

and accuracy of video comparison is to reasonably extract and compare features of video frames and determine similarities.

[0004] In order to improve the speed and accuracy of comparison, it is necessary to generate, for one video, a video feature that can characterize the video, so as to determine the similarity of the videos by comparing the video features of two videos. An extraction method of the video features and the excellence degree of the video features determine the efficiency and accuracy of the video comparison.

Summary

[0004a] According to a first aspect, the present disclosure refers to a video feature extraction method, comprising: extracting one or more frames from a video object to obtain one or more frames of images; stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes to obtain an image feature of the one or more frames of images; wherein the multi-typed pooling processes comprise a maximum pooling process, a minimum pooling process and an average pooling process; and determining a video feature according to the image feature of the one or more frames of images; wherein the stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes to obtain an image feature of the one or more frames of images comprises: determining a matrix according to the one or more frames of images, stage-by-stage generating a smaller matrix

by utilizing the multi-typed pooling processes until the matrix is minimized to be a matrix merely comprising one point, and determining the image feature according to the matrix merely comprising one point, with following steps: (a) according to one of the one or more frames of images, determining a first matrix having a first matrix dimensionality and a second matrix dimensionality; wherein points in the first matrix and pixels in the one of the one or more frames of images are corresponding; values of the points in the first matrix are first vectors, and the first vectors are three-dimensional vectors, configured for indicating brightness of three color channels of a corresponding pixel; (b) disposing a plurality of first blocks in the first matrix; wherein each of the plurality of first blocks comprises the first vectors; the number of the plurality of first blocks on the first matrix dimensionality is smaller than the number of the points of the first matrix comprised on the first matrix dimensionality, and the number of the plurality of first blocks on the second matrix dimensionality is smaller than the number of the points of the first matrix comprised on the second matrix dimensionality; with respect to each of the plurality of first blocks, respectively calculating a maximal value, a minimal value and an mean value of each dimensionality of the first vectors comprised in the each of the plurality of first blocks to obtain a nine-dimensional second vector; (c) according to the nine-dimensional second vector corresponding to the plurality of

first blocks, determining a second matrix; wherein points in the second matrix and the plurality of first blocks are corresponding, and a value of the points in the second matrix is the nine-dimensional second vector; and (d) repeating step (b) and step (c) until the first matrix is minimized to be a point whose value is a 3^N -dimensional vector, where N is a positive integer; and determining the 3^N -dimensional vector to be the image feature of the one or more frames of images.

[0004b] According to a second aspect, the present disclosure refers to a video feature database construction method, comprising: extracting a video feature of a video object according to any one of the video feature extraction method in the first aspect; and storing the video feature in a video feature database.

[0004c] According to a third aspect, the present disclosure refers to a video feature extraction device, comprising: a frame extractor, configured for extracting one or more frames from a video object to obtain one or more frames of images; an image feature determiner, configured for stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes to obtain an image feature of the one or more frames of images; wherein the multi-typed pooling processes comprise a maximum pooling process, a minimum pooling process and an average pooling process; a video feature determiner, configured for determining a video feature according to the image feature of the one or

more frames of images; and a module performing steps in the first aspect.

[0004d] According to a fourth aspect, the present disclosure refers to a video feature database construction device, comprising: a video feature extractor, configured for extracting a video feature of a video object according to any one of the video feature extraction method in the first aspect; a video feature memory, configured for storing the video feature into a video feature database; and the video feature database, configured for storing the video feature.

[0004e] According to a fifth aspect, the present disclosure refers to a video feature extraction hardware device, comprising: a memory, configured for storing non-transitory computer readable instructions; and a processor, configured for executing the non-transitory computer readable instructions such that the processor performs the video feature extraction method according to the first aspect.

[0004f] According to a sixth aspect, the present disclosure refers to a computer readable storage medium, configured for storing non-transitory computer readable instructions, when the non-transitory computer readable instructions are executed by a computer, causing the computer to perform the video feature extraction method according to the first aspect.

[0004g] According to a seventh aspect, the present disclosure refers to a terminal device, comprising the video feature extraction device according to the third aspect.

[0005] The object of one or more ~~embodiments~~examples of the disclosure is to provide a new video feature extraction method and device.

[0006] The purpose of one or more ~~embodiments~~examples of the disclosure is achieved by the following technical solutions. The video feature extraction method according to one or more ~~embodiments~~examples of the disclosure includes the following steps: extracting one or more frames from a video object to obtain one or more frames of images; stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes to obtain an image feature of the one or more frames of images, and determining a video feature according to the image feature of the one or more frames of images. The multi-typed pooling processes include a maximum pooling process, a minimum pooling process and an average pooling process.

