COLOR BASED VIDEO RETRIEVAL USING BLOCK AND GLOBAL METHODS

Smita Chavan
Information Technology Department, Government Engineering College, Osmanpura, Aurangabad, India

ABSTRACT

Interacting with multimedia data and video in particular, requires more than connecting with data banks and delivering data via networks to customer’s homes or offices. We still have limited tools and applications to describe, organize, and manage video data. The fundamental approach is to Index video data and make it a structured media manually generating video. Color features are key-elements which are widely used in video retrieval, video content-analysis, and object recognition. In the present study we are using the video retrieval technique, based on color content, the standard way of extracting a color from an image is to generate a color histogram. Therefore, the core research in content-based video retrieval is developing technologies to automatically parse video, audio, and text to identify meaningful composition structure and to extract and represent content attributes of any video sources. A typical scheme of video-content analysis and indexing, as proposed by many researchers, involves four primary processes: feature extraction, structure analysis, abstraction, and indexing. Each process poses many challenging research problems. The main processes in a proposed system includes search image, apply two methods for color feature extraction, find matching videos as compared with image input given, calculation of distance metrics using Euclidean distance measurement.

Keywords: Absolute Difference, Distance Metrics, Euclidean Distance, Feature Extraction, Similarity Measurement.

1. INTRODUCTION

Content-Based video search is according to the user-supplied in the bottom Characteristics, directly find out images containing specific content from the image library the basic process: First of all, do appropriate pre-processing of images like size and image
transformation and noise reduction is taking place, and then extract image characteristics needed from the image according to the contents of images to keep in the database. When we retrieve to identify the image, extract the corresponding features from a known image and then retrieve the image database to identify the images which are similar with it, also we can give some of the characteristics based on a query requirement, then retrieve out the required images based on the given suitable values. In the whole retrieval process, feature extraction is essential it is closely related to all aspects of the feature, such as color, shape, texture and space. A color image is a combination of some basic colors. In MATLAB breaks each individual pixel of a color image termed true color down into Red, Green and Blue values. We are going to get as a result, for the entire image is 3 matrices, each one representing color features. The three matrices are arranging in sequential order, next to each other creating a 3 dimensional m by n by 3 matrixes. An image which has a height of 5 pixels and width of 10 pixels the resulting in MATLAB would be a 5 by 10 by 3 matrixes for a true color image. For improved retrieval performance, results from the different content retrieval methods may be combined. Fusion of multimedia search results is query-dependent, and an area of ongoing research. Besides merging results from different content retrieval methods, when retrieving whole programs, it is necessary to combine shot-level results from within the program. There has not been much work in this area, and approaches consist of assigning binary values to each shot and aggregating these to the program level, averaging the scores of all shots for a particular feature, using the maximum score of a shot in a program, or simply using the central shot in a program to represent the whole video[1]. Overall paper contains section I introduction, section II history, section III methods, section IV implementation, section V experimental Result with BCVR and GCVR, section VI conclusion.

1.1 Color Features

![Fig-1.1: Basic Content Based Video Retrieval Method](image-url)
Color is one of the most widely used visual features in multimedia context and image/video retrieval. To support communication over the Internet, the data should compress well and be suitable for heterogeneous environment with a variety of the user platforms and viewing devices, large scatter of the user's machine power, and changing viewing conditions.

The CBVR systems are not aware usually of the difference in original, encoded, and perceived colors, e.g., differences between the colorimetric and device color data. It is also an important feature of perception, comparing with other image features such as texture and shape etc, color features are very stable and robust[2].

It is not sensitive to rotation, translation and scale changes. Moreover, the color feature calculation is relatively simple. A color histogram is the most used method to extract color features.

2. HISTORY

Shape-based image/video retrieval is one of the hardest problems in general image/video retrieval. This is mainly due to the difficulty of segmenting objects of interest in the images. Consequently, shape retrieval is typically limited to well distinguish objects in the image. In order to detect and determine the border of an object an image may need to be preprocessed. The preprocessing algorithm or filter depends on the application. Different object types such as skin lesions, brain tumors, persons, flowers, and airplanes may require different algorithms. If the object of interest is known to be darker than the background, then a simple intensity thresholding scheme may isolate the object. For more complex scenes, noise removal and transformations invariant to scale and rotation may be needed. Once the object is detected and located, its boundary can be found by edge detection and boundary following algorithms. The detection and shape characterization of the objects becomes more difficult for complex scenes where there are many objects with occlusions and shading. Once the object border is determined the object shape can be characterized by measures such as area, eccentricity (e.g., the ratio of major and minor axes), circularity (closeness to a circle of equal area), shape signature (a sequence of boundary-to-center distance numbers), shape moments, curvature (a measure of how fast the border contour is turning), fractal dimension (degree of self similarity) and others. All of these are represented by numerical values and can be used as keys in a multidimensional index structure to facilitate retrieval.

