

CLAIMS1. A system for segmenting videos, comprising:

at least one processing component;

at least one memory component;

a video;

an extraction component configured to:

extract image data and text data from the video;

identify at least two entities in the image data;

assign at least one entity relation to the at least two entities in the image data;

identify at least two entities in the text data; and

assign at least one entity relation to the two or more entities in the text data; and

a graphing component configured to:

generate an image knowledge graph for the at least one entity relation assigned to the at least two entities in the image data;

generate a text knowledge graph for the at least one entity relation assigned to the at least two entities in the text data; and

generate a weighted knowledge graph based on the image knowledge graph and the text knowledge graph.

2. The system of claim 1, wherein the weighted knowledge graph includes relation weights for the at least one entity relation assigned to the at least two entities in the image data and the at least one entity relation assigned to the at least two entities in the text data.

3. The system of claim 2, further comprising a grouping component configured to:

identify a top relation in the at least one entity relation assigned to the at least two entities in the image data and the at least one entity relation assigned to the at least two entities in the text data, wherein the top relation is an entity relation having a relation weight greater than a threshold relation weight;

select frames of the video that correspond to the top relation; and

group the frames into a video segment.

4. The system of claim 3, wherein the grouping component is further configured to:

determine that there are remaining frames of the video that do not include the top relation;

determine that the frames in the video segment are nearest to the remaining frames; and
group the remaining frames with the video segment.

5. The system of claim 1, wherein the video is divided into pictures, wherein each picture includes a set of frames.

6. The system of claim 1, wherein the text data is captions.

7. The system of claim 1, wherein the text data is extracted from speech data.

8. The system of claim 1, wherein the at least two entities in the image data are identified based on facial recognition.

9.

~~4.~~ 9. A method, comprising:
 receiving a video;
 extracting image data and text data from the video;
 identifying at least two entities in the image data;
 assigning at least one entity relation to the at least two entities in the image data;
 identifying at least two entities in the text data;
 assigning at least one entity relation to the at least two entities in the text data;
 generating an image knowledge graph for the at least one entity relation assigned to the at least two entities in the image data;
 generating a text knowledge graph for the at least one entity relation assigned to the at least two entities in the text data; and
 generating a weighted knowledge graph based on the image knowledge graph and the text knowledge graph.

~~2.~~—10. The method of claim ~~4~~9, wherein the weighted knowledge graph includes relation weights for the at least one entity relation assigned to the at least two entities in the image data and the at least one entity relation assigned to the at least two entities in the text data.

3. ~~11.~~ The method of claim ~~210~~, further comprising:

identifying a top relation in the at least one entity relation assigned to the at least two entities in the image data and the at least one entity relation assigned to the at least two entities in the text data, wherein the top relation is an entity relation having a relation weight greater than a threshold relation weight;

selecting frames of the video that correspond to the top relation; and
grouping the frames into a video segment.

4. ~~12.~~ The method of claim ~~311~~, further comprising:

determining that there are remaining frames of the video that do not include the top relation;
determining that the frames in the video segment are nearest to the remaining frames; and
grouping the remaining frames with the video segment.

5. ~~13.~~ The method of claim ~~49~~, wherein the video is divided into pictures, wherein each picture includes a set of frames.

6. ~~14.~~ The method of claim ~~49~~, wherein the text data is captions.

7. ~~15.~~ The method of claim ~~49~~, wherein the at least two entities in the image data are identified based on facial recognition.

~~8. A system comprising means adapted for carrying out all the steps of the method according to any preceding method claim.~~ 16. A computer program product for segmenting videos, the computer program product comprising a computer readable storage medium having program instructions embodied therewith, the program instructions executable by a processor to cause a device to perform a method, the method comprising:

receiving a video;

extracting image data and text data from the video;

identifying at least two entities in the image data;

assigning at least one entity relation to the at least two entities in the image data;

identifying at least two entities in the text data;

assigning at least one entity relation to the at least two entities in the text data;

generating an image knowledge graph for the at least one entity relation assigned to the at least two entities in the image data;

generating a text knowledge graph for the at least one entity relation assigned to the at least two entities in the text data; and

generating a weighted knowledge graph based on the image knowledge graph and the text knowledge graph.

17. The computer program product of claim 16, wherein the weighted knowledge graph includes relation weights for the at least one entity relation assigned to the at least two entities in the image data and the at least one entity relation assigned to the at least two entities in the text data.

18. The computer program product of claim 17, further comprising:

identifying a top relation in the at least one entity relation assigned to the at least two entities in the image data and the at least one entity relation assigned to the at least two entities in the text data, wherein the top relation is an entity relation having a relation weight greater than a threshold relation weight;

selecting frames of the video that correspond to the top relation; and

grouping the frames into a video segment.

19. The computer program product of claim 18, further comprising:

determining that there are remaining frames of the video that do not include the top relation;

determining that the frames in the video segment are nearest to the remaining frames; and

grouping the remaining frames with the video segment.

20. The computer program product of claim 16, wherein the at least two entities in the image data are identified based on facial recognition.

~~9. — A computer program comprising instructions for carrying out all the steps of the method according to any preceding method claim, when said computer program is executed on a computer system.~~