ROBOMECH Journal

RESEARCH ARTICLE

Open Access

Projection-mapping-based object pointing using a high-frame-rate camera-projector system

Deepak Kumar[1], Sushil Raut[2], Kohei Shimasaki[2], Taku Senoo[3] and Idaku Ishii[3]*

Abstract

The novel approach to physical security based on visible light communication (VLC) using an informative object-pointing and simultaneous recognition by high-framerate (HFR) vision systems is presented in this study. In the proposed approach, a convolutional neural network (CNN) based object detection method is used to detect the environmental objects that assist a spatiotemporal-modulated-pattern (SMP) based imperceptible projection mapping for pointing the desired objects. The distantly located HFR vision systems that operate at hundreds of frames per second (fps) can recognize and localize the pointed objects in real-time. The prototype of an artificial intelligence-enabled camera-projector (AiCP) system is used as a transmitter that detects the multiple objects in real-time at 30 fps and simultaneously projects the detection results by means of the encoded-480-Hz-SMP masks on to the objects. The multiple 480-fps HFR vision systems as receivers can recognize the pointed objects by decoding pixel-brightness variations in HFR sequences without any camera calibration or complex recognition methods. Several experiments were conducted to demonstrate our proposed method's usefulness using miniature and real-world objects under various conditions.

Keywords: Physical security, Cyber-physical systems, High-speed camera-projector, Informative projection mapping, Real-time pattern recognition, visible light communication

Introduction

Physical security using cyber-physical systems (CPS) involves numerous interconnected systems to monitor and manipulate real objects and processes. CPS integrates the information and communication systems into the physical objects by active feedback from the physical environment. They incorporate CPS systems to exchange various types of data and confidential information in real-time to play a vital role in Industry v4.0 [1] and infrastructure security, enabling smart applications services to operate accurately. The infrastructures' physical security requires various measures to prevent unauthorized access to facilities, equipment, and resources. These measures are interdependent systems that mainly include surveillance by vision-based object detection and recognition to ease observation of distributed systems. However, these technologies are computationally and economically expensive. Most of the infrastructures are equipped with multiple cameras and alarm systems, which increase installation and maintenance costs. If the security equipment is enabled with artificial intelligence (AI), then the cost goes higher.

The proposed novel approach integrates CPS in physical security using VLC to transmit and receive information about static and moving environmental objects without causing any distraction to the human visual system (HVS). As a transmitter, the AiCP system broadcasts the CNN-detection results using an SMP-based encoded projection mapping, referred to as informative

*Correspondence: iishii@hiroshima-u.ac.jp
[3] Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima, Japan
Full list of author information is available at the end of the article

Kumar *et al. Robomech J*   (2021) 8:8

Page 2 of 21

object-pointing (IOP) in this study. Whereas distantly located HFR vision-based multiple receivers decode each pattern captured frame-by-frame and identify the objects simultaneously. Hence, this approach reduces the computational load on the receivers required for complex recognition algorithms in a distributed processing. The imperceptibility of IOP to an HVS is achieved by mapping the SMP-based projection masks on desired objects facilitated by a projection system of a higher projection-rate than the critical frequency of HVS. The SMP projected at hundreds of fps can only be perceived by a vision system of equivalent frame rate or higher than that of a projection system. We observed that the information to be communicated is constrained by the transmitter and the receiver's framerate. Therefore, a higher framerate can transfer a higher amount of information. We used a high framerate projector with color wheel filters that projects each red, green, and blue light at 480 Hz sequentially. A single filter of the color wheel represents a single bit of information. Hence, using a combination of three color-filters of the high-speed projector, we can send 2^3-bits sequentially that resemble the information of a maximum of eight objects.

Related works

In this study, we primarily focus on prototyping CNN-object detection assisted projection mapping that can encode information on the environmental object and HFR vision-based decoding while maintaining the confidentiality of data to be communicated. Projection mapping has been used in the entertainment industries and in scientific research as a surface-oriented video projection system for augmenting realistic videos onto the desired surfaces [2–4]. They are widely used for visual augmentation in buildings, rooms, and parks [5–7]. Projection-mapping-based systems are mainly classified as static and dynamic projection mapping. Static projection mapping is usually preferred in industries and scientific researches for shape analysis using structured light projection mapping [8, 9]. It involves a projection of light patterns by manually aligning the objects and projectors [10–13]. In dynamic projection mapping (DPM), a system tracks the desired surfaces' positions and shapes using a marker [14–16] followed by model [17–19] tracking methods to project videos onto the moving surfaces. Asayama et al. [20] proposed an approach on visual markers for the projection of dynamic spatial augmented reality (AR) on fabricated objects. DPM requires heavy computation to acquire dynamic, realistic effects in real-time [21, 22]. Narita et al. [23] have explained using a dot cluster marker for DPM onto a deformable nonrigid surface. The use of an RGB depth sensor-assisted projector with a DPM to render surfaces of complex geometrical shapes