In dimensionality color histograms may include hundreds of colors. In order to efficiently compute histogram distances, the dimensionality should be reduced. Transform methods such as K-L and Discrete Fourier Transform (DFT), Discrete Cosine Transform (DCT) or various wavelet transforms can be used to reduce the number of significant dimensions. Another idea is to find significant colors by color region extraction and compare images based on presence of significant colors only partitioning strategies such as the Pyramid-Technique map n-dimensional spaces into a one-dimensional space and use a B+ tree to manage the transformed data. The resulting access structure shows significantly better performance for large number of dimensions compared to methods such as R-tree variants.

Texture and shape retrieval differ from color in that they are defined not for individual pixels but for windows or regions. Texture segmentation involves determining the regions of image with homogeneous texture. Once the regions are determined, their bounding boxes may be used in an access structure like an R-tree. Dimensionality and cross correlation problems also apply to texture and can be solved by similar methods as in color. Fractal codes capture the self similarity of texture regions and are used for texture based indexing and retrieval. Shapes can be characterized by methods and represented by n-dimensional
numeric vectors which become points in n-dimensional space. Another approach is to approximate the shapes by well defined, simpler shapes. For instance, triangulation or a rectangular block cover can be used to represent an irregular shape. In this latter scheme, the shape is approximated by collection of triangles or rectangles, whose dimensions and locations are recorded. The advantage of this approach is that storage requirements are lower, comparison is simpler and the original shape can be recovered with small error [3].

2.1 Color based video retrieval using DCT

The DCT block computes the unitary discrete cosine transform (DCT) of each channel in the M-by-N input matrix, \( y = \text{dct}(u) \) % Equivalent MATLAB code When the input is a sample-based row vector, the DCT block computes the discrete cosine transform across the vector dimension of the input. For all other N-D input arrays, the block computes the DCT across the first dimension. The size of the first dimension (frame size), must be a power of two. To work with other frame sizes, use the Pad block to pad or truncate the frame size to a power-of-two length. When the input to the DCT block is an M-by-N matrix, the block treats each input column as an independent channel containing M consecutive samples. The block outputs on-by-N matrix whose lth column contains the length-M DCT of the corresponding input column [4].

![Fig 2.1: Comparison of DCT with row and column](image)

3. METHODS

Content-Based video search is according to the user-supplied in the bottom Characteristics, directly find out images containing specific content from the image library the basic process. First of all, do appropriate pre-processing of images like size and image transformation and noise reduction is taking place, and then extract image characteristics needed from the image according to the contents of videos to keep in the database. When we retrieve to identify the video, extract the corresponding features from a known image and then retrieve the video database to identify the images which are similar with it, also we can give some of the characteristics based on a query requirement, then retrieve out the required
videos based on the given suitable values. In the whole retrieval process, feature extraction is essential it is closely related to all aspects of the feature, such as color, shape, texture and space.

Feature extraction based on the color characteristics in the proposed method is classified into two types.

3.1 Global color based method

Because an RGB color space does not meet the visual requirements of people, for video retrieval, the video normally converts from an RGB space to other color spaces. The HSV space is used as it is a more common color space.

The global color is a simple way of extracting image features. High effective calculation and matching are its main advantage. The feature is invariable to rotation and translation.

The drawback is that the global color only calculates the frequency of colors. The spatial distribution of color information is lost. Two completely different images can get the same global color difference graph, which will cause retrieval errors.

3.2 Block color based method

For the block color, an image is separated into \( n \times n \) blocks. Each block will have less meaning if the block is too large, while computation of retrieval process will be increased if the block is too small. A two-dimensional space divided into \( 3 \times 3 \) will be more effective. For each block, we carry out a calculation of the color space converting and color quantization. The normalized color features for each block can be calculated. Usually in order to highlight the specific weight of different blocks, a weight coefficient is distributed to each block, and the weight of middle block is often larger [5].

3.3 Similarity of Frame sequence

Video access is one of the important design issues in the development of multimedia information system, video-on-demand and digital library. Video can be accessed by attributes of traditional database techniques, by semantic descriptions of traditional information retrieval technique, by visual features and by browsing. Access by attributes of database technique and access by semantic descriptions of information retrieval technique are insufficient for video access owing to the numerous interpretations of visual data. Besides, the automatic extraction of semantic information from general video programs is outside the capability of the current technologies of machine vision.

