Image Retrieval from Video using Color Based Image Retrieval Algorithm

T. Durai Murugan¹  K. Balasubramannian²
¹PG Scholar  ²Assistant Professor
¹²Department of Computer Science and Engineering
E.G.S. Pillay Engineering College, Nagapattinam

Abstract—The proposed technique concurrently provided very good segmentations of the video into objects as image. Furthermore, capability of the proposed method to take into account several object-of-interest was also tested on the following experiments concerning video portioning, object detection, object motion tracking and cropping image as object image retrieval tasks, and concerning multi view of image retrieval. This project concurrently solves image retrieval problems. The proposed method provides three main benefits with respect to traditional retrieval approaches: Color based Image retrieval may be useful in many applications (e.g. video editing). It improves the retrieval process by enforcing the matches to fulfill a set of geometric constraints. The model used to represent the matching process allows us to consider more than one image region being matched in a reference image.

Key words: Image retrieval, Object Segmentation, Object Motion Tracking

I. INTRODUCTION

An document image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. Most traditional and common methods of image retrieval utilize some method of adding metadata such as captioning, keywords, or descriptions to the images so that retrieval can be performed over the annotation words. Manual image annotation is time-consuming, laborious and expensive; to address this, there has been a large amount of research done on automatic image annotation. Additionally, the increase in social web applications and the semantic web have inspired the development of several web based image annotation tools. Several methods for retrieving images on the basis of color similarity are being used. A color histogram is computed which shows the proportion of pixels of each color within the image. During the search time, the user specifies the desired proportion from which a boundary is formed in which the process uncovers the object within the boundary.

A. Image Processing:

Image processing is a method to convert an image into digital from and perform some operations on it, in order to get an enhanced image or to extract some useful information from it. It is a type of signal dispensation in which is image, like video frame or photograph and output may be image or characteristics associated with that image. Usually Image Processing system includes treating images as two dimensional signals while applying already set signal processing methods to them.

B. Color Based Image Retrieval:

Color-Based Image Retrieval technique uses three primitive features like color, texture and shape which play a vital role in image retrieval. This technique presents a novel framework using color and shape features by extracting the different components of an image using the Lab and HSV color spaces to retrieve the edge features. Invariant moments are then used to recognize the image. In this proposed work, the performance of the HSV and Lab color space approach gives better performance than RGB and HSV. The experiments carried out on the bench marked Wang’s dataset, comprising Corel images, demonstrate the efficiency of this method.

C. Object Tracking:

Our survey is focused on methodologies for tracking objects in general and not on trackers tailored for specific objects, for example, person trackers that use human kinematics as the basis of their implementation. There has been substantial work on tracking humans using articulated object models. We will however, include some works on the articulated object trackers that are also applicable to domains other than articulated objects. We will follow a bottom-up approach in describing the issues that need to be addressed when one set out to build an object tracker. The first issue is defining a suitable representation of the object. We will describe some common object shape representations, for example, points, primitive geometric shapes and object—contours, and appearance representations. The issue is the selection of image features used as an input for the tracker. We discuss various image features, such as color, motion, edges, etc., which are commonly used in object tracking.

D. Object Representation:

An object can be defined as anything is of interest for further analysis. For instance, boats on the sea, fish inside an aquarium, vehicles on a road, planes in the air, people walking on a road, or bubbles in the water are a set of objects that may be important to track in a specific domain. Objects can be represented by their shapes and appearances. In this section, we will first describe the object shape representations commonly employed for tracking and then address the joint shape and appearance representations:

1) Points

The object is represented by a point that is the centered or by a set of points. In general, the point representation is suitable for tracking objects that occupy small regions in an image.

2) Primitive Geometric Shapes

Object shape is represented by a rectangle, ellipse, etc., Object motion for such representations is usually modeled by translation, affine or projective (homographs) transformation. Though primitive geometric shapes are more suitable for representing simple rigid objects, they are also used for tracking not rigid objects.
3) **Object Silhouette and Contour**
Contour representation defines the boundary of an object. The region inside the contour is called the silhouette of the object. Silhouette and contour representations are suitable for tracking complex no rigid shapes.

4) **Skeletal Models**
Object skeleton can be extracted by applying medial axis transform to the object silhouette. This model is commonly used as a shape representation can be used to model both articulated and rigid objects.

5) **Articulated Shape Models**
Articulated objects are composed of body parts that are held together with joints. For example, the human body is an articulated object with torso, legs, hands, head, and feet connected by joints. The relationships between the parts are governed by kinematic motion models, for example, joint angle, etc. In order to represent an articulated object one can model the constituent parts using cylinders or ellipses.