was reported [24] for developing an interactive system of surface reconstruction. Several approaches using nonintrusive and imperceptible patterns have been presented using projection mapping. Lee et al. [25] have proposed a location tracking method based on a hybrid infrared and visible light projection system. Their system has the unique capabilities of providing location discovery and tracking simultaneously. Visible light-emitting projection devices such as high-speed digital light projection (DLP) systems are enabled with a high-frequency digital micromirror device (DMD) to project binary image patterns at thousands of fps [26–29]. DMD projectors have been used in structured-light-based three-dimensional (3D) sensing, interactive projection mapping, and other geometric and photometric applications [30–33]. Daniel et al. [34] presented a simultaneous acquisition and display method that can embed imperceptible patterns in projected images. High-speed switching between the projected pattern and its complementary pattern with DLP is used in their research, indistinguishable by HVS. However, the resultant projection leads to lower brightness, and hardware modification is required in such a system [35]. High-speed projection systems that can emit light at a higher frequency than the HVS have been used in numerous AR applications. The projection patterns and their complementarity at 120 Hz are sufficient to generate uniform brightness projections to the HVS [36–38]. Color-wheel filter-based 3D projectors with the DLP principle can emit 120-Hz color-plane patterns [39, 40]. Projection mapping based VLC has been used to establish a wireless link between projection and sensing systems to transmit anticipated information [41–43]. Kodama et al. [44] have designed a VLC position detection system embedded in single-colored light using a DMD projector. They used photodiodes as sensors to decode the projected area location for IoT applications. However, the photodiode-based sensor cannot obtain complete projection information at an instant. Conventional vision systems that operate at tens of fps cannot capture temporal changes in high-speed projection. They lead to a severe loss of temporal information. Hence, an HFR vision system to sense temporal alterations in high-speed projection data is required. With millisecond-level accuracy, HFR vision systems operated at hundreds or thousands of fps have been used for various industrial applications [45–48]. A saccade mirror and HFR cameras have been used to add visual information in real-time for the projection-based mixed reality of dynamic objects [49]. An HFR camera-projector depth vision system has been used for simultaneous projection mapping of RGB light patterns augmented on 3D objects by computing the depth using a camera projector system [50]. Temporal dithering of high-speed illumination was reported for

fast active vision [51]. HFR vision systems have also been used as sensing devices in many applications such as optical flow [52], color histogram-based cam-shift tracking [53, 54], face tracking [55], image mosaicking, and stabilization [56, 57]. Hence, HFR vision systems can be used as an environment sensing device in CPS.

Concept

Recently emerging AI-vision systems are most informative for human operators and an automatic alarming system for physical security by reducing the efforts required for manual monitoring. The AI-enabled camera-projector system is used in this research to reduce the computational cost required in multiple surveillance vision systems by broadcasting the CNN-object detection results. Hence numerous cost-efficient systems can synthesize the same results as illustrated in Fig. 1. The proposed active projection mapping and simultaneous recognition system consists of three parts, (1) a smart object pointing using AiCP system as a transmitter, (2) an HFR vision-based object recognition system as a receiver, and (3) encoding and decoding protocol in VLC.

Smart object pointing using AiCP system

Various studies reported that HVS could not resolve rapid visual changes beyond the critical frequency F_{cf} of 60 Hz except the subconscious effects under most conditions. In the proposed system, we used a DLP projector of projection frequency higher than the F_{cf} of HVS. The temporal sensitivity of the HVS is subtle, with bright components of the light. However, the sensitivity decreases as contrast reduces. The DLP projector can control the light to be passed, resulting in an overall image to appear as an integrated image. The DLP projector emits a series of light pulses at variable time intervals to obtain the desired light intensity. The object detection system outputs classes of detected objects and their region of interest (ROI) using a complex algorithm. The smart object pointing system transmits informative light using the AiCP system based on object pointing code (OPC) for a particular object, which is unique for every input intensity, known as temporal dithering of the illumination. A unique SMP color mask based on OPC is projected onto the same objects like a spotlight to catch HFR vision systems' attention while maintaining imperceptibility.