[0007] The object of one or more ~~embodiments~~examples of the disclosure can also be further achieved by the following technical scheme.

[0008] In the foregoing video feature extraction method, stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes includes stage-by-stage processing each of

is smaller than the number of the points of the first matrix contained on the first matrix dimensionality, and the number of the first blocks on the second matrix dimensionality is smaller than the number of the points of the first matrix contained on the second matrix dimensionality; (c) according to the nine-dimensional second vector corresponding to the first blocks, determining a second matrix. Points in the second matrix and the first blocks are corresponding, and a value of the points in the second matrix is the nine-dimensional second vector; (d) repeating step (b) and step (c) until the first matrix is minimized to be a point whose value is a 3^N -dimensional vector, where N is a positive integer; and determining the 3^N -dimensional vector to be the image feature of the one or more frames of images.

[0011] In the foregoing video feature extraction method, determining a video feature according to the image feature of the one or more frames of images contains processing the image feature by binarization to obtain a binarization image feature; and according to the binarization image feature of the one or more frames of images, determining the video feature.

[0012] In the foregoing video feature extraction method, processing the image feature by binarization to obtain a binarization image feature contains following steps: generating groups according to the image feature; respectively adding up the elements in each of the groups to obtain a total number of each of the groups; pairing the groups to obtain group pairs; with respect to each of the plurality of groups, comparing total numbers of two groups in each of the group pairs, and generating a binarization image feature bit according to a comparison result; and according to binarization image feature bits of the group pairs, determining the binarization image feature of the one or more frames of images. Each of the groups includes elements in the image feature.

Claims

WHAT IS CLAIMED IS:

1. A video feature extraction method, comprising:

extracting one or more frames from a video object to obtain one or more frames of images;

stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes to obtain an image feature of the one or more frames of images; wherein the multi-typed pooling processes comprise a maximum pooling process, a minimum pooling process and an average pooling process; and

determining a video feature according to the image feature of the one or more frames of images;

wherein the stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes to obtain an image feature of the one or more frames of images comprises:

determining a matrix according to the one or more frames of images, stage-by-stage generating a smaller matrix by utilizing the multi-typed pooling processes until the matrix is minimized to be a matrix merely comprising one point, and determining the image feature according to the matrix merely comprising one point, with following steps:

(a) according to one of the one or more frames of images, determining a first matrix having a first matrix dimensionality and a second matrix dimensionality; wherein points in the first matrix and pixels in the one of the one or more frames of images are corresponding; values of the points in the first matrix are first vectors, and the first vectors are three-dimensional vectors, configured for indicating brightness of three color channels of a corresponding pixel;

(b) disposing a plurality of first blocks in the first matrix; wherein each of the plurality of first blocks comprises the first vectors; the number of the plurality of first blocks on the first matrix dimensionality is smaller than the number of the points of the first matrix comprised on the first matrix dimensionality, and the number of the plurality of first blocks on the second matrix dimensionality is smaller than the number of the points of the first matrix comprised on the second matrix dimensionality; with respect to each of the plurality of first blocks, respectively calculating a maximal value, a minimal value and an mean value of each dimensionality of the first vectors comprised in the each of the plurality of first blocks to obtain a nine-dimensional second vector;

(c) according to the nine-dimensional second vector corresponding to the plurality of first blocks, determining a second matrix; wherein points in the second matrix and the plurality of first blocks are corresponding, and a value of the points in the second matrix is the nine-dimensional second vector; and

(d) repeating step (b) and step (c) until the first matrix is minimized to be a point whose value is a 3^N -dimensional vector, where N is a positive integer; and determining the 3^N -dimensional vector to be the image feature of the one or more frames of images.

2. The video feature extraction method according to claim 1, wherein the stage-by-stage processing each of the one or more frames of images by multi-typed pooling processes comprises:

stage-by-stage processing each of the one or more frames of images by the multi-typed pooling processes based on a plurality of color channels of the one or more frames of images.

