Shape-Based Retrieval of Arbitrarily Shaped Video Objects

Mohammed Jassim Mohammed Jassim

Abstract- The increasing availability of object-based video content requires new technologies for automatically extracting and matching of the low level features of arbitrarily shaped video. In this paper, methods for shape retrieval of arbitrarily shaped video objects are proposed. Along with still shape features the shape deformations that may occur in an object's life span is also taken. In this paper a novel method is used for measuring shape similarities between arbitrarily shaped video objects by comparing the low-level still shape features of the representative frames of the video objects.

Keypads:- object-based, video, extracting, shape similarities, video objects.

I. INTRODUCTION

The advancements in technologies for extracting video objects in a scene, such as cameras equipped with depth sensors that capture three-dimensional (3-D) video and naturally segmented objects, combined with the advancements in video object segmentation technologies, and efficient object-based video representations availability, such as MPEG-4 results in increased availability of arbitrarily shaped digital video content. The most recent MPEG video coding standard, MPEG-4, offers an object-based representation of video, where individual video objects (Vos) are coded into separate bit streams. In the MPEG-4 terminology, temporal instances of video objects are referred to as video object planes (VOPs). Similar to key frames, key VOPs can be used for visual summarization of the video object content in an object-based framework. Shape matching is an important part of content-based visual data retrieval as how humans recognize objects primarily by their shape. Much effort has been taken for finding the shape features and similarity matching techniques that closely resemble human perception.

In the existing video object retrieval systems, shapes of objects are represented with one set of features, generally obtained by averaging the features that represent each temporal instant of video objects. While such a representation would work well for video objects that have constant shapes during their life spans, it is insufficient in the cases where the shapes of a video object changes. In this paper, the above described problem is overcome by a novel method for measuring shape similarities between two arbitrarily shaped video objects by comparing the low-level still shape features of the representative frames of the video objects. A video object sequence consists of video object planes (VOPs) that describe the shape and texture that the object has in a particular instant. As the shape of a video object could vary in the temporal dimension, the shape deformation's type that an object goes through could offer valuable information about the object's occlusion, local motion, articulated parts, and elasticity.

In these paper domain-independent descriptors for representing the local motion of arbitrarily shaped video objects is taken. These descriptors were based on the fact that any significant motion of video objects is very likely to result in changes in the object's shape and the variance of this change can potentially describe the type of object motion. As digital video is mostly available in compressed forms, most object-based video retrieval systems perform analysis in the uncompressed domain.

In this paper also the compressed domain retrieval of MPEG-4video objects has been addressed. The MPEG-4 compressed domain framework is employed mainly because of MPEG-4's support of an object-based representation, which helps to separate the segmentation problem from the retrieval problem. This method offers good retrieval performance and match closely with human ranking.

Three main sections in this paper are presented as follows.

- Shape matching of video objects is done by a similarity measure that is based on matching a subset of representative video object planes of a video object.
- Then extraction and matching of the new descriptors representing video object motion is presented. The shape of each temporal instant of a video object is represented with Fourier descriptor. These coefficients also change as the shape of video object changes with time.

Then another descriptor Angular Circular Local Motion descriptor is computed to capture the location as well as the type of object motion.

- In this final section descriptors are computed in the MPEG-4 compressed domain. A rough approximation of the video object contour by defining intracoded shape blocks as the contour point is obtained.

II. MATERIALS AND METHODOLOGY

The first step in retrieving the arbitrarily shaped video objects is collection of database. The database contains 50 MPEG-4 video object streams, including more than 20 arbitrarily shaped video objects. After collecting the database the video objects are matched based on its shape. Initially for matching the video objects by its shape a distance measure has been employed. The Euclidean distance is measured as,

$$d(VO_A,VO_B)=$$

$$1/N\sum_{k}^{N}\min_{VOP_b\in VO_B}\{d_{vop}(VOP_{ak},VOP_b)\}$$

N-number of representative VOPs of VO_A, VOPak, is the kth representative video object plane of VO_A, in the above distance measure for every VOP of $VO_{A,}$ the smallest distance to any representative VOP in VO_B is found. After then the sum of these distances is divided by the number of VOPs in VO_A to find the distance between two video objects.

Revised Version Manuscript Received on November 09, 2015.

Dr. Mohammed Jassim Mohammed Jassim, College of Science and Information Technology, Iraq.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

The Euclidian distance between two VOPs is calculated as,

$$d_{vop}(VOP_a, VOP_b) = \| \xrightarrow[R_a]{} - \xrightarrow[R_b]{} \|$$

Where, $\xrightarrow[R_a]{}$ *and* $\xrightarrow[R_b]{}$ are the feature vectors of the VOP$_a$ and VOP$_b$ respectively.