HFR Vision-based recognition system

An HFR vision system of equivalent framerate, same as the AiCP system or higher, can perceive the temporal changes (i.e., time-varying photometric properties) in the projection area. The HFR vision system acquires informative SMP projected onto the objects frame-by-frame in the form of a sequence of packets of information. The
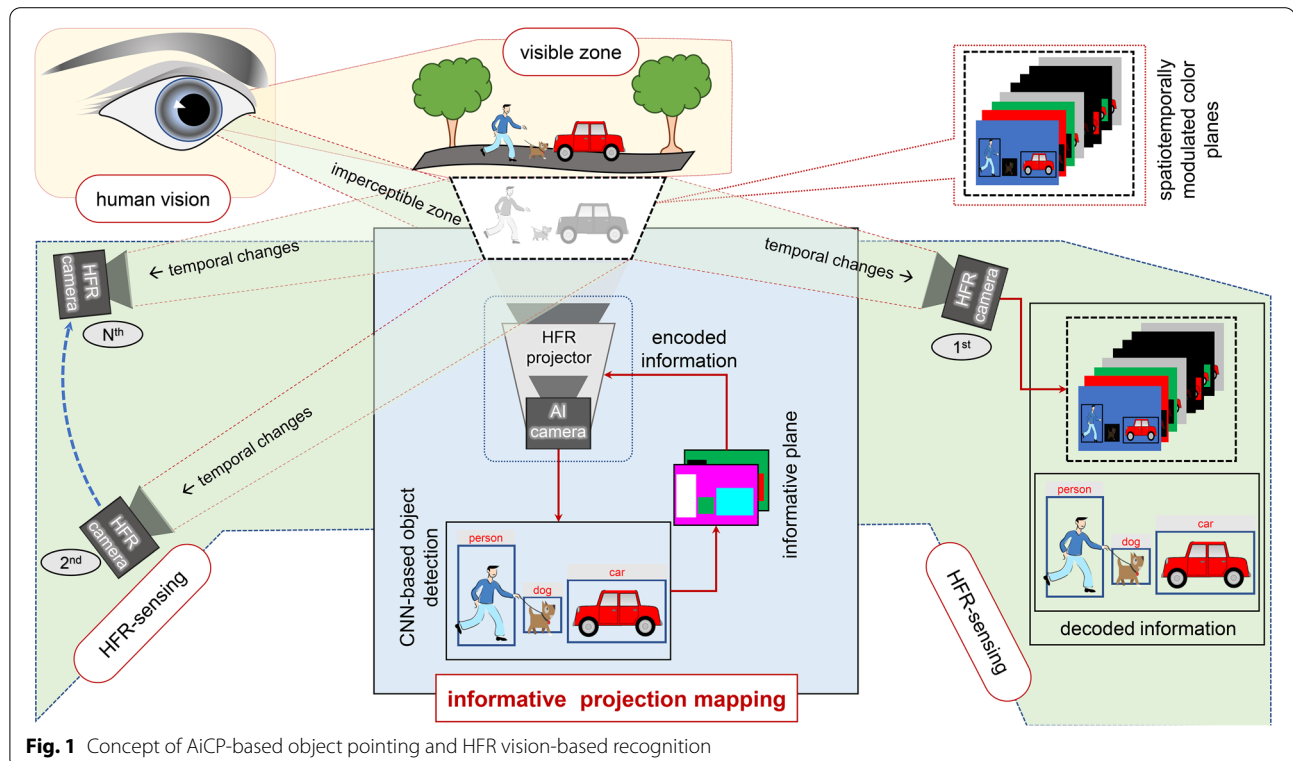

Fig. 1 Concept of AiCP-based object pointing and HFR vision-based recognition

OPC in a packet of information is encoded in a time-varying color mask. Both transmitter and receivers should be synchronized to know the start and end of a packet of information. In this study, an HFR vision system is distantly located without any wired synchronization and acts as a receiver of informative light. Both the systems are optically synchronized by projecting a spatial header with the status of the projection plane. Hence, any number of remotely located HFR vision systems observing the projection area can recognize the pointed objects using computationally efficient methods.

Encoding and decoding protocol in VLC

The functional blocks of the VLC-based transmitter and receiver are shown in Fig. 2. The smart object pointing using AiCP system as a transmitter consists of an AI-enabled camera and a spatiotemporal encoding block.

