CLAIMS:

1. An active audio-visual surveillance system for identifying events occurring at a location-of-interest, the system comprising:

a Fisheye image capturing device provided at the location-of-interest, the device configured to continuously capture wide-angle image frames and convert the captured wide-angle image frames into non-distorted perspective images using a set of calibration parameters generated by a fisheye correction scheme, and store the non-distorted perspective images;

an audio module provided at the location-of-interest, the module configured to capture audio signals and to determine a Direction of Arrival (DOA) of the audio signals;

a computing module communicatively coupled to the audio module and the Fisheye image capturing device, the computing module being configured to:

receive a notification from the audio module when the audio module detects an event having an audio signal at the location-of-interest, whereby the notification includes data of a DOA of the audio signal;

instruct the Fisheye image capturing device to generate, using a trained single-image-super-resolution (SISR) scheme, high-resolution perspective images based on the stored non-distorted perspective images associated with the DOA of the audio signal,

wherein the detecting the event having an audio signal at the location-of-interest by the audio module comprises the audio module being configured to:

generate a discrete set of DOA points that cover a DOA region from which the audio signal of the event was detected; and

embed the discrete set of DOA points in the notification that is sent to the computing module, and

wherein the generating the high-resolution perspective images based on the stored non-distorted perspective images by the Fisheye image capturing device comprises the Fisheye image capturing device being configured to:

extract the discrete set of DOA points received from the computing module, wherein the discrete set of DOA points are provided in the instructions sent by the computing module;

generate a unified region of interest from the stored non-distorted perspective images that correspond to the extracted set of DOA points; and

generate, using the trained single-image-super-resolution (SISR) scheme, high-resolution perspective images of the unified region of interest.

2. The system according to claim 1 wherein the fisheye correction scheme comprises a fisheye rectilinear transformation scheme.

3. The system according to claim 1 wherein the audio module comprises:

a plurality of sensor arrays configured to utilize a beamforming technique to estimate a direction of arrival of captured audio signals.

~~4. The system according to claim 1 wherein the detecting the event having an audio signal at the location-of-interest by the audio module comprises:~~

~~the audio module being configured to:~~

~~generate a discrete set of DOA points that cover a DOA region from which the audio signal of the event was detected; and~~

~~embed the discrete set of DOA points in the notification that is sent to the computing module.~~

~~5. The system according to claim 4 wherein the generating the high-resolution perspective images based on the non-distorted perspective images by the Fisheye image capturing device comprises:~~

~~the Fisheye image capturing device being configured to:~~

~~extract the discrete set of DOA points received from the computing module, wherein the discrete set of DOA points are provided in the instructions sent by the computing module;~~

~~generate a unified region of interest of the non-distorted perspective images that corresponds to the extracted set of DOA points; and~~

~~generate, using the trained single-image-super-resolution (SISR) scheme, high-resolution perspective images of the unified region of interest.~~

~~6.~~4. The system according to claim ~~5~~1 wherein the generating the unified region of interest of the captured wide-angle image frames comprises:

the Fisheye image capturing device being configured to:

for every DOA point, extract a corresponding θ±36 degree region of interest from the non-distorted perspective images; and

compile all the extracted θ±36 degree regions of interest to form the unified region of interest.

~~7.~~5. The system according to claim 1 further comprising:

another Fisheye image capturing device provided at the location-of-interest, the device configured to continuously capture wide-angle image frames and convert the captured wide-angle image frames into non-distorted perspective images using another set of calibration parameters generated by a fisheye correction scheme;

another audio module provided at the location-of-interest, the module configured to capture audio signals and to determine a Direction of Arrival (DOA) of the audio signals;

the computing module communicatively coupled to the another audio module and the another Fisheye image capturing device, the computing module being configured to:

receive a notification from the another audio module when the another audio module detects the event having the audio signal at the location-of-interest, whereby the notification includes data of another DOA of the audio signal;

instruct the another Fisheye image capturing device to generate, using the trained single-image-super-resolution (SISR) scheme, another high-resolution perspective images based on the captured wide-angle image frames associated with the another DOA of the audio signal;

an event identification module configured to:

receive the high-resolution perspective images, and the another high-resolution perspective images from the computing module; and

stitch the received images to form a single set of high-resolution perspective images.

