Spatio-Temporal Lifelog Using a Wearable Compound Omnidirectional Sensor

Haruka Azuma, Yasuhiro Mukaigawa, Yasushi Yagi

Spatio-Temporal Lifelog Using a Wearable Compound Omnidirectional Sensor

Haruka Azuma, Yasuhiro Mukaigawa, and Yasushi Yagi

The Institute of Scientific and Industrial Research, Osaka University, Japan
{azuma, mukaigaw, yagi}@am.sanken.osaka-u.ac.jp

Abstract. In this paper, we propose a spatio-temporal lifelog, that enables us intuitively to understand both the spatial and temporal aspects of a situation. We have designed a new wearable compound omnidirectional sensor comprising a hyperboloidal mirror and multiple paraboloidal mirrors. Using this sensor, visual information is recorded in all directions around the user. Objects are classified as foreground or background according to their distance. The stored log can be viewed using a browser that supports two different views, a spatial view and temporal view, which allow an effective visualization of the situation based on the classification. The user can switch between views depending on the purpose.

1 Introduction

Throughout human history, man has attempted to record his activities to enable these to be recalled at a later stage. Traditional recording methods, such as writing a diary or taking photos, are now maintained by digital devices. Recently there has been a trend to record all our daily activities as digital data without any human intervention. By storing such information automatically it would be possible to search for a person you once met or relive an ordinary day from the past as if you watch a movie of yourself.

Previous studies have focused on achieving an effective automatic lifelog. Microsoft Research developed a wearable camera containing other sensors such as a motion sensor [1]. The camera with a fish-eye lens covers almost all of the user's view. In one of the studies using the sensor, captured images are browsed with different sized icons ranked by frequency of appearance [2]. Aizawa et al. [3] used brainwave data to distinguish an image including the object which caught the user's attention. In all these studies, captured images have been used to re-create or complement a user's visual memory.

These previous lifelog systems have not, however, focused on spatial and temporal understanding of the event. Spatial information such as the directions of the surrounding people acts as a clue when recalling the past. Temporal information such as continuous relationships with different people is also useful if it is given all together. To develop a spatio-temporal lifelog, we need to acquire visual information that surpasses that which is detectable by the user's eye. Moreover,

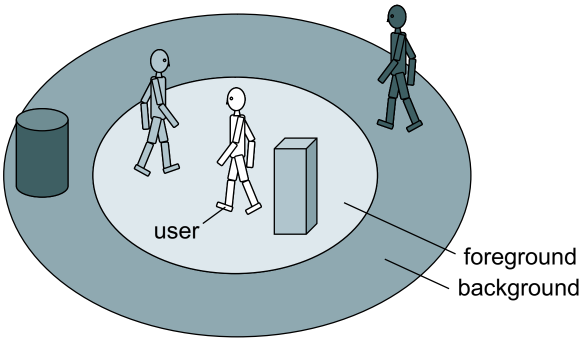


Fig. 1. Model of the surrounding environment. The foreground contains a group of objects that could be elements included in an activity related to the user, while the background contains a group of objects specifying the location in which the user exists.

it is necessary to visualize such rich information in a way that allows the user to understand the situation intuitively.

In this paper, we propose a spatio-temporal lifelog using a compound omnidirectional sensor that can acquire visual information around a user and can classify the objects in the environment according to distance. We have designed a new wearable sensor that consists of a hyperboloidal mirror and multiple paraboloidal mirrors. We have also developed a browser with two different views aimed at enhancing spatial and temporal understanding, respectively.

2 Spatio-Temporal Lifelog

A spatio-temporal lifelog is a tool to aid the understanding of a situation from both spatial and temporal aspects. We need spatial information to perceive the situation surrounded by other objects such as friends or buildings. The directions from where we are to these objects and positional relations among the objects are key to recalling the situation precisely. Additionally to ascertain how situations have changed over time without spending too much time checking, we need to see all the changes in one glance. Temporal information specifies locations which change continuously as we walk and as a result of the motion of objects around us.