A color-wheel-based high-speed single-chip DLP projector consists of spectral distributions with segments of blue filter (B, 460 nm), red filter (R, 623 nm), green filter (G, 525 nm), and a blank transparent filter. The wheel rotates at high-speed to generate various combinations of RGB planes for each image as modulated and emitted color plane slices of blue, red, and green patterns. The blank transparent filter adds the overall brightness to the projection area. The imperceptibility can be achieved when a packet of *N*-light planes has a combination of each *N*/2 spatiotemporal color plane and its inverse planes in sequential order. Thus, due to temporal dithering, the mapped informative color masks as a pointer

should be a non-flickering light source to the HVS and the conventional 25 to 30 fps vision systems. As illustrated in Fig. 3, we encode the information in two phases of the projection as a forward projection phase (FPP) and inverse projection phase (IPP). The AiCP system generates three color planes of FPP along with an embedded color mask and another three planes of IPP with the complementary color mask onto the objects to be pointed. The combination of color masks in FPP and IPP is chosen to visualize the accumulated light as a uniform bright gray level light within the projection area.

At the receiver end, HFR vision-based recognition consists of spatiotemporal decoding and object recognition blocks. HFR vision acquires all the SMP-planes frame-by-frame and decodes the packets of information embedded in each frame by temporally observing each pixel in an image that corresponds to the projection area. The amplitude of pixel brightness determines the information encoded in the image. All the pixels covering the projection area have the same high projection frequency; however, variation in phase values helps to decode the accumulated information based on OPC. Pixels with the same phase are segregated as a single object; however, the remaining pixels correspond to the non-projection area are referred to as zero-pixel value. In this way, HFR vision systems can accumulate frame-by-frame data, decode the embedded information to recognize globally pointed objects, simultaneously.

The communication protocol and data transaction from the VLC-based transmitter to the receiver are
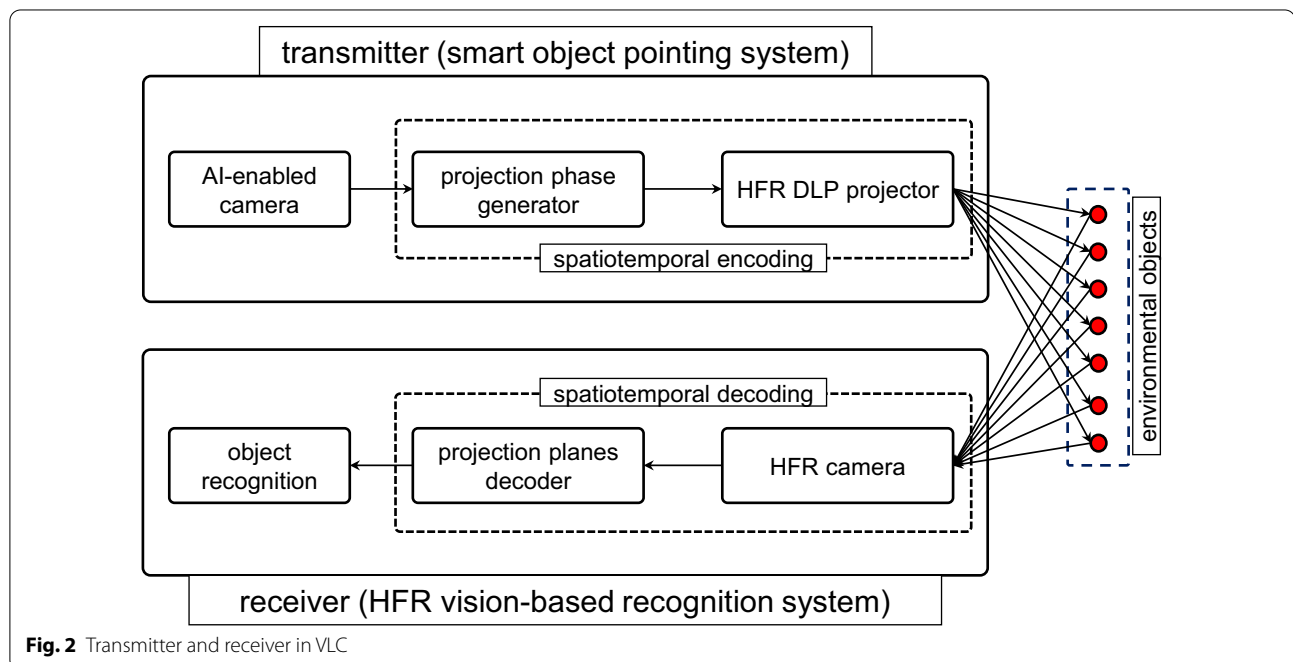