3.3.1 Symmetric similarity measures

A similarity measure is symmetric if \( D(U, V) = D(V, U) \). The straightforward measure of similarity is the one to one optimal mapping. In the one to one mapping, we try to map as many pairs of frames as possible under the constraint that each frame \( u_i \) in \( U \) corresponds to only one frame \( v_j \) in \( V \). Obviously, the maximal number of mapping pairs is equal to the number of frames of shorter shot (shot with less number of frames). The mapping with minimum sum of frame distance is selected as the optimal mapping. The formal definitions are given as follows. The effect of video segmentation also affects the performance of sequence mapping. Video segmentation with lower threshold produces shots with much variation. It is possible that the shot \( U \) is very similar to the \( V \) except that some frames are very dissimilar. Using OM, these dissimilar frame pairs produce large distance. Therefore, in the next definition, the mapping is constrained by a threshold. Two frames with
distance larger than are not allowed to match. In addition, each unmatched frame is associated with a penalty value v in the computation of shot distance. Otherwise, the result of the distance-constrained optimal mapping with minimum cost would be no matching at all. No matching produces the distance of zero.

3.3.2 Asymmetric similarity measures

A similarity measure is asymmetric if $D(\mathbf{U}, \mathbf{V}) \neq D(\mathbf{V}, \mathbf{U})$. In general, asymmetric similarity measure is used to map between the query sequence of frames and video sequence of frames. The simplest proposed asymmetric similarity measure is the Optimal Subsequence Mapping (OSM). The algorithm of OSM is similar to that of OM except that the query video sequence must be the shorter sequence. Therefore, it is not necessary to compare the length between the video shot $\mathbf{U}$ and query video shot $\mathbf{Q}$ [6].

3.4 Find matching videos

Here absolute difference (AD) is used for similarity measurement of feature vectors of the videos in content based video retrieval. In this method, the feature vectors of a query video are compared with feature vectors of all the videos stored in the database by taking absolute difference between them. The differences will be ordered in increasing fashion so as to show most relevant searches on top, as the videos with which the absolute difference will be lower will be most relevant ones [4].

3.5 Calculation of Distance metrics

In the image database and video database proposed method compares with the images given in query and compare with video database features as with each frames from the video. for the each frames method calculates Euclidean distance takes the mean of the distances calculated and then it is stored into the feature vector.

4. IMPLEMENTATION

For the purpose of implementation MATLAB 2012 is used. MATLAB is a "Matrix Laboratory", and as such it provides many convenient ways for creating vectors, matrices, and multi-dimensional arrays. In the MATLAB vernacular, a vector refers to a one dimensional (1×N or N×1) matrix, commonly referred to as an array in other programming languages. A matrix generally refers to a 2-dimensional array, i.e. an m x n array where m and n are greater than or equal to 1. Arrays with more than two dimensions are referred to as multidimensional arrays. Operating system windows 7 is used for the proposed method.

The Software Requirement: Matlab (2012a), The Hardware Requirement: The RAM: 2 GB with Operating System Windows 7/XP.

The first step for video retrieval based on the partitioning of a video sequence into shots. A shot is defined as an image sequence. Key-frames are images extracted from original video data. Key-frames have been frequently used to supplement the text of a video log, though they were selected manually in the past. Key-frames, if extracted properly, are a very effective visual abstract of video contents and are very useful for fast video browsing. key-frames can also be used in representing video in retrieval video index may be constructed based on visual features of key-frames, and queries may be directed at key-frames using query by retrieval algorithms.

Next step is to extract features. The features are typically extracted with the above mentioned methods i.e. For color feature extractions methods used is global and block color
based system. Low-level features such as object motion, color, shape, texture, loudness, power spectrum, bandwidth, and pitch are extracted directly from video in the database. High-level features are also called semantic features. Features such as timbre, rhythm, instruments.

Next step is to find matching videos with both methods using Euclidean distance formula. Each video is divided into five frames. For each frame Euclidean distance is calculated take mean of that Euclidean distance, arrange it into ascending order, we will get most closed matching videos to the given image.