~~3. The video feature extraction method according to claim 1, wherein the stage by stage processing each of the one or more frames of~~

images by multi-typed pooling processes to obtain an image feature of the one or more frames of images comprises:

~~determining a matrix according to the one or more frames of images, stage by stage generating a smaller matrix by utilizing the multi-typed pooling processes until the matrix is minimized to be a matrix merely comprising one point, and determining the image feature according to the matrix merely comprising one point.~~

4. ~~The video feature extraction method according to claim 3, wherein the determining a matrix according to the one or more frames of images, stage by stage generating a smaller matrix by utilizing the multi-typed pooling processes until the matrix is minimized to be a matrix merely comprising one point, and determining the image feature according to the matrix merely comprising one point comprises following steps:—~~

~~(a) according to one of the one or more frames of images, determining a first matrix having a first matrix dimensionality and a second matrix dimensionality; wherein points in the first matrix and pixels in the one of the one or more frames of images are corresponding; values of the points in the first matrix are first vectors, and the first vectors are three-dimensional vectors, configured for indicating brightness of three color channels of a corresponding pixel;—~~

~~(b) disposing a plurality of first blocks in the first matrix; wherein each of the plurality of first blocks comprises the first vectors; the number of the plurality of first blocks on the first matrix dimensionality is smaller than the number of the points of the first matrix comprised on the first matrix dimensionality, and the number of the plurality of first blocks on the second matrix dimensionality is smaller than the number of the points of the first matrix comprised on the second matrix dimensionality; with respect to each of the plurality of first blocks, respectively calculating a~~

~~maximal value, a minimal value and an mean value of each dimensionality of the first vectors comprised in the each of the plurality of first blocks to obtain a nine dimensional second vector;~~

~~(c) according to the nine dimensional second vector corresponding to the plurality of first blocks, determining a second matrix; wherein points in the second matrix and the plurality of first blocks are corresponding, and a value of the points in the second matrix is the nine dimensional second vector; and~~

~~(d) repeating step (b) and step (c) until the first matrix is minimized to be a point whose value is a 3^N -dimensional vector, where N is a positive integer; and determining the 3^N -dimensional vector to be the image feature of the one or more frames of images.~~

53. The video feature extraction method according to claim 1, wherein the determining a video feature according to the image feature of the one or more frames of images comprises:

processing the image feature by binarization to obtain a binarization image feature; and

according to the binarization image feature of the one or more frames of images, determining the video feature.

64. The video feature extraction method according to claim 53, wherein the processing the image feature by binarization to obtain a binarization image feature comprises following steps:

generating a plurality of groups according to the image feature; wherein each of the plurality of groups comprises a plurality of elements in the image feature;

respectively adding up the plurality of elements in each of the plurality of groups to obtain a total number value of each of the plurality of groups;

pairing the plurality of groups to obtain a plurality of group pairs,
wherein every two adjacent groups of the plurality of groups are paired to
obtain a group pair of the plurality of group pairs;

with respect to each of the plurality of groups, comparing total
values ~~numbers~~ of two groups in each of the plurality of group pairs, and
generating a binarization image feature bit according to a comparison
result; and

according to binarization image feature bits of the plurality of
group pairs, determining the binarization image feature of the one or
more frames of images.

~~75.~~ A video feature database construction method, comprising:

extracting a video feature of a video object according to any one of
the video feature extraction method in claims 1-~~64~~; and

storing the video feature in a video feature database.

~~86.~~ A video feature extraction device, comprising:

a frame extractor, configured for extracting one or more frames
from a video object to obtain one or more frames of images;

an image feature determiner, configured for stage-by-stage
processing each of the one or more frames of images by multi-typed
pooling processes to obtain an image feature of the one or more frames of
images; wherein the multi-typed pooling processes comprise a maximum
pooling process, a minimum pooling process and an average pooling
process; ~~and~~

a video feature determiner, configured for determining a video
feature according to the image feature of the one or more frames of
images; and;

~~9. The video feature extraction device according to claim 8,
wherein the video feature extraction device further comprises a module
performing steps in any one of claims 2-~~64~~.~~

~~107~~. A video feature database construction device, comprising:

a video feature extractor, configured for extracting a video feature of a video object according to any one of the video feature extraction method in claims 1-~~64~~;

a video feature memory, configured for storing the video feature into a video feature database; and

the video feature database, configured for storing the video feature.

~~118~~. A video feature extraction hardware device, comprising:

a memory, configured for storing non-transitory computer readable instructions; and

a processor, configured for executing the non-transitory computer readable instructions such that the processor performs any one of the video feature extraction methods according to claims 1-~~64~~.

~~129~~. A computer readable storage medium, configured for storing non-transitory computer readable instructions, when the non-transitory computer readable instructions are executed by a computer, causing the computer to perform any of the video feature extraction methods according to claims 1-~~64~~.

~~1310~~. A terminal device, comprising ~~any of~~ the video feature extraction devices according to claims ~~8-6 or 9~~.