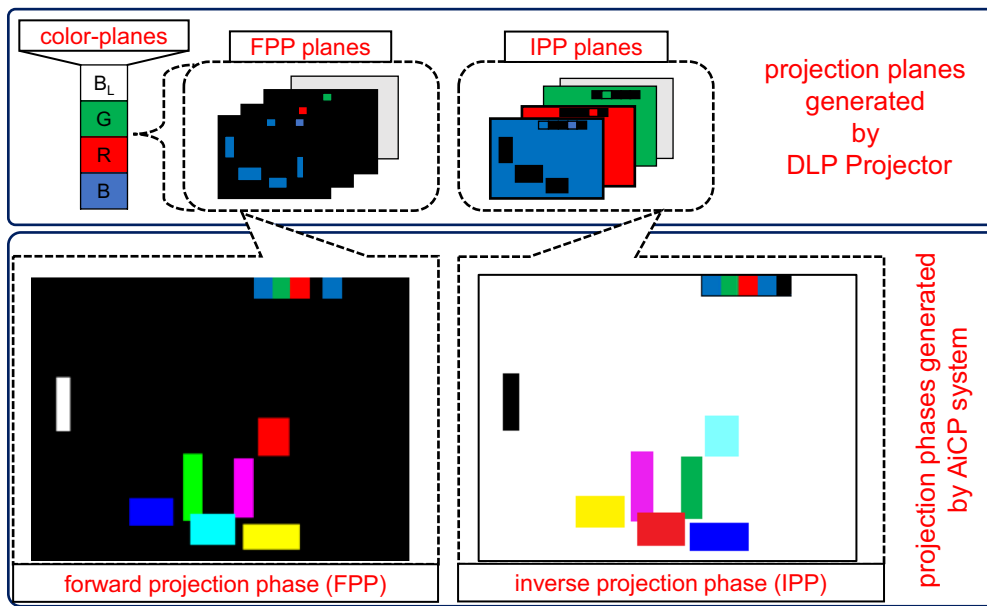**Fig. 2** Transmitter and receiver in VLC

Fig. 3 Projection phases and corresponding projection planes of AiCP system

depicted in Fig. 4. The spatially distributed information in a single detection frame is represented by a 24-bit 3-channel RGB image which is used for OPC generation. We produce two more 24-bits 3-channel RGB images representing FPP and IPP frames. Each FPP and IPP image is supplied to the DLP projector for time-varying projection. The DLP projector projects four planes of each projection phase consisting of blue, red, green, and blank planes. A total of eight 1-bit colored projection planes are required for transmitting a single detection frame. The combination of the projected eight projection-planes generates uniform gray-level brightness results in spatiotemporal encoding at VLC-based transmitter. An HFR vision system captures all the transmitted planes frame-by-frame in the same projection sequence at the receiver side. The eight 8-bit 1-channel monochrome images are binarized, weighted, and accumulated for spatiotemporal decoding. After decoding, a single 8-bit monochrome image is generated to represent the recognized objects. Thus, spatiotemporal encoding and decoding can be achieved using the VLC system of the HFR projector and camera, respectively.

Smart projection mapping and HFR vision-based recognition methodology

CNN-based object detection

The camera module of the AiCP system detects the objects in the scene using a CNN-based object detection algorithm, You Only Look Once (YOLO) [58]. It predicts the class of an object and outputs a rectangular bounding box specifying the object location at detection time δt_D. $B(I_{yolo})$ contains the bounding-boxes $bb^1; bb^2; ..; bb^N$ of the top-scored candidate class of all N-objects detected in the acquired image I_{yolo}. Each bb^n have four parameters: centroid coordinates b^n_{xc}, b^n_{yc} of the detected top-scored candidate class along with the width (b^n_w) and height (b^n_h), expressed as,

$$B(I_{yolo}(x, y, t)) = (bb^1; bb^2; ..; bb^N), \tag{1a}$$

$$bb^n = \{b^n_{xc}, b^n_{yc}, b^n_w, b^n_h\}. \tag{1b}$$

Informative masking

The FPP and IPP images are cumulatively generated based on the detected objects and their bounding box $B(I_{yolo}(x, y, t))$; they are then passed through the