To clarify the requirements for a spatio-temporal lifelog, we first model the surrounding environment as illustrated in Fig.1. The lifelog user is in the center, surrounded by two categories of objects, namely foreground and background objects, according to the context of association with the user's event. The foreground comprises a group of objects that could be elements included in an activity related to the user. On the other hand, the background comprises a group of objects specifying the location in which the user exists. In other words, objects are distinguished according to their roles with respect to the user. For example, while working in an office, a computer screen in front of you or a colleague sitting next to you would be considered part of the foreground. In contrast, the walls of the room surrounding you or some colleagues working farther from you would be background objects.

To build a spatio-temporal lifelog based on this environmental model, visual information from all directions around the user is necessary. The objects in the environment also need to be classified as foreground and background. Finally the acquired information needs to be presented in a meaningful way to enable spatial and temporal understanding of the situation.

3 Compound Omnidirectional Sensor

3.1 Sensor Design

When recording, it is necessary to capture the scene around the user together with information to classify objects as foreground or background. We distinguish foreground and background according to the distance from the sensor, with near objects being classified as foreground and far objects as background. This solution is valid where an important object such as a friend, who is talking to you and is immediately in front of you, is classified as foreground.

To obtain omnidirectional view and distance information, various catadioptric stereo systems have been proposed. Jang et al. [4] proposed a small sensor with two paraboloidal mirrors upside down, but which only has a single baseline for stereo matching. Several compound omnidirectional sensors have been proposed to allow multiple baselines for stable distance estimation. Kojima et al. [5] developed a compound omnidirectional sensor with a spherical mirror in the middle and smaller ones around it. The problem with this sensor is that a complete perspective image cannot be produced because a spherical mirror does not maintain a single viewpoint. Ngo et al. [6] proposed a sensor with seven paraboloidal mirrors, all of which have their respective single viewpoint. However, the telecentric lens used in this sensor resulted in the sensor being too big to be attached to a human body.

For use in a lifelog, we have designed a new wearable compound omnidirectional sensor, that consists of a hyperboloidal mirror, a concave paraboloidal mirror and four paraboloidal mirrors. The optical part of our sensor is a combination of two layers as shown in Fig.2. The larger mirror, illustrated in (a), is hyperboloidal and is used to obtain a high resolution image. A ray going to a focal point of the hyperboloidal mirror, depicted as F_h in (a), and which is a viewpoint, is reflected to a single point F'_h, which is another focal point. On the other hand, the smaller paraboloidal mirrors are used to supply multiple viewpoints and are illustrated in (b). A ray going to a focal point of the convex hyperboloidal mirror, depicted as F_p in (b), is reflected vertically to the

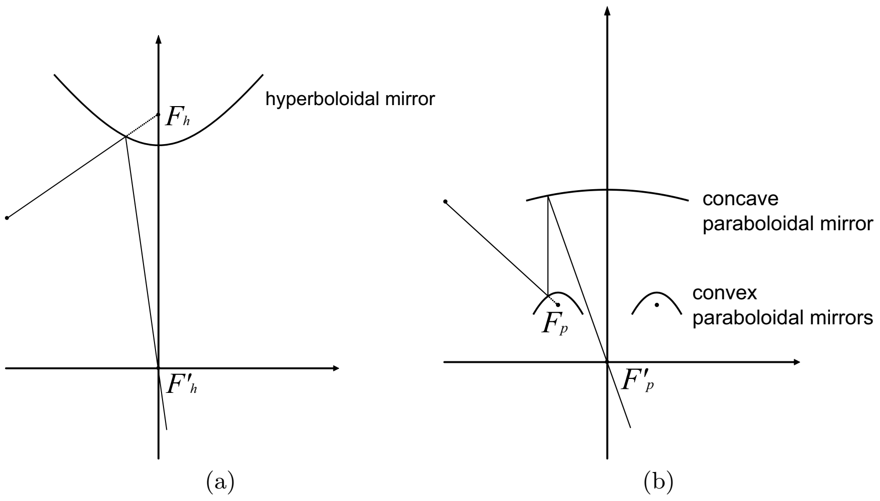


Fig. 2. Configuration of two layers: (a) the layer with a hyperboloidal mirror, and (b) the layer with paraboloidal mirrors. Both configurations retain a single viewpoint (F_h/F_p) and converge a ray to a single point (F'_h/F'_p).