~~8.~~6. A method for identifying events occurring at a location-of-interest comprising:

capturing continuously, using a Fisheye image capturing device provided at the location-of-interest, wide-angle image frames and converting the captured wide-angle image frames into non-distorted perspective images using a set of calibration parameters generated by a fisheye correction scheme, and store the non-distorted perspective images;

capturing, using an audio module provided at the location-of-interest, audio signals and determining a Direction of Arrival (DOA) of the audio signals;

receiving, using a computing module communicatively coupled to the audio module and the Fisheye image capturing device, a notification from the audio module when the audio module detects an event having an audio signal at the location-of-interest, whereby the notification includes data of a DOA of the audio signal;

instructing, using the computing module, the Fisheye image capturing device to generate, using a trained single-image-super-resolution (SISR) scheme, high-resolution

perspective images based on the stored non-distorted perspective images associated with the DOA of the audio signal,

wherein the step of detecting the event having an audio signal at the location-of-interest comprises:

generating, using the audio module, a discrete set of DOA points that cover a DOA region from which the audio signal of the event was detected; and

embedding, using the audio module, the discrete set of DOA points in the notification that is sent to the computing module

wherein the step of generating the high-resolution perspective images based on the non-distorted perspective images by the Fisheye image capturing device comprises:

extracting, using the Fisheye image capturing device, the discrete set of DOA points received from the computing module, wherein the discrete set of DOA points are provided in the instructions sent by the computing module;

generating, using the Fisheye image capturing device, a unified region of interest from the stored non-distorted perspective images that correspond to the extracted set of DOA points; and

generating, using the Fisheye image capturing device and the trained single-image-super-resolution (SISR) scheme, high-resolution perspective images of the unified region of interest.

~~9.~~7. The method according to claim ~~8~~6 wherein the fisheye correction scheme comprises a fisheye rectilinear transformation scheme.

~~10.~~8. The method according to claim ~~8~~6 wherein the audio module comprises:

a plurality of sensor arrays configured to utilize a beamforming technique to estimate a direction of arrival of captured audio signals.

~~11. The method according to claim 8 wherein the step of detecting the event having an audio signal at the location-of-interest comprises:~~

~~generating, using the audio module, a discrete set of DOA points that cover a DOA region from which the audio signal of the event was detected; and~~

~~embedding, using the audio module, the discrete set of DOA points in the notification that is sent to the computing module.~~

~~12. The method according to claim 11 wherein the step of generating the high-resolution perspective images based on the non-distorted perspective images by the Fisheye image capturing device comprises:~~

extracting, using the Fisheye image capturing device, the discrete set of DOA points received from the computing module, wherein the discrete set of DOA points are provided in the instructions sent by the computing module;

generating, using the Fisheye image capturing device, a unified region of interest of the non-distorted perspective images that corresponds to the extracted set of DOA points; and

generating, using the Fisheye image capturing device and the trained single-image-super-resolution (SISR) scheme, high-resolution perspective images of the unified region of interest.

~~13.~~9. The method according to claim ~~12~~ 6 wherein the step of generating the unified region of interest of the captured wide-angle image frames comprises:

for every DOA point, extracting, using the Fisheye image capturing device, a corresponding θ±36 degree region of interest from the non-distorted perspective images; and compiling all the extracted θ±36 degree regions of interest to form the unified region of interest.

~~14.~~10. The method according to claim ~~8~~6 further comprising the steps of:

capturing continuously, using another Fisheye image capturing device provided at the location-of-interest, wide-angle image frames and converting the captured wide-angle image frames into non-distorted perspective images using another set of calibration parameters generated by ~~a~~the fisheye correction scheme;

capturing, using another audio module provided at the location-of-interest, audio signals and determining a Direction of Arrival (DOA) of the audio signals;

receiving, using the computing module communicatively coupled to the another audio module and the another Fisheye image capturing device, a notification from the another audio module when the another audio module detects the event having the audio signal at the location-of-interest, whereby the notification includes data of another DOA of the audio signal;

instructing, using the computing module, the another Fisheye image capturing device to generate, using the trained single-image-super-resolution (SISR) scheme, another high-resolution perspective images based on the captured wide-angle image frames associated with the another DOA of the audio signal;

receiving, using an event identification module, the high-resolution perspective images, and the another high-resolution perspective images from the computing module; and stitching the received images to form a single set of high-resolution perspective images.