The above computed distance measure is asymmetric. Thus the final distance is calculated as,
D$_{VO}$ (VO$_A$, VO$_B$)=max {d$_{ov}$ (VO$_A$, VOB), dvo (VO$_B$, VO$_A$)}

As the shape of the video objects vary with time, their local motion is described using two main descriptors

1. Fourier Coefficient Based Local motion Descriptor
2. Angular Circular Local Motion Descriptor

Fourier Coefficient Based Local motion Descriptor finds the local motion based on the variances of the Fourier Coefficients. The variance vector is computed as follows, $\xrightarrow[F\sigma^2]{} = 1/K\sum_{k=0}^{K}(\xrightarrow[F_k]{} - \xrightarrow[F_\mu]{})^2$

and the mean Fourier vector is computed as, $\xrightarrow[F_\mu]{} = 1/K\sum_{k=0}^{K}\xrightarrow[F_k]{}$

Where K is the number of VOPs in a video object. $\xrightarrow[F_k]{}$ is the kth VOPs Fourier vector. $\xrightarrow[F_\mu]{}$ is the average of the feature vectors of all VOPs in a video object.

Every element of $\xrightarrow[F_\sigma 2]{}$ is divided by $\max\limits_{VOP_k \in VO}\{F_1(VOP_k)^2\}$, where F$_1$ (VOP$_k$) is the first non-zero Fourier coefficient of the kth VOP.The magnitudes of the complex coefficients are used as motion features. Angular Circular Local Motion Descriptor finds the location as well as the type of object motion. Information about the shape deformation present in a video object is offered by the variance of the object area. Based on this the shape mask of the video object is divide in to M angular and N circular segments and the variances of the pixels that fall into each segment describes the local motion. The variance is computed as follows,

$$\sigma^2{}_{n,m} = 1/S(n,m)^2 K\sum_{k=0}^{K-1}(P_{n,m} - \mu_{n,m})^2$$

where K is the number of the VOPs of the Video object, VOP$_k$ is the binary shape mask of the video object at that instant.

$$\mu_{n,m} = 1/K\sum_{k=0}^{K-1}P_{n,m} \quad ,$$

$$Pn,m = \sum_{\theta=\theta m}^{\theta m+1}\sum_{\rho=\rho n}^{\rho n+1} VOP_k(\rho,\theta)$$

VOP$_k$ (ρ , θ) is the binary shape mask value in VOPk at position (θ , ρ) at the center of VOP$_k$. S (n, m) is the area, θ_m is the start angle and ρ_n is the start radius of the angular circular segment (n, m). They are defined by, S (n, m)= $\Pi(\rho^2{}_{n+1}\text{-}\rho^2{}_n)$/M, where θ_m=m*2Π/M, and ρ_n=n*ρ_{max}/N, ρ_{max}= $\max\limits_{VOP_k \in VO}$ { ρ VOP$_k$} where M and N are the angular and circular sections, VOP$_k$ is the kth video object plane of the video object and ρ VOP$_k$ is the radius of the circle around VOP$_{k.}$ Then the local motion feature matrix R is formed by the variance $\sigma^2{}_{n,m}$ that fall in to the segment (n, m).

Then the descriptors in the MPEG-4 compressed domain are computed by shape coding method. The video object contour is obtained by defining the intracoded shape blocks as the contour points.

III. HOW IT WORKS

In this paper the system for retrieving the shape of video objects works as follows and it consists of the following step

1. Shape Matching of video objects
2. Local Motion Descriptors Based on Shape Deformations
3. Computation of the Descriptors in the MPEG-4 compressed domain.

The first step, shape matching of video objects is done by a similarity measure between the video objects. K-means clustering is performed before computing the similarity distance.

In the second step, two main descriptors are computed.

a. Fourier Coefficient Based Local motion Descriptor finds the local motion based on the variances of the Fourier Coefficients. The variance vector and the mean Fourier vector is computed.
b. Angular Circular Local Motion Descriptor finds the location as well as the type of object motion. The variance and mean vector is computed here.

Then in the final step, shape-coding method is used to compute the descriptors. This is done by first dividing the video object into 16*16 blocks then the blocks are transmitted that are inside the VOP as opaque, the blocks that are outside the VOP as transparent, and the boundary blocks, i.e., blocks that contain pixels both inside and outside the VOP, as intracoded with context-based arithmetic coding. Then a rough approximation of the video object contour by defining the intracoded shape blocks as the contour points is obtained. Defining the intra and opaque coded blocks as inside the shape approximates the object’s binary shape mask. Here, the object’s area is approximated finally using the formula as, A=O+0.5I where O is the number of opaque coded shape blocks and I is the number of intracoded shape blocks.

IV. COMPARISON WITH EXISTING SYSTEMS

In the existing video object retrieval systems, shapes of objects are represented with one set of features, generally obtained by averaging the features that represent each temporal instant of video objects. While such a representation would work well for video objects that have constant shapes during their lifespan, it is insufficient in the cases where the shape of a video objects changes. In this paper, this problem is overcome by using a novel method for measuring shape similarities between two arbitrarily shaped video objects by comparing the low-level still shape features of the representative frames of the video objects. As the shape of a video object could vary in the temporal dimension, the type of shape deformations that an object goes through could offer valuable information about the object's occlusion, local motion, articulated parts, and elasticity.

