CLAIMS

[Claim 1]

A monitoring device comprising:

a voice acquisition means for acquiring a predetermined voice uttered by a person due to the occurrence of an abnormal situation in a monitoring target area;

a person identification means for identifying the person who uttered the predetermined voice based on features obtained from the predetermined voice;

an analysis means for searching for the identified person from a video from a camera shooting the monitoring target area and analyzing a facial expression or a motion of the person; and

an abnormal situation evaluation means for evaluating the abnormal situation in the monitoring target area based on a result of the analysis.

[Claim 2]

The monitoring device according to claim 1, wherein the analysis means analyzes whether or not the facial expression of the person is a predetermined facial expression.

[Claim 3]

The monitoring device according to claim 1 or 2, wherein the analysis means analyzes whether or not the motion of the person is similar to a predefined series of motions.

[Claim 4]

The monitoring device according to any one of claims 1 to 3, wherein the analysis means analyzes whether or not the motion of the person is similar to a predefined gesture.

[Claim 5]

The monitoring device according to any one of claims 1 to 4, wherein analysis processing by the analysis means is executed when the predetermined voice is detected, and is not executed before the predetermined voice is detected.

[Claim 6]

The monitoring device according to any one of claims 1 to 5, further comprising a sound source position estimating means for estimating a sound source position of the predetermined voice,

wherein the analysis means performs analysis processing only on video data from the camera shooting the area including the sound source position, among a plurality of the cameras.

[Claim 7]

The monitoring device according to claim 6, wherein the analysis means performs analysis processing only for the partial image including the sound source position in the image constituting the video.

[Claim 8]

The monitoring device according to any one of claims 1 to 7, further comprising a signal output means for outputting a predetermined signal when the evaluation of the abnormal situation satisfies a predetermined criterion.

~~[Claim 9]~~

~~A monitoring system comprising:~~

~~a camera configured to shoot a monitoring target area;~~

~~a sensor configured to detect a sound generated in the monitoring target area; and~~

~~a monitoring device,~~

~~wherein the monitoring device includes:~~

~~a voice acquisition means for acquiring, from the sensor, a predetermined voice uttered by a person due to the occurrence of an abnormal situation in the monitoring target area;~~

~~a person identification means for identifying the person who uttered the predetermined voice based on features obtained from the predetermined voice;~~

~~an analysis means for searching for the identified person from a video from the camera and analyzing a facial expression or a motion of the person; and~~

~~an abnormal situation evaluation means for evaluating the abnormal situation in the monitoring target area based on a result of the analysis.~~

~~[Claim 10]~~

~~The monitoring system according to claim 9, wherein the analysis means analyzes whether or not the facial expression of the person is a predetermined facial expression.~~

~~[Claim 11]~~

~~The monitoring system according to claim 9 or 10, wherein the analysis means analyzes whether or not the motion of the person is similar to a predefined~~

~~series of motions.~~

~~[Claim 12]~~

~~The monitoring system according to any one of claims 9 to 11, wherein the analysis means analyzes whether or not the motion of the person is similar to a predefined gesture.~~

[Claim ~~13~~9]

A monitoring method comprising:

acquiring a predetermined voice uttered by a person due to the occurrence of an abnormal situation in a monitoring target area;

identifying the person who uttered the predetermined voice based on features obtained from the predetermined voice;

searching for the identified person from a video from a camera shooting the monitoring target area and analyzing a facial expression or a motion of the person; and

evaluating the abnormal situation in the monitoring target area based on a result of the analysis.

[Claim ~~14~~10]

A non-transitory computer-readable medium storing a program causing a computer to execute:

a voice acquisition step of acquiring a predetermined voice uttered by a person due to the occurrence of an abnormal situation in a monitoring target area;

a person identification step of identifying the person who uttered the predetermined voice based on features obtained from the predetermined voice;

an analysis step of searching for the identified person from a video from a camera shooting the monitoring target area and analyzing a facial expression or a motion of the person; and

an abnormal situation evaluation step of evaluating the abnormal situation in the monitoring target area based on a result of the analysis.