E. **Selection for Tracking:**
Selecting the right features plays a critical role in tracking. In general, the most desirable property of a visual feature is its uniqueness so that he objects can be easily distinguished in the feature space. Feature selection is closely related to the object representation. For example, color is used as a feature for histogram based appearance representations, while for contour based representation, object edges are usually used as features. In general, many tracing algorithms use a combination of these features. The details of common visual features are as follows:

1) **Color**
The apparent color of an object is influenced primarily by two physical factors. 1) the spectral power distribution of the illuminant and 2) the surface reflectance properties of the object. In image processing, the RGB (red, green, blue) color space is usually used to represent color. However the RGB space is not a perceptually units form color space; that is the differences between the colors in the RGB space do not correspond to the color differences perceived by humans. Additionally, the RGB dimensions are highly correlated. In contrast, L*a*b* space is perceptually uniform color spaces, while HSV (Hue, Saturation, Value) is an approximately uniform color space. However, these color spaces are sensitive to noise.

2) **Edges**
Object boundaries usually generate strong changes in image intensities. Edge detection is used to identify these changes. An important property of edges is that they are less sensitive to illumination changes compared to color features. Algorithms that track the boundary of the objects usually use edges as the representative feature. Because of its simplicity and accuracy, the most popular edge detection approach is the Canny Edge Detector. An evaluation of the edge detection algorithms.

3) **Optical Flow**
Optical flow is a dense field of displacement vectors which defines the translation of each pixel in a region. It is computed using the brightness constraint, which assumes brightness constancy of corresponding pixels in consecutive frames. Optical flow is commonly used as a feature in motion-based segmentation and tracking applications.

Popular techniques for computing dense optical flow methods; we refer the interested reader to the survey.

4) **Texture**
Texture is a measure of the intensity variation of a surface which quantifies properties such as smoothness and regularity. Compared to color texture requires a processing step to generate the descriptors. There are various texture descriptors. There are various texture descriptors: Gray-Level concurrence Matrices (GLCM) [Haralick] (a 2D histogram which shows the concurrence of intensities in a specified direction and distance), Law’s texture measures (twenty-five 2D filters generated from five 1D filters corresponding to level, edge, spot, wave, and ripple), wavelets [Mallet] (orthogonal bank of filters), and steerable pyramids [Greenspan]. Similar to edges features, the texture features, the textures features are less sensitive to illumination changes compared to color.

F. **Segmentation:**
Segmentation algorithms are to partition he image into perceptually similar regions. Every segmentation algorithm addresses two problems, the criteria for a good partition and the method for achieving efficient partitioning:

1) **Mean-Shift Clustering**
For the image segmentations problem, propose the mean-shift approach to find clusters in the joint spatial + color space [I, u, v, x, y], where [I, u, v] represents the color and [x, y] represents the spatial location. The algorithm is the initialized with a large number of hypothesized cluster center is moved to the mean of the data lying inside the multidimensional ellipsoid centered on the cluster center. The vector defined by the old and the new cluster center is called the mean - shift vector. The mean – shift vector is computed iteratively until the cluster center does not change their positions. Note that during the mean - shift iterations, some using may get merged. We show the segmentation using he mean-shift approach generated using the source code available at mean shift segment. Mean-shift based segment requires fine tuning of various parameters to obtain better segmentation, for instance, selection of the color and spatial kernel bandwidths, and the threshold for the minimum size of the region considerably effect the resulting segmentation.

2) **Image Segmentation Using Graph-Cuts**
Image segmentation can also be formulated as a graph partitioning problem, where the vertices (pixels), V= {u, v,} of a graph image, G are partitioned into N disjoint sub graphs (regions), A, i=1; A[u]=V, i=j; by pruning the weighted edges between two sub graphs is called a cut. The weight is typically computed by color, brightness, or texture similarity between the nodes. The weights are defined based on the color similarity. One limitation of minimum cut is its bias towards over segmenting the images. This effect is due the increase in cost of a cut with the number of edges going across the two portioned segments. The normalized cut to overcome the over segmentation problem. In their approach, the cut not only depends on the sum of edge weights in the cut, but also on the ratio of the total connection weights of nodes in each partition to all nodes of the graph. For image-based segmentation, he weights between the nodes are defined by the product of the color similarity and the spatial proximity. Once the weights between each pair of nodes are
computed, a weight matrix W and a diagonal Matrix D, where \( D_{i,j} = -N_{i,j} \); \( W_{i,j} \) are constructed. The segment is performed first by computing the eigenvectors and the Eigen values of the generalized Eigen systems \( (D-W) \), and then the second smallest Eigen vectors is used to divide the image into two segments. For each new segment this process is recursively performed until a threshold is reached.  