In a typical retrieval system, a feature vector is formed for each object and then the similarities between objects are found by computing the distance between these feature vectors. A video object sequence consists of video object planes (VOPs) that describe the shape and texture that the object has in a particular instant. In the existing systems, the feature vector of one key object plane is used to represent the whole video object sequence or by computing the average of the feature vectors belonging to all object planes. Such techniques work well for representing the color content of video objects, which tend to remain somewhat consistent during an object's lifespan. However, these techniques can fail to accurately represent an object's shape content, considering that the object's shape may change significantly during its existence. In this paper the above problem is overcome by using a similarity measure that is based on matching a subset of representative video object planes of a video object.

V. CONCLUSION

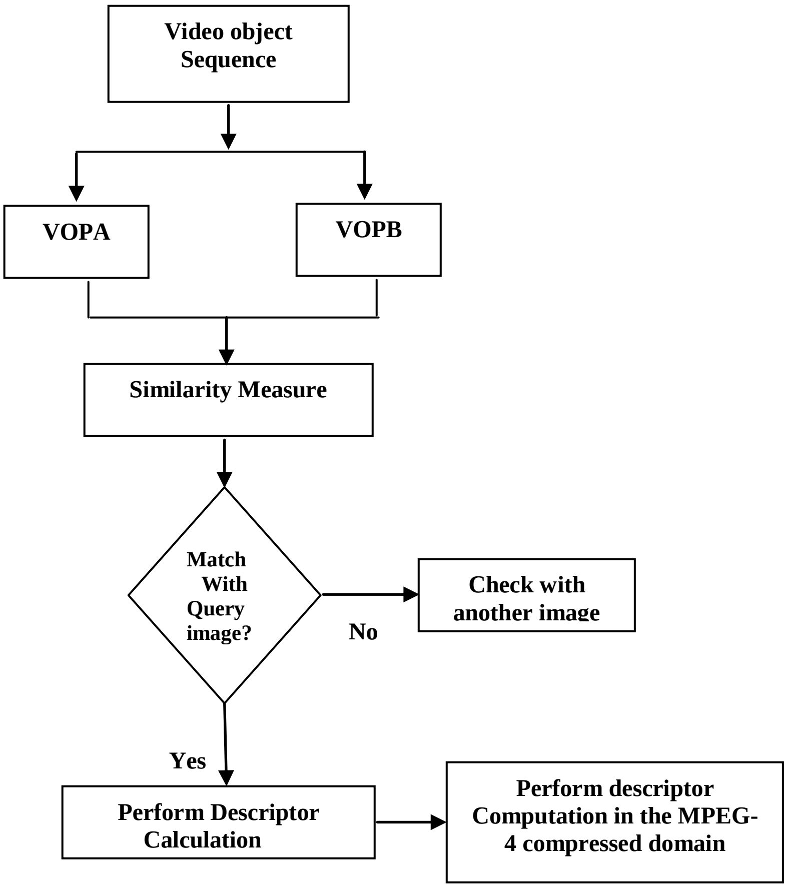
Matching of video objects based on their shape information is a very important component of any video object retrieval system. The possible variations in a video object's shape in time makes matching of objects using theirs shapes a challenging task Thus, in this paper shape matching via comparing a subset of representative planes of video objects is performed which outperformed the commonly used averaging technique. In this paper, the motion of objects based on variations of still shape descriptors is also described. The descriptors have been shown to successfully classify local object motion. The speed of the motion of these descriptors is not captured. Here shape and local motion descriptors are computed using the low-resolution shape data obtained from the MPEG-4 bit stream.

6.

REFERENCES

1. H.T.Nguyen, M. Worring, and A.Dev, “Detection of moving objects in video using a robust motion similarity measure,” IEEE Trans. Image Process., vol. 9, no.1, pp. 88–101,Jan. 2000.
2. MPEG-4 Video Group,” Coding of Audio-Visual Objects: Video,” SO/IEC 14 496-2, 2000.
3. S. F. Chang, W. Chen, H. J. Meng, H. Sundaram, and D. Zhong, “A Fully automated content-based video search engine supporting spatiotemporalqueries,” IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, pp.602–615, Sep. 1998.
4. Y. Deng and B. S.Manjunath, “NeTra-V: Toward an object-based video representation,” IEEE Trans. Circuits Syst. Video Technol., vol. 8, pp.616–627, Sep. 1998.
5. B. Erol and F.Kossentini, “Automatic key video object plane selection using the shape information in the MPEG-4 compressed domain,” IEEE Trans. Multimedia, vol. 2, no. 2, pp. 129–138, Jun. 2000.

FLOWCHART



Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication