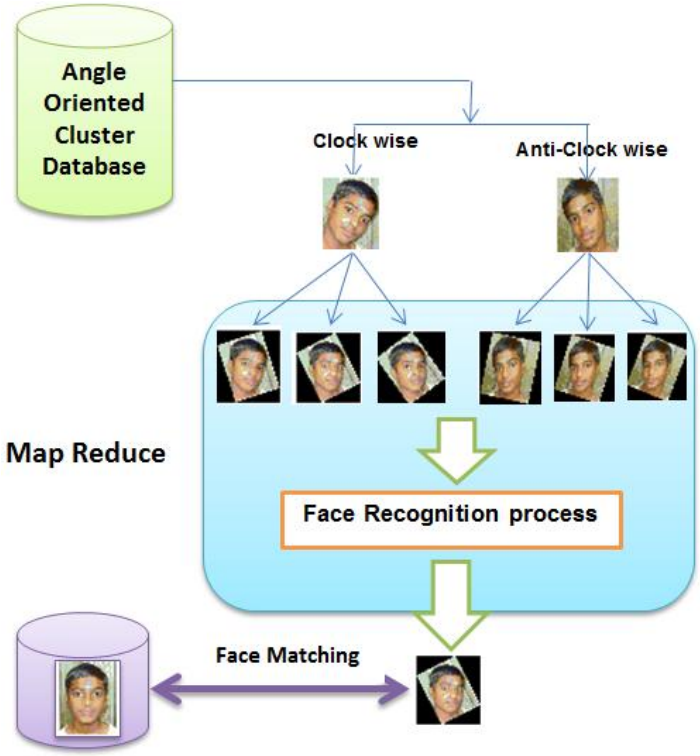


Figure 3.1: Map Reduce for Face Recognition

4 Query by Example

The Similarity with query by example is suggested by different authors. The main goal of similarity based query optimization, when users submit an image as a query to the system to get in return this most similar image to the image query according to particular visual principles. User may want to know, how to select visual descriptors. However, it's generally that do not have any idea of the most perceptive descriptor to use for their query. In many cases, our approach finds assigns priorities on clusters of image descriptions and consequently proposes continuously cluster based query optimization the query by example execution.

4.1 Similarity Based Query Optimization

Earlier works on similarity based query optimization is proposed by various authors are as follows. Solomon Atnafu,

Lionel Brunie and Harald Kosch, 2001[12], T. Seidl and H.P. Kriegel, 1997[14] working on large image database however, as a substitute of exact match; all images that are similar to the query image with a specified degree of similarity are returned as a query result. The above cited in query processing is based on similarity of images on their features. When an image is inserted into database its feature is extracted using feature extraction functions and stored with the image in database as an additional attribute of the images. The features of the image may be indexed by using multi-dimensional indexing methods to facilitate image database query processing. Similarity-based query processing is then based on the image features not on the actual image.

A query posed on image database among other things contains the query image, value of features and the threshold values. During similarity based query processing inside of the query image is extracted and then matched with the feature of images in the image database. Those images are stored in the image database that match the query image with a given threshold will be retrieved as the result of the query. As it was discussed above, query processing on image database is not exact match rather it is based on degree of similarity between images. Hence the query model of traditional database systems is not directly applicable. To commonly used query models for image database are nearest neighbor query models.

Previous works on nearest neighbor search is recommended by various authors J.H. Lee, G.H. Cha and C.W. Chung, 1999[9]. The nearest neighbor query model returns number of images that are more similar to the query image than any other images in the image always get images in response to his query even though they are not similar. On the other hand, returns all images in the image database that falls in a given threshold. The threshold value is given by the user and database. In this method processing of query of the form retrieve images similar to a query image, involves extracting the features of query image, matching it with feature of all images in the database and retrieving images that are similar to query image in sorted order. The nearest neighbor model always returns output images even though they are not similar. Hence the user can it shows the maximum distance between the value of feature of the query image and the result images. The query may return no image at all or all images depending on the value of the threshold and the distribution of the image collection.

5 Classifier for Distance Methods

The feature vectors obtained by familiarizing various transformations algorithms like DCT, KLT, DWT etc., have certain confines such as poor discriminatory power and ability to handle large computational load. While studying of this section, Tanimatto distance classifier methods for image matching are discussed.

Image matching classification system analyzes the numerical properties of various image features and organizes and data into categories. This classification includes a wide range of theoretic-decision approaches for the identification of images for each image debits one or more features and each of these features belongs to one of several distinct classes. In practice, the minimum distance classifier works well, provided the distance between means is large when compare to the randomness of each class with respect to its mean. The minimum distance classifier is used to categories unknown image data classes and therefore, the minimum distance between the image data and the classes in multi feature space exists. The distance is defined as an index of similarity and in consequence the minimum distance is identical to the maximum similarity. This paper describes Tanimatto distance as follows shown in below.

5.1 Tanimatto Distance Measure

The Tanimatto distance classifier is a comprehensive of jaccard coefficient and can be used for document data but reduce the jaccard coefficient in the case of binary attributes. This is represented as T.

$$\text{Tanimatto Distance} \quad \frac{X.Y}{||X||^2+||Y||^2-X.Y}$$

When the "similarity ratio" is given over bitmaps, where each bit of a fixed size array represents the presence or absence of a character being modeled. The definition of the ratio is the number of common bits divided by the number of bits set in either sample. The same calculation is expressed in terms of vector product and magnitude. This representation relies on the fact that, for a bit vector (where the value of each dimension is either 0 or 1).

6 The Basic Algorithm for Similarity Based Query Optimization

1) Select normalized input image classification i.e., Clock wise rotation and Anti-clock wise rotation.
2) Choose the no.of clusters
3) Select the each cluster in no.of image objects
4) Select the database image from database as the initial centroid points
5) Choose DCT based feature extraction of feature vectors in x (input) and y (Database).
6) for each x and y calculate the distances (using any distance classifier method)
7) Subtract y from x i.e., equal to c=x-y;
8) Calculate the mean of z
9) Verify mean value of z either positive or negative.
10) If $z \leq 0$ then
    Return" we may say that match between images".
  Else if z>1 then
    Return "No match between images".
  End if
 End for
11) Re-compute the centroid of each cluster.
12) until the centroids don't change.

7 Experimental Results for Angle Oriented Image Database

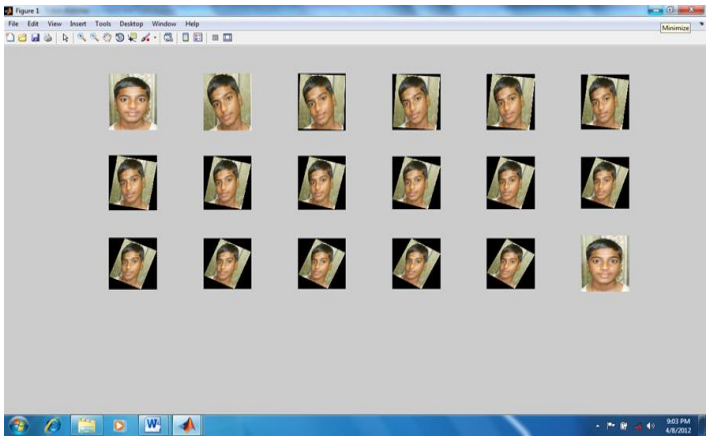
In this section, we discuss the retrieval plan of stability be-

tween computational complexity and retrieval accuracy. We focus on the comparison of two important retrieval things, namely retrieval accuracy and the speed. The test image is classified into two groups, namely clockwise rotation and anticlockwise rotation for each of them various θ values of both images. The Recall Rate of software reliability is used for efficiency of the angle oriented image database.

$$\text{Recall Rate} = \frac{\text{Number of relevant items retrieved}}{\text{Total number of relevant items in collection}}$$

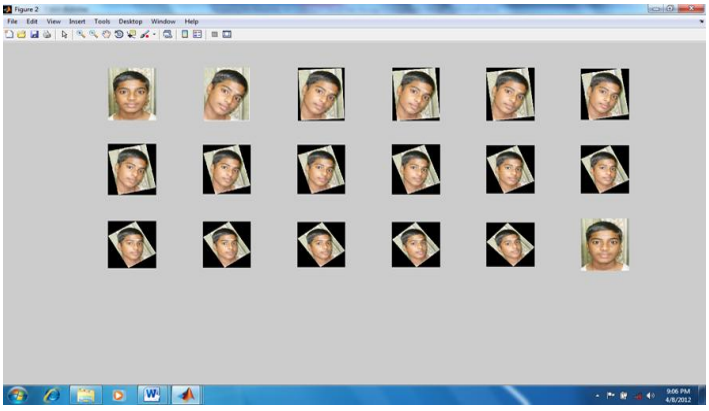
7.1 Anticlockwise Rotations

If test image is not equivalent to the database pose then test image is rotated to the horizontal axis to vertical axis and compared to database pose.