3) K-means Clustering Algorithm  
K-means algorithm is an unsupervised clustering algorithm that classifies the input data points into multiple classes based on their inherent distance from each other. The algorithm assumes that the data features form a vector space and tries to find natural clustering in them. The point are clustered around \( \mu_i, \forall i = 1...k \) which are obtained by minimizing the objective where there are \( K \) clusters \( S_i; i = 1...k \) and \( \mu_i \) is the centered or mean point of all the points \( x_i \in S_i \).

\[
V = \sum_{i=1}^{k} \sum_{x_i \in S_i} (x_i - \mu_i)^2
\]  

4) Managing the flexibility of Geometric Transformations  
Using a Gaussian distribution to model the transformation based location provides certain degree of flexibility since the covariance matrix \( \Sigma_i \) allows us to relax the geometric constraints imposed by transformation when necessary. However, it would be desirable to have control on the flexibility of the model, so that it could fit to either more or less similarity demanding tasks.

G. Scope of the Project:  
This project deals with image retrieval from video and detecting object motion from the video. Color based image retrieval algorithm is used to find the object. Detecting image is compare with the database and it retrieve likely image from the database.

H. Organization of the Report:  
The project report is organized as follows: In Chapter 2 the literature survey and existing system is discussed. The proposed system, system architecture and system specification are in Chapter 3. The implementation is discussed in chapter 4. The report is concluded in Chapter 5.

II. EXISTING SYSTEM  
The matching process between two objects in different images, namely: objects undergoing a geometric transformation, typical spatial location of the region of interest, and visual similarity. In this manner, our approach improves the reliability of detected true matches between any pair of images. Furthermore, by taking advantages of the links to the provided by the true matches, the proposed method is able to work when there is more than one ROI in the query image. Our experiments on two challenging image retrieval datasets proved that our approach clearly outperforms the most prevalent approach for geometrically constrained matching and compares favorably to most of the state-of-the-art methods furthermore, the proposed technique concurrently provided very good segmentations of the ROI. Furthermore, the capability of the proposed methods to take into account several objects-of-interest was also tested on three experiments: two of them concerning image segmentation and object detection in multi-object image retrieval tasks, and another concerning multi View image retrieval. These experiments proved the ability of our approach to handle scenarios in which more than one object of interest is present in the query.

III. PROPOSED SYSTEM  
The proposed model provides a unified framework that takes into different stages for image retrieval:

- Video Partitioning
- Object Detection
- Object Motion tracking
- Cropping Image
- Concurrent Image retrieval from Database

The proposed technique concurrently provided very good segmentations of the video into objects as image. Furthermore, the capability of the proposed methods to take into account several objects-of-interest was also tested on the following experiments concerning video partitioning, object detection, object motion tracking and cropping image in object image retrieval tasks, and concerning multi View of image retrieval. It decreases the computation time. Further color features, Corner detection, Filtering, Similarity measures, design and development of a multistage model for image retrieval. To improve the retrieval accuracy by filtering down irrelevant images at each stage. The accuracy of region based image retrieval system is improved by introducing region based state art matching scheme. The proposed scheme several properties of two different images from the data set to make similarity of image properties in the both compare image using geometric transformation, resize the image. We except true match result only in proposed scheme so that we use spatial location of the region of interest, and visual similarity [that means same image with shape variation from differential image set]. We give input query like image or video it will match with dataset image from the system. The structural and statistical approaches of texture description are utilized to develop a single feature which can thoroughly describe the correlation between color and texture properties of image. A region based image retrieval system is developed using state-of-art method based representation of color, texture and shape features based on geometric transformation method. The notation of relative spatial location of different regions in the images is also exploited using improved region state-of-art method based matching scheme.
IV. SYSTEM ARCHITECTURE

![Diagram of system architecture](image)

V. MODEL DESCRIPTION

For implementing this system the entire system has been divided into the following modules:

- Video Partitioning
- Dividing video Frames
- Object Boundary Marking
- Retrieving the object
- Noise Reduction
- Color Features
- Harris Corner Detector
- Median Filtering
- Similarity Measures

A. Video Partitioning:

Video parsing, or called syntactic segmentation, involves temporal partitioning of the video sequence into meaningful units which then serve as the basis for descriptor extraction and semantic notation. Partitioning is done by pointing he mark-in and mark-out values based on timing running. The process of manipulating video into images is termed as video partitioning. Once the province of expensive machines called video editors, video editing software is now available for personal computers and workstations. Video editing includes cutting segments, re-sequencing clips, video partitioning and adding transitions and other special effects.