4.1 Color description

In general, color is one of the most dominant and distinguishable low-level visual features in describing video image. Many CBIR systems employ color to retrieve images. In theory, it will lead to minimum error by extracting color feature for retrieval using real color image directly, but the problem is that the computation cost and storage required will expand rapidly. In fact, for a given color video image, the number of actual colors only occupies a small proportion of the total number of colors in the whole color space. In the HSV color space, each component occupies a large range of values. If we use direct values of H, S and V components to represent the color feature, it requires lot of computation. So it is better to quantify the HSV color space into non-equal intervals. At the same time, because the power of human eye to distinguish colors is limited, we do not need to calculate all segments. Unequal interval quantization according the human color perception has been applied on H, S and V components [7].

4.1.1. Color set notation

Transformation

The color set representation is defined as follows: given the color triple (r; g; b) in the 3-D RGB color space and the transform T between RGB and a generic color space denoted XY Z, for each (r; g; b) let the triple (x; y; z) represent the transformed color such that (x; y; z) = T(r; g; b):

Quantization

Let QM be a vector quantizer function that maps a triple (x; y; z) to one of M bins. Then m, where m is the index of color (x; y; z) assigned by the quantizer function and is given by m = QM(x; y; z):

Binary Color Space

Let BM be an M dimensional binary space that corresponds to the index produced by QM such that each index value m corresponds to one axis in BM.

Definition 1: Color Set. A color set is a binary vector in BM. A color set corresponds to a selection of colors from the quantized color space. For example, let T transform RGB to HSV and let QM where M = 8 vector quantize the HSV color space to 2 hues, 2 saturations and 2 values. QM assigns a unique index m to each quantized HSV color. Then B8 is the 8-dimensional binary space whereby each axis in B8 corresponds one of the quantized HSV colors. Then a color set c contains a selection from the 8 colors. If the color set corresponds to a unit length binary vector then one color is selected. If a color set has more than one non-zero value then several colors are selected. For example, the color set c = (10010100)
corresponds to the selection of three colors, \( m = 7; m = 4 \) and \( m = 2 \), from the quantized HSV color space.

4.2 Database
Small Training Databases: The videos are categorized in number of frames as each video is divided into five frames assorted categories containing 8 similar videos. So, in total there are 8 videos in data set for experimentation. All the videos are normalized in each category. The videos are chosen such that those of same categories differ very minutely in color content so as to find accurate result.

Test Database: In proposed method test database contains 8 videos. Search image is compared with the video database for color feature extraction. Comparative difference graph is plotted by Euclidean distance.

Created video database is used for the proposed method.

Large Training Database: the images are considered as large training database. There are 15 search images in the image database. All the images are compared with video database for color feature extraction. Following image database is used in the proposed method.

4.3 Video retrieval algorithm is as follows
Step 1: First divide the video into no of frames.
Step 2: Uniformly divide each video in the database and the target image
Step 3: Extract Feature with color based methods for Efficient Video retrieval
Step 4: Construct a combined feature vector for color and differential graph
Step 5: Find the distances between feature vector of query Video and the feature vectors of
target video using Euclidean distance.
Step 6: Sort the Euclidean distances by comparing the query image and database video in
ascending order
Step 7: Retrieve most similar Video with minimum Distance from Video data.

5. EXPERIMENTAL RESULT WITH BCVR AND GCVR

5.1 Comparisons

In the proposed method results are shown with block color based video retrieval and
global color based video retrieval including calculation of Euclidean distances.

Table 5.1: Search image results

<table>
<thead>
<tr>
<th>Search image 12</th>
<th>Search image 3</th>
<th>Search image 14</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCVR</td>
<td>BCVR</td>
<td>GCVR</td>
</tr>
<tr>
<td>2.15444e+09</td>
<td>4.13454e+08</td>
<td>5.11385e+09</td>
</tr>
<tr>
<td>2.78191e+09</td>
<td>7.09396e+08</td>
<td>6.7226e+08</td>
</tr>
<tr>
<td>3.85552e+09</td>
<td>7.26439e+08</td>
<td>6.83174e+09</td>
</tr>
<tr>
<td>4.06727e+09</td>
<td>7.39199e+08</td>
<td>7.70484e+09</td>
</tr>
<tr>
<td>4.86135e+09</td>
<td>7.41091e+09</td>
<td>8.94948e+09</td>
</tr>
</tbody>
</table>

