CLAIMS

1. A video call device comprising:

a communications unit for supporting a video call service between a plurality of user terminals through a communication network;

an extraction unit for generating an image file and an audio file using a video call-related video file collected from each of the plurality of user terminals, and extracting original language information for each user using the image file and the audio file;

a translation unit for generating translation information from the original language information; and

a control unit for controlling transmission of an interpreted/translated video, in which the extracted original language information and the translation information are mapped to the video call-related video file:

wherein the original language information includes voice original language information and text original language information, and the translation information includes voice translation information and text translation information;

wherein the extraction unit extracts voice original language information for each caller by applying a frequency band analysis process to the audio file, maps the extracted voice original language information to specific character information and stores the mapped information, wherein the mapping is performed by identifying a user terminal that transmitted a specific voice and mapping a preset ID for the corresponding user terminal or a nickname preset by a user to the voice original language information;

wherein the extraction unit generates text original language information by further applying a voice recognition process to the extracted voice original language information;

wherein the extraction unit determines whether a sign language pattern exists in the image file by applying an image processing process to the image file, and generates text original language information based on a detected sign language pattern if the sign language pattern exists; and

wherein the translation unit generates voice translation information with a voice most similar to a speaker's voice among preset voices based on characteristics for each voice analyzed by applying the frequency band analysis process to the audio file by the extraction unit, wherein the characteristics include a gender, an age group, and a pronunciation tone of a voice.

2. A control method of a video call device, the method comprising the steps of:

receiving a video call-related video file from a plurality of user terminals through a communication network;

extracting original language information for each caller using an image file and an audio file generated from the video call-related video file, wherein the original language information includes voice original language information and text original language information;

generating translation information of the original language information translated according to a language of a selected country, wherein the translation information includes voice translation information and text translation information; and

controlling to transmit an interpreted/translated video, in which the original language information and the translation information are mapped to the video call-related video file:

wherein the extracting step includes:

applying a frequency band analysis process to the audio file to extract voice original language information for each caller, mapping the extracted voice original language information to specific character information and storing the

mapped information, wherein the mapping is performed by identifying a user terminal that transmitted a specific voice and mapping a preset ID for the corresponding user terminal or a nickname preset by a user to the voice original language information;

applying a voice recognition process to the extracted voice original language information to generate text original language information; and

applying an image processing process to the image file to determine whether a sign language pattern exists in the image file, and generating text original language information based on a detected sign language pattern if the sign language pattern exists; and

wherein the generating translation information step includes generating voice translation information with a voice most similar to a speaker's voice among preset voices based on characteristics for each voice analyzed by applying the frequency band analysis process to the audio file in the extracting step, wherein the characteristics include a gender, an age group, and a pronunciation tone of a voice.