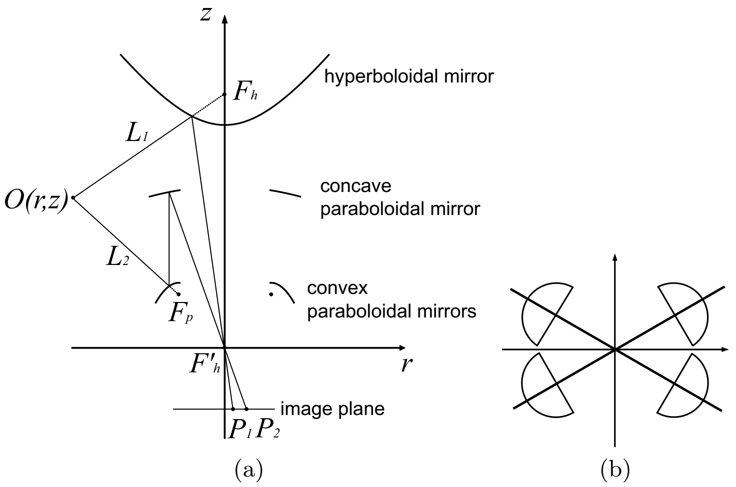
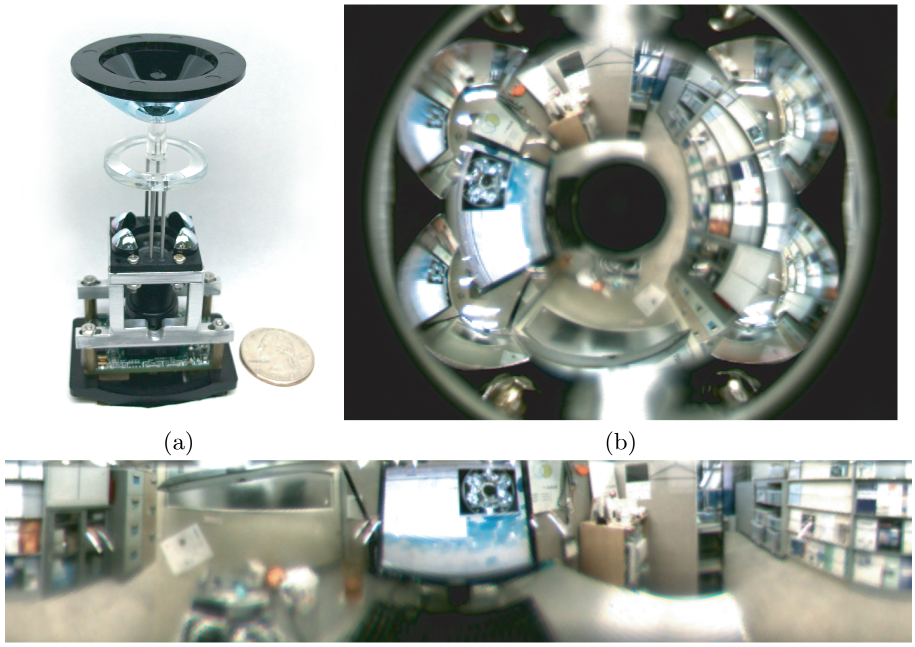


Fig. 3. Configuration of the compound omnidirectional sensor: (a) a cross-section view of the overall optical part along the thick diagonal line indicated in (b), which is a top view of the four convex paraboloidal mirrors. The ray from $O(r, z)$ follows two different light paths L_1 and L_2, that meet at F'_h, which is a focal point of both the hyperboloidal mirror and the concave paraboloidal mirror.