Fig-5.1: Search image comparison
5.2 Accuracy with GCVR and BCVR

Table 5.2: Search image accuracy percentage

<table>
<thead>
<tr>
<th>Search Image Method</th>
<th>GCVR Accuracy %</th>
<th>BCVR Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image 1</td>
<td>75</td>
<td>79</td>
</tr>
<tr>
<td>Image 2</td>
<td>65</td>
<td>69</td>
</tr>
<tr>
<td>Image 3</td>
<td>85</td>
<td>90</td>
</tr>
<tr>
<td>Image 4</td>
<td>90</td>
<td>95</td>
</tr>
<tr>
<td>Image 5</td>
<td>50</td>
<td>55</td>
</tr>
<tr>
<td>Image 6</td>
<td>90</td>
<td>91</td>
</tr>
<tr>
<td>Image 7</td>
<td>45</td>
<td>41</td>
</tr>
<tr>
<td>Image 8</td>
<td>89</td>
<td>92</td>
</tr>
<tr>
<td>Image 9</td>
<td>65</td>
<td>76</td>
</tr>
<tr>
<td>Image 10</td>
<td>32</td>
<td>43</td>
</tr>
<tr>
<td>Image 11</td>
<td>67</td>
<td>53</td>
</tr>
<tr>
<td>Image 12</td>
<td>89</td>
<td>92</td>
</tr>
<tr>
<td>Image 13</td>
<td>91</td>
<td>96</td>
</tr>
<tr>
<td>Image 14</td>
<td>92</td>
<td>96</td>
</tr>
<tr>
<td>Image 15</td>
<td>54</td>
<td>60</td>
</tr>
</tbody>
</table>
6. CONCLUSION

Despite the considerable progress of academic research in video retrieval, there has been relatively little impact of content based video retrieval research on commercial applications with some niche exceptions such as video segmentation. Choosing features that reflect real human interest remains an open issue. One promising approach is to use Meta learning to automatically select or combine appropriate features. Another possibility is to develop an interactive user interface based on visually interpreting the data using a selected measure to assist the selection process. We would be interested to see how these techniques can be incorporated in an interactive video retrieval system. The usefulness of color correlation statistics in semantic retrieval of video and images. It shows HSV Color is to be efficient feature over the other more traditional color-based features. MPEG-7 color structure descriptor is similar to HSV color correlogram and is very efficient in retrieval of images/videos. This proposed method covers the following tasks: image search that means search is based on the images, extraction of features of static key frames using color features specifically HSV color features using conversion of RGB to HSV, video may be TV video, natural images, collection of frames, video search including interface, similarity measure, video retrieval and distance metrics, we made program in two parts database creation, videos, samples, one program is for video file reading then save into database. Search images option will search images which are closer to the video. Advanced filtering and retrieval methods for all types of multimedia data are highly needed. Current image and video retrieval systems are the results of combining research from various fields. Hence color-based retrieval has been the backbone for content-based image and video retrieval compared with the other methods like shape or texture.

6.1 Future Scope

Many issues are still open and deserve further research, especially in the following areas. Most current video indexing approaches depend heavily on prior domain knowledge. This limits their extensibility to new domains. The elimination of the dependence on domain
knowledge is a future research problem. Fast video search using hierarchical indices are all interesting research questions. Video indexing and retrieval in the cloud computing environment, where the individual videos to be searched and the dataset of videos are both changing dynamically, will form a new and flourishing research direction in video retrieval in the very near future. Video affective semantics describe human psycho-logical feelings such as romance, pleasure, violence, sadness, and anger.

6.2 Applications

Specially used in supercomputers and mainframe computers, because they have high computation time. In professional and educational activities that involve generating or using large volumes of video and multimedia data are prime candidates for taking advantage of video-content analysis techniques. Automated authoring of web content media organizations and TV broadcasting companies have shown considerable interest in presenting their information on the web. Automated authoring of web content searching and browsing large video archives major news agencies and TV broadcasters own large archives of video that have been accumulated over many years. Intelligent video segmentation and sampling techniques can reduce the visual contents of the video program to a small number of static images. Easy access to educational material the availability of large multimedia libraries that we can efficiently search has a strong impact on education. Indexing and archiving multimedia presentation.

Consumer domain applications, video content analysis research are geared toward large video archives. However, the widest audience for video content analysis is consumers. We all have video content pouring through broadcast TV and cable. Also, as consumers, we own unlabeled home video and recorded tapes. To capture the consumer’s perspective, the management of video information in the home entertainment area will require sophisticated yet feasible techniques for analyzing, filtering, and browsing video by content [8].

7. REFERENCES


[10] Che-Yen Wen, Liang-Fan Chang, Hung-Hsin Li, Content based video retrieval with motion vectors and the RGB color model, Received: October 31, 2007 / Accepted: November 12, 2007.