7.2 Clockwise Rotation

If test image is compared with the equivalent database image pose. Otherwise, it is not equivalent then test image rotates from vertical axis to horizontal axis to match with the database pose.



8 Conclusion and Future Works

The perceptive of this paper, Map Reduce with angle oriented face recognition using Tanimatto distance algorithm is proposed. While studying of this paper the input image is classified into anti-clock wise rotation and clock wise rotation and also the cluster classification is studied. We introduced the Map Reduce algorithm Cluster based angle oriented images is adopted. The efficient cluster reliability of cluster recall rate on face matching like similarity based algorithm is performed.

9 References

[1] A.Thousoo, R.Murthy, J.S.Sarma, Z.Shao, N.Jain, P.Chakka, S.Anthony, H.Liu and N.Zhang. Hive-a petabyte scale data warehousing using hadoop. In ICDE, 2010.

[2] A.Thousoo, J.S.Sarma, N.Jain, Z.Shao, P. Chakka, S.Anthony, H.Liu, P.Wychoff, and R.Murthy. Hive- a warehousing solution over a map-reduce framework. In VLDB, 2009.

[3] C. Olston, B.Read, U.Srivastava, R.Kumar, and A.Tomkins.Pig Latin: a not-so-foreign language for data processing. In SIGMOD, 2008.

[4] Chellappa, R., Wilson, .C, and Sirohey, S. 1995. Human and machine recognition of faces: A survey. In Proc. IEEE, 83(5):705-740.

[5] Ferenc Peter Pach and Janos Abonyi.: Association Rule and Decision Tree based Methods for Fuzzy Rule Base Generation, World Academy of Science, Engineering and Technology 2006.

[6] Jagan Mohan R.N.V. and Subbarao. R.,: Increasing the Reliability of Angle Oriented Face Recognition using DCT, published In the Proceedings of International Congress on Productivity, Quality, Reliability, Optimization and Modeling, Allied Publishers 2011.

[7] Janikow. C.Z.: Fuzzy decision trees: issues and methods. IEEE Trans. Systems Man Cybernet. *Part B (Cybernetics)*, 28(1):1–14, 1998.

[8] Janikow, C.Z.: Fuzzy partitioning with fid 3.1. In Proc.18th Internat. *Conf. of the North American Fuzzy Information Processing Society, NAFIPS99*, pages 467–471, 1999.

[9] J.H. Lee, G.H. Cha and C.W. Chung. A model for k-Nearest Neighbor query processing cost in multidimensional data space. Information Processing Letters, 69(2), 1999.

[10] Kouomou-Choupo .A, Berti-Equille L., Visual Feature Mining for Adapting Query-By-Example over Large Multidisciplinary, Video, and Audio retrieval and Mining, Canada, 2004.

[11] Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, CA 1993.

[12] S. Atnafu, L. Brunie and H. Kosch. Similarity-based operators in image database systems. In Proceedings of the International Conference on Advances in Web-Age Information Management (WAIB), pages 14-25, Xi'an, China, July 2001. LNCS 2118, Springer Verlag.

[13] Taffee T. Tanimatto and David J. Rogers. A Computer Program for Classifying Plants, Science Vol 132 pp. 1115-1118, October, 1960.

[14] T. Seidl and H.P. Kriegel. Efficient user-adaptable similarity search in large multimedia databases. Proceedings of the VLDB Conference, pages 506-515, August 1997.

[15] Ting Shan, Brain, Lovell, C., and Shaokang Chen. "Face Recognition to Head Pose from One sample Image" proceedings of the 18th International conference on Pattern Recognition, 2006.

[16] Ziad Hafed, M., and Martin Levine, "Face Recognition using Discrete Cosine Transform", International Journal of Computer Vision 43(3), 167-188, 2001.

[17] Zenon A. Sosnowskic Witold Pedrycz. : The design of decision trees in the framework of granular data and their application to software quality models. Fuzzy Sets and Systems, 123:271290, 2001.