(a)

(b)

(c)

Fig. 4. The omnidirectional sensor developed: (a) the sensor in relation to a 25-cent coin, (b) a captured image, and (c) a panoramic image.

upper concave paraboloidal mirror, and then folded to the point F'_p, which is a focal point of the concave paraboloidal mirror. A folded catadioptric system[7] adopted in layer (b) shrinks the overall sensor size.

The integrated optical schematic of the sensor is shown in Fig.3. (a) is a cross-section view of the overall optical part and (b) is a view from the top of four paraboloidal mirrors, with diagonal lines indicating the cross section. A light ray L_1 from $O(r, z)$ going to the focal point F_h, which is also a viewpoint, is reflected to another focal point F'_h, and finally projected to P_1 on the image plane. On the other hand, a light ray L_2 going to the focal point F_p is reflected in parallel to the optical axis z, then folded by the concave paraboloidal mirror, passes the focal point F'_h, which is also a focal point of the concave paraboloidal mirror, and finally reaches P_2 on the image plane. Multi-baseline stereo is achieved among the hyperboloidal mirrors and four convex paraboloidal mirrors.

3.2 Sensor Specification

Here we describe the setup of our compound omnidirectional sensor in detail. Figure 4 (a) is the overall view of the sensor. The height of the optical part is 50mm and the diameter of the central hyperboloidal mirror is 40mm. The

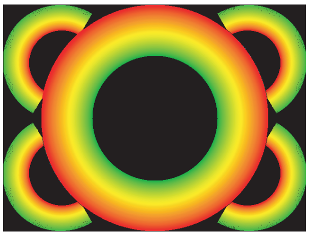


Fig. 5. Correspondence between the hyperboloidal mirror and surrounding paraboloidal mirrors when every ray is from infinity.

overall height including the camera is 85mm. The weight of the sensor is only 50g because the mirrors are made of plastic with an evaporation coating. The camera has a 1/3 inch CCD with 480 by 640 pixels.

The vertical baseline between the hyperboloidal mirror and the convex paraboloidal mirrors is 30.2 mm long, while the shortest and longest horizontal baselines of the convex hyperboloidal mirrors are 8.3mm and 14.0mm, respectively. A vertical overlap of the field of view among the hyperboloidal mirror and convex paraboloidal mirrors is between –20.2 degrees and 11.5 degrees.

Figure 4 (b) shows an example of a captured image, with (c) depicting its panoramic image. The convex paraboloidal mirrors are visible at the four corners in the image in (b).

3.3 Near Object Detection

The objects can be classified by distance with the compound omnidirectional sensor. In principle, it is possible to calculate the precise depth by searching the corresponding points between mirrors[8]. However we only need two classifications based on whether an object is near or far from the sensor. For this reason, we have adopted a method of discriminating between only two alternatives, near or far [5].

If we assume that every ray to an image plane comes from infinity, there is no disparity among mirrors. The resolution of the camera defines a minimum distance that is sufficiently far and treated as infinity. Corresponding points between mirrors are calculated based on their positional relation, assuming that a ray from infinity reaches each mirror in parallel. Finally we know whether a ray comes from closer or further than the minimum distance by comparing the pixel values between these corresponding points. If the difference of the pixel values is less than a threshold, a far object is seen in the area. In contrast, if the difference is more than the threshold, there is a near object.