Fig. 4 VLC-based encoding and decoding protocol
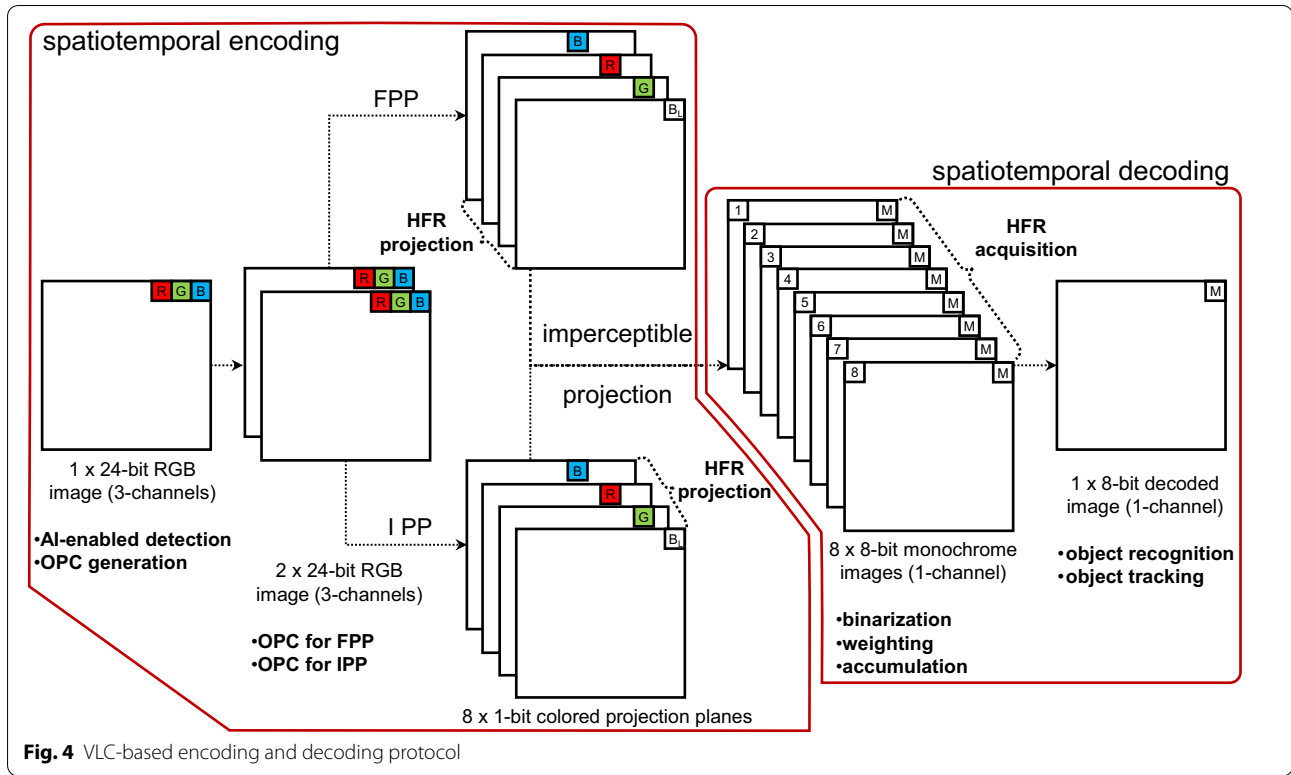
DLP projector and later projected as series of modulated colored planes on the respective objects. This is expressed as,

$$P_{fpp}(x, y, t) = \int_0^{\frac{T}{2}} \lambda_t \alpha_t(u, v, t) dt, \quad (2a)$$

$$P_{ipp}(x, y, t) = \int_{\frac{T}{2}}^{T} \lambda_t \alpha'_t(u, v, t) dt. \quad (2b)$$

$P_{fpp}(x, y, t)$ and $P_{ipp}(x, y, t)$ are the forward and inverse projection light planes that encode the information of the pointed bounding box of objects $B(I_{yolo}(x, y, t))$. The values of $P_{fpp}(x, y, t)$ and $P_{ipp}(x, y, t)$ are determined by combination of color wheel filter, temporal dithering and F_{cf} of DLP projector in place. The $\alpha_t(u, v, t)$ is DMD mirror angle in DMD mirror position (u, v) and (λ_t) is