What is claimed is:

1. An activity detection device in an actively detection system comprising:

an input unit that inputs an image sequence including at least a first image and a second image;

an object detection unit that detects a first object in the image sequence;

a component model unit that generates first characteristic information characterizing the first object and includes a plurality of individually trainable component models, wherein training of each individually trainable component model in the plurality of individually trainable component models is individually performed using training data of different domains; and

an activity detection unit that generates a first object state corresponding to the first object in the first image and a second object state corresponding to the first object in the second image based on the first characteristic information, determines an activity related to the first object based on the first object state and the second object state, and outputs a first activity detection result indicating the determined activity,

wherein the plurality of individually trainable component models includes at least two individually trainable component models selected from an optical flow

model, an object tracking model, an object re-identification model, a neural network feature model, an object attribute model, and an object skeleton model, and

a number and a type of the at least two individually trainable component models are selected according to a purpose of use of the activity detection system,

wherein the activity detection unit generates the first object state and the second object state in a Markov model, and determines the activity related to the first object based on a relationship between the first object state and the second object state in the Markov model,

wherein the Markov model is trained by using outputs of the plurality of individually trainable component models.

2. The activity detection device according to claim 1,

wherein the object detection unit detects a second object in the image sequence,

wherein the component model unit determines second characteristic information characterizing the second object, and

wherein the activity detection unit generates a third object state corresponding to the second object in the first image based on the second characteristic information, determines an activity related to the first object based on the first object state, the second object state, and the third object state, and outputs a second activity detection

result indicating the determined activity.

3. The activity detection device according to claim 1,

wherein each individually trainable component model in the plurality of individually trainable component models generates information related to a generation process of the first characteristic information as an explanation notification, and

wherein the activity detection unit outputs the explanation notification together with the first activity detection result.

4. The activity detection device according to claim 3,

wherein the first object state and the second object state show a characteristic characterizing the first object in a semantic binary format.

5. The activity detection device according to claim 1,

wherein the plurality of individually trainable component models includes a convolutional neural network and a skeleton detection model,

wherein the first characteristic information generated includes a feature quantity of the first object generated by the convolutional neural network and skeleton information of the first object generated by the skeleton detection model, and

wherein the activity detection unit generates the first object state corresponding to the first object in the

first image and the second object state corresponding to the first object in the second image based on the feature quantity of the first object and the skeleton information of the first object as the first characteristic information.

6. The activity detection device according to claim 1,

wherein the activity detection unit determines the activity related to the first object based on a transition from the first object state to the second object state in the Markov model.

7. The activity detection device according to claim 6,

wherein the activity detection unit derives an object state corresponding to an event not shown in the image sequence based on a causal relationship between the first object state and the second object state in the Markov model.

8. The activity detection device according to claim 7,

wherein the activity detection unit outputs the relationship between the first object state and the second object state used in the determination of the activity related to the first object as the explanation notification, as a reason for supporting the detection result.

9. An activity detection system in which a client terminal and an activity detection device are connected to each other via a communication network,

wherein the activity detection device comprises:

an input unit that inputs an image sequence including

at least a first image and a second image from the client terminal;

an object detection unit that detects a first object in the image sequence;

a component model unit that generates first characteristic information characterizing the first object and includes a plurality of individually trainable component models, wherein training of each individually trainable component model in the plurality of individually trainable component models is individually performed using training data of different domains; and

an activity detection unit that generates a first object state corresponding to the first object in the first image and a second object state corresponding to the first object in the second image based on the first characteristic information, determines an activity related to the first object based on the first object state and the second object state, and outputs an activity detection result indicating the determined activity to the client terminal,

wherein the plurality of individually trainable component models includes at least two individually trainable component models selected from an optical flow model, an object tracking model, an object re-identification model, a neural network feature model, an object attribute model, and an object skeleton model, and

a number and a type of the at least two individually trainable component models are selected according to a purpose of use of the activity detection system,

wherein the activity detection unit generates the first object state and the second object state in a Markov model, and determines the activity related to the first object based on a relationship between the first object state and the second object state in the Markov model,

wherein the Markov model is trained by using outputs of the plurality of individually trainable component models.

10. The activity detection system according to claim 9,

wherein the object detection unit detects a second object in the image sequence, and

wherein the component model unit determines second characteristic information characterizing the second object, and

wherein the activity detection unit generates a third object state corresponding to the second object in the first image based on the second characteristic information, and determines an activity detection result related to the first object based on the first object state, the second object state, and the third object state.

11. An activity detection method executed by an activity detection device in an activity detection system comprising:

a step of inputting an image sequence including at

least a first image and a second image;

a step of detecting a first object in the image sequence;

a step of generating first characteristic information characterizing the first object by using a plurality of individually trainable component models, wherein training of each individually trainable component model in the plurality of individually trainable component models is individually performed using training data of different domains; and

a step of generating a first object state corresponding to the first object in the first image and a second object state corresponding to the first object in the second image based on the first characteristic information, determining an activity related to the first object based on the first object state and the second object state, and outputting an activity detection result indicating the determined activity,

wherein the plurality of individually trainable component models includes at least two individually trainable component models selected from an optical flow model, an object tracking model, an object re-identification model, a neural network feature model, an object attribute model, and an object skeleton model, and

a number and a type of the at least two individually trainable component models are selected according to a purpose of use of the activity detection system,

wherein the activity detection method further comprises a step of generating the first object state and the second object state in a Markov model, and a step of determining the activity related to the first object based on a relationship between the first object state and the second object state in the Markov model,

wherein the Markov model is trained by using outputs of the plurality of individually trainable component models.

12. The activity detection method according to claim 11, further comprising:

a step of detecting a second object in the image sequence;

a step of generating second characteristic information characterizing the second object; and

a step of generating a third object state corresponding to the second object in the first image based on the second characteristic information, determining an activity related to the first object based on the first object state, the second object state, and the third object state, and outputting a second activity detection result indicating the determined activity.

13. The activity detection method according to claim 11,

wherein the step of generating first characteristic information characterizing the first object by using the plurality of individually trainable component models further

comprises a step of generating information related to a generation process of the first characteristic information as an explanation notification.