[CLAIMS]

1. A face detection method, which is performed by a server associated with a user terminal, comprising:

receiving video data and voice data from the user terminal;

deriving a first interval related to a predetermined message based on the received voice data;

configuring a second interval based on the derived first interval, wherein the second interval is located after the first interval in time series within the voice data;

extracting a part of the video data corresponding to the second interval;

deriving a video frame which satisfies a predetermined criterion from the extracted video data based on the second interval; and

detecting a facial image included in each of the derived video frame based on the second interval.

2. The face detection method of claim 1, wherein the deriving the first interval comprises:

generating a spectrogram by converting the voice data into frequency domain in a predetermined time unit;

generating a frequency pattern of the voice data including the predetermined message; and

selecting an interval having a highest similarity to the frequency pattern in the spectrogram as the first interval.

3. The face detection method of claim 2, wherein the generating the spectrogram comprises:

generating a first spectrum by converting first voice data corresponding to a first window configured in the predetermined time unit into frequency domain;

generating a second spectrum by converting second voice data corresponding to a second window, which is different from the first window and configured in the predetermined time unit, into frequency domain; and

generating the spectrogram by merging the first spectrum and the second spectrum.

4. The face detection method of claim 3, wherein the first window and the second window partially overlap each other on time domain of the voice data.

5. The face detection method of claim 1, wherein the deriving the first interval comprises:

sampling the voice data in intervals of a predetermined time unit;

generating a voice pattern including a predetermined message;

extracting voice similarity for each interval based on the sampled voice data for each interval and the voice pattern using a deep learning module; and

selecting an interval in which the voice similarity is higher than a predetermined threshold as the first interval.

6. The face detection method of claim 5, wherein the deep learning module comprises:

an input layer including the sampled voice data for each interval and the voice pattern as input nodes;

an output layer including the voice similarity as an output node; and

one or more hidden layers disposed between the input layer and the output layer,

wherein weights of nodes and edges between the input node and the output node are updated by learning processes of the deep learning module.

7. The face detection method of claim 1, wherein the predetermined message related to the first interval comprises a first message,

wherein an end point of the second interval is set to a point where a predetermined second message different from the first message is outputted.

8. The face detection method of claim 1, wherein a part of the second interval overlaps a part of the first interval.

9. The face detection method of claim 1, wherein the deriving the at least one video frame comprises deriving one or more frames for the second interval by using a predetermined period, or deriving a frame in which an optical flow is smaller than a threshold in the second interval.

10. The face detection method of claim 1, wherein the detecting the facial image comprises:

deriving facial landmarks for each of the at least one derived video frame;

performing correction for facial alignment based on the derived facial landmarks; and

extracting feature points from the corrected facial image.

11. A face detection method, which is performed by a server associated with a user terminal, comprising:

receiving video data and voice data from the user terminal;

deriving an interval related to a predetermined first message based on the received voice data;

extracting a part of the video data within a predetermined range based on the derived interval;

deriving a video frame which satisfies a predetermined criterion from the extracted video data; and

detecting a facial image included in the derived video frame,

wherein the extracting a part of the video data comprises extracting the part of the video data from video data existing from an end point of the interval related to the first message to before a point where a second message different from the first message is outputted.

12. The face detection method of claim 11, wherein the deriving the interval comprises:

generating a spectrogram by converting the voice data into frequency domain in a predetermined time unit;

generating a frequency pattern of the voice data including the predetermined message; and

selecting an interval having a highest similarity to the frequency pattern in the spectrogram as the interval.

13. The face detection method of claim 11, wherein the deriving the interval comprises:

sampling the voice data in intervals of a predetermined time unit;

generating a voice pattern including a predetermined message;

extracting voice similarity for each interval based on the sampled voice data for each interval and the voice pattern using a deep learning module; and

selecting an interval in which the voice similarity is higher than a predetermined threshold as the interval.

FACE DETECTION METHODS USING VOICE

[ABSTRACT]

Face detection methods using voice are provided. The face detection method comprises the: receiving video data and voice data from a user terminal; deriving a first interval related to a predetermined message based on the received voice data; configuring a second interval based on the derived first interval; extracting a part of the video data corresponding to the second interval; deriving at least one video frame which satisfies a predetermined criterion from the extracted video data; and detecting a facial image included in each of the at least one derived video frame.

Figure 2 accompanies.