B. Dividing Video Frames:

In essence, the differences between consecutive frames are compared in terms of their pixel values, segmented regions characteristics and foreground objects size/location, and shot boundary is detected when the difference reaches a certain threshold. There are three basic common types of coded frames:

- Intra-coded frames, where the frames are coded independently of all other frames.
- Predicatively coded, or p-frames, where the frame is coded based on a previously coded frame.
- Bi-directionally predicted frames, or b-frames where the frame is coded using both previous and future coded frames. Hysteresis thresholding finds where edges begin and end.

C. Object Boundary Marking:

Based on the color based image retrieval algorithm, the objects get retrieved. This algorithm is based on color histogram is computed which shows the proportion of pixels of each color within the image. An image is basically a long string of pixels for which each pixel is identified by its place in the image matrix, its color and its intensity. An analysis of the pixel set can give information about the distribution of dominant colors, the image texture, and the shapes formed by marked change in neighboring colors.

D. Noise Reduction:

Image noise is random (not present in the object imaged) variation of brightness or color information in images, and is usually an aspect of electronic noise. Image noise is an undesirable by product of image capture that adds spurious and extraneous information. The original meaning of noise was and remains “unwanted signal”; unwanted electrical fluctuations in signals received by a radios caused audible acoustic noise “static”. By analogy unwanted electrical fluctuations themselves came to be known as “noise”, image noise is inaudible.

E. Color Features:

One of the most commonly used visual feature for image retrieval is he color feature. There are many color models like RGB, HSV, HIS used to represent the color models. Since the HSV color space is more similar to human vision system, this colorimetric approach is been used the global color histogram for extracting the color features has been adopted. The image in the general RGB (Red, Green, and Blue) color space is converted into the HSV (Hue, Saturation, Value) color space:

\[ s(H, S, V) = \sum_{x, y, z \in Z} H_{x,y,z} \]

Where X, Y, Z are the arguments of the discreditied color channels. This metric satisfies the associatively condition, finally the distance \( d_C(Q, B) \) between the query image and the database image according to the extracted color feature is given by the equation:

\[ d_C(Q, B) = 1 - S(H_Q, H_S) \]

F. Harris Corner Detector:

The Harris corner detector concept was proposed as he interest points for the image retrieval by Schmidt and Mohr. The core idea in the Harris corner detector is the usage of the auto correlation function to find the pixels were the signals change in the two directions. A matrix is constructed based on the auto correlation function, where denotes the derivation scale, denotes the integration scale, Gaussian kernel smoothed image. This matrix is based on the signal’s first derivatives. The auto correlation function’s principal curvatures are the Eigenvectors of the matrix. If both the Eigenvector values are large and distinct, then there exists a corner.

\[ E(X, c, dB) = dB2G(X, cA) = \begin{bmatrix} L_x, (X, cB) \\ L_y, (X, cB) \end{bmatrix} \]

G. Median Filtering:

The edges feature extraction process goes as given bellow:

- Convert the input image into gray scale image.
On the gray scale image apply the Histogram equalization technique.
- The empirical mode decomposition process detects the edge.
- The median filter of size 3*3 is used to smooth the histogram equalized image.
- Thus generated image is converted into the 64 bits feature vector and is put aside in the database for further usage.

H. Similarity Measures:
The corners of an image are extracted and this data is stored in the feature vector FH. Since the corners of each image differ from each other image the feature vector FH is of varying size.

\[
difference(n) = \text{abs}(H(n) - Q(n))
\]

Where \(n = 1, 2, \ldots, 64\)

This value is recorded in the feature vector FE. The HSV color features are extracted and quantized to form 11 pins for each component. The color features are stored in the FC feature vector. The final distance between the query and the input image is computed by the weighted sum of the three features vectors FH, FC and FE, with the similarity measure given below.

\[
SM = \alpha \cdot FE + \beta \cdot FH + \gamma \cdot FC
\]

Where the sum of \(\alpha, \beta, \gamma\) will be 1.

The experiments were carried out with different values of \(\alpha, \beta, \gamma\). Good results were obtained with \(\alpha = 0.5, \beta = 0.1, \gamma = 0.4\), since the edge feature is predominant in the case of scaling and rotation than the corners and also global color feature.

VI. SYSTEM MODELING

![Fig. 2:](image)

VII. CONCLUSION

A generative probabilistic solves the problem of image retrieval from video. By jointly modeling several properties of true matches namely: object undergoing a geometric transformation, typical spatial location of the region of interest, and visual similarity. Our approach improves the reliability of detected true matches between any pair of images. Furthermore, the method associates the true matches with any of the considered foreground components in the image and assigns the rest of the matches to a background region, what allows it to perform a suitable ROI segmentation.

REFERENCES