Figure 5 shows the correspondence between the hyperboloidal mirror and paraboloidal mirrors when every ray comes from infinity. The corresponding pixels are marked with the same color. Image size is normalized to fit the area of the mirrors. The inner zone of the hyperboloidal mirror corresponds to the outer

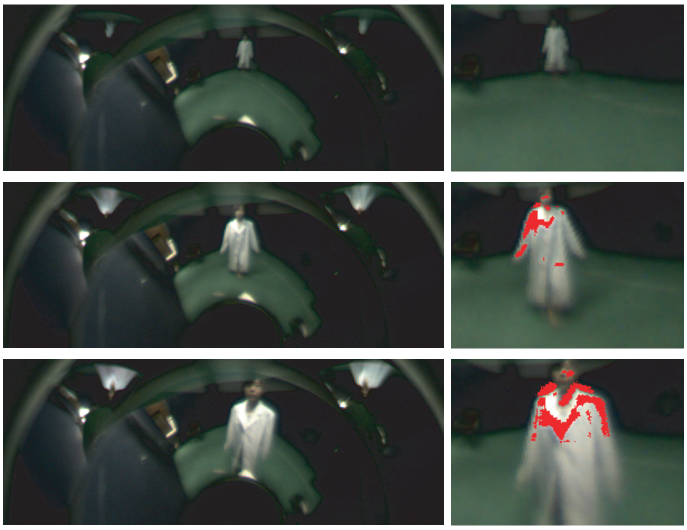


Fig. 6. Three examples of near object detection. The left column shows the inputs, while the right column illustrates the resulting perspective images where the area detected as near is marked in red.

rim of the paraboloidal mirrors since the hyperboloidal mirror and the convex paraboloidal mirrors are arranged upside down.

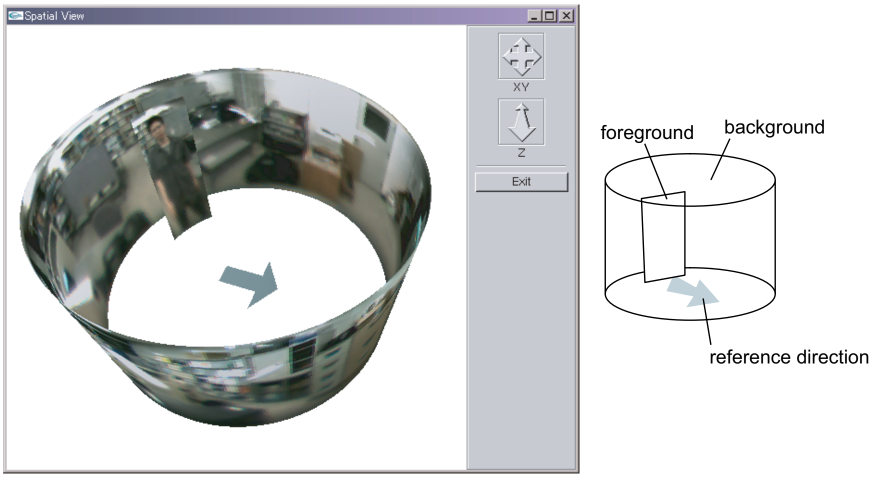
Three examples of near object detection are shown in Fig.6. The left column shows the inputs, while the right column illustrates the resulting perspective images where the area detected as near is marked in red. In the top image, where the person is at a distance of 4m, no near area is detected. In the middle and bottom images, where the person is at a distance of 2m and 50cm respectively, the area around the person is detected as near.

4 Lifelog Browser

4.1 Multiple Views

We now possess omnidirectional visual information around the user and a classification of foreground and background. Next, we focus on a means of browsing which is suitable for understanding the situation both spatially and temporally. The guiding principles we followed to build the browser are given below:

- Spatial view
  - The positional relations between the user and other objects are recognizable.
  - The positional relations among foreground objects are recognizable.
- Temporal view



(a) (b)

Fig. 7. An example of spatial view: (a) the browser, and (b) a description thereof. A user can easily recognize the local relation between a person and the foreground/background objects or among foreground objects.

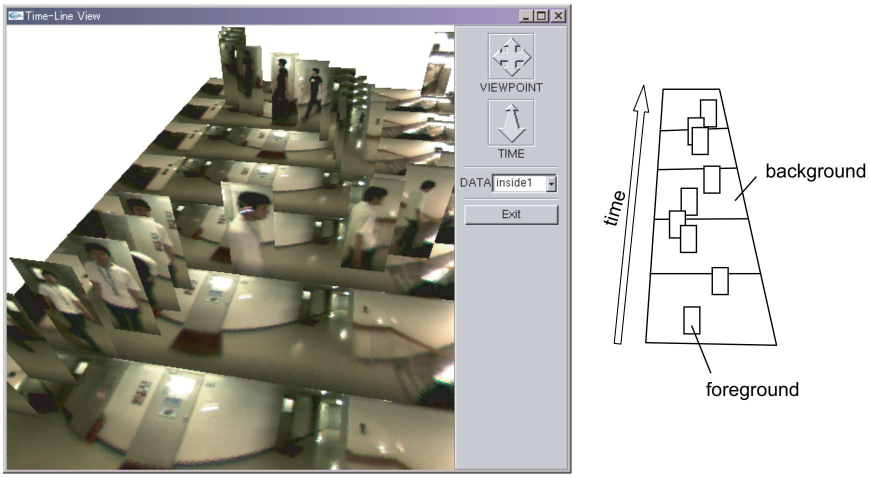
- The changing situation in the foreground is recognizable at a glance without checking sequential images from a video.
- The location where a person exists can be perceived.

Omnidirectional images are often viewed as a panoramic image, which achieves the same perspective view as that from our eyes. However, when we think of spatial understanding, a panoramic image does not suffice since it does not depict any spatial relations among objects in the image. To perceive a situation spatially, three-dimensional expression is needed. Meanwhile, to enable us to deal with a situation temporally without having to check a sequence of images, three-dimensional expression with time as the third axis is crucial.

Consequently, we propose a browser with two different views that facilitate spatial and temporal understanding, respectively. The two views have been designed to enable the user to understand the situation intuitively from different aspects rather than presenting the whole omnidirectional image captured by the sensor. The user can switch between these views depending on the purpose. In addition, the viewpoint can be freely controlled to display the area the user wishes to see.

4.2 Spatial View

As shown in Fig.7, spatial view depicts the background as a cylinder and foreground objects as billboards inside. (a) is our browser showing spatial view and (b) is an explanation thereof. The cylinder represents one of the backgrounds.



(a) (b)

Fig. 8. An example of temporal view: (a) our browser, and (b) the description thereof. A user can easily recognize the changing situation at a glance.

Foreground objects appear and disappear with the automatic progression of time. The arrow at the bottom indicates the reference direction related to the position of the sensor. In this view temporal progress is merely shown as a series of images as in a movie. Dragging the cylinder with a mouse causes it to be rotated in any direction for better visibility. The buttons on the right of the window can be used to translate the viewpoint into a three dimensional space.

In this view, the positional relations between the user and foreground or background objects, or the relations among foreground objects are easily recognized. By representing images as a three-dimensional form, the environment is visualized realistically and can be perceived intuitively. This view allows a user to concentrate on understanding the situation spatially, while some areas are designed not to be seen because of the shape of the displayed image.

4.3 Temporal View

The temporal view represents the background as a road and the foreground as billboards on it as shown in Fig.8. (a) is our browser showing the temporal view and (b) is the explanation thereof. Temporal progress is expressed as depth, with later time on the far side. The background appears continuously in the road to inspire the user to recall the place where he was standing. Foregrounds appear as billboards walking along the background road. Multiple billboards represent a single person or object as they remained continuously around the user. In this view, as the user walks or as the foreground/background moves around the user whilst recording, a continuous change in textures is shown on the road or strings of billboards. As in the spatial view, the entire foreground and background can

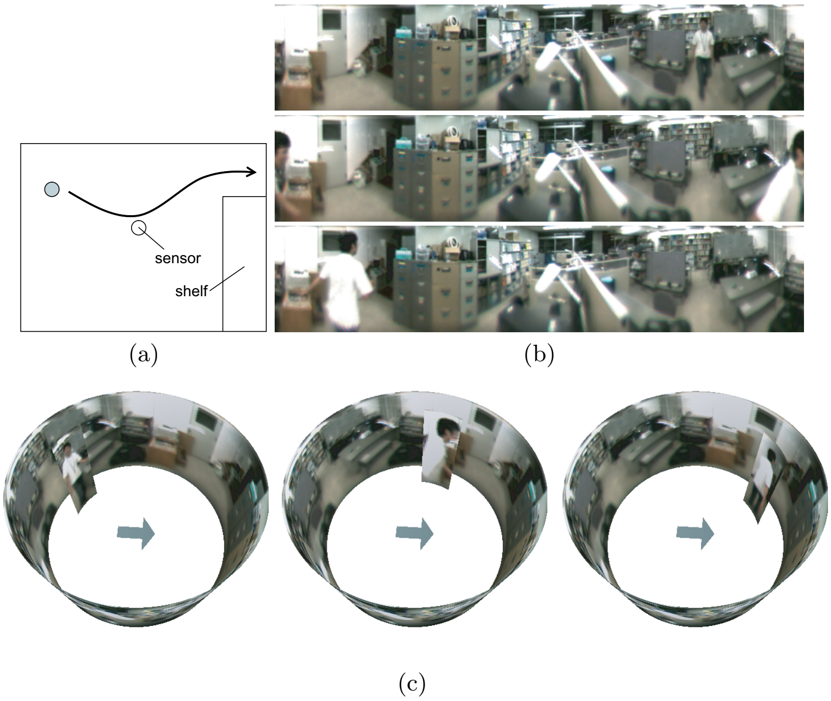


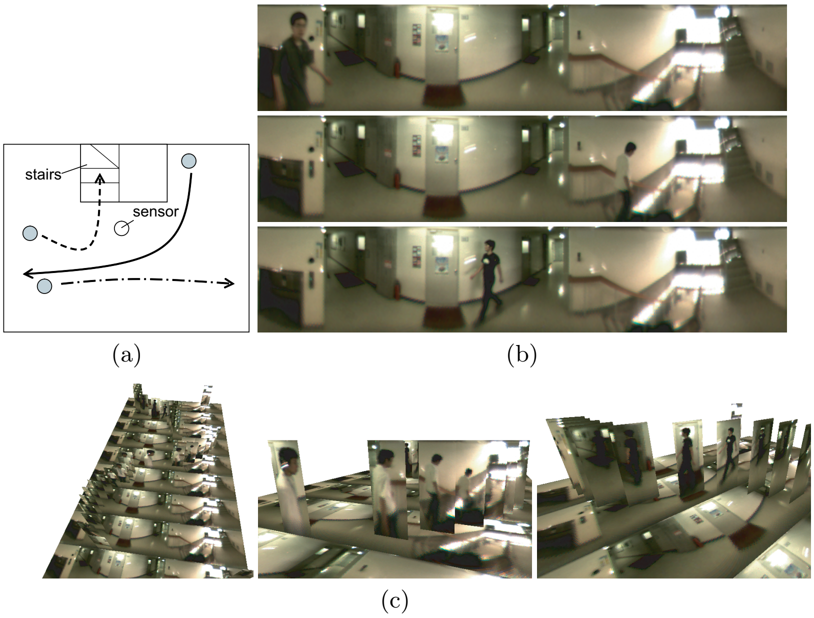
Fig. 9. Experimental results of spatial view: (a) the situation, (b) the panoramic images and (c) the spatial view. The compound omnidirectional sensor is in the center, a person approaches the sensor and then steps away along the line.

be rotated by dragging with a mouse, and the viewpoint can also be translated. The main operation in this view is translation along the depth combined with temporal progress.

In this view, a user can easily recognize the changing situation at a glance without checking it in a movie. This view omits spatial representation in a three-dimensional space to allow a user to focus on the temporal changes in the foreground and background.

4.4 Experimental Results

Using the compound omnidirectional sensor we recorded two situations in an indoor environment. For these experiments, the sensor was fixed to a tripod. Our compound omnidirectional sensor can classify objects as foreground and background, but the accuracy of the classification depends on the alignment. To demonstrate the effectiveness of the browser, foreground and background areas in the images are chosen manually. Foreground areas are demarcated with rectangles to be easily recognized.



(a) (b)

(c)

Fig. 10. Experimental results of temporal view: (a) the situation, (b) panoramic images, and (c) the temporal view. The compound omnidirectional sensor is in the center, and three people approach the sensor and step away along the lines.

Figure 9 shows experimental results that demonstrate the advantage of the spatial view over panoramic images. The situation with a sensor in the center of a room and a person walking towards the sensor and then stepping away along the path indicated by the line, is illustrated in (a),. In the three panoramic images shown in (b), it is difficult to follow the person knowing his direction from the sensor. A further problem here is that his trajectory is segmented on the way. In contrast, the three representative images in the spatial view shown in (c) allow us to see the person passing on the left of the sensor. His movement is correctly recognizable in terms of the positional relation to both the sensor and the background.

Figure 10 shows the experimental results that demonstrate the advantage of the temporal view over panoramic images. The situation in which a sensor is located in a corridor and three people walk towards the sensor and then step away along the path indicated by the three different lines is illustrated in (a),. In the panoramic images shown in (b), we can see that three people appear around the sensor, but the three images are not enough to understand their motion. The representative images snapshot from the three different viewpoints in the temporal view are shown in (c). As observed in the first image, the continuous movement of multiple people is recognized at the same time with a single still image. In the second image we can see that one of the people is about to descend the stairs. The third image depicts a series of continuous motion of another person. As demonstrated in these examples, temporal view is useful to understand temporal change without time consuming viewing.

5 Conclusions

In this paper we have proposed a new spatio-temporal lifelog which enables us to understand a situation both spatially and temporally. We modeled the environment as a combination of foreground objects, which are directly related to a user's event, and backgrounds, which specify the location in which the user exists. We designed a new wearable compound omnidirectional sensor that can acquire visual information around the user and can classify the objects in the environment as foreground and background objects based on their distance. We developed a browser with both spatial and temporal views, which allows a user to understand a situation from two aspects.

Future work includes accurate calibration and alignment of the mirrors and the camera for better accuracy of near and far object classification. Focusing will be improved by replacing the current lens with another which has the wider focal depth. We also plan to add other views to the browser, making use of omnidirectional visual information and separation of foregrounds and backgrounds.

This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (S).

References

1. Gemmell, J., Williams, L., Wood, K., Lueder, R., Bell, G.: Passive capture and ensuing issues for a personal lifetime store. In: CARPE04. (2004) 48–55
2. Smeaton, A.F.: Content vs. context for multimedia semantics: The case of sensecam image structuring. In: SAMT2006. (2006) 1–10
3. Aizawa, K., Hori, T., Kawasaki, S., Ishikawa, T.: Capture and efficient retrieval of life log. In: Pervasive 2004 Workshop on Memory and Sharing Experiences. (2004) 15–20
4. Jang, G., Kim, S., Kweon, I.: Single-camera panoramic stereo system with single-viewpoint optics. Optics Letters **31** (2006) 41–43
5. Kojima, Y., Sagawa, R., Echigo, T., Yagi, Y.: Calibration and performance evaluation of omnidirectional sensor with compound spherical mirrors. In: OMNIVIS2005. (2005)
6. Thanh, T.N., Nagahara, H., Sagawa, R., Mukaigawa, Y., Yachida, M., Yagi, Y.: Robust and real-time rotation estimation of compound omnidirectional sensor. In: ICRA2007. (2007) 4226–4231
7. Nayar, S.K., Peri, V.: Folded catadioptric cameras. CVPR **02** (1999) 2217
8. Gluckman, J., Nayar, S.K., Thoresz, K.J.: Real-time omnidirectional and panoramic stereo. In: DARPA1998 Image Understanding Workshop. (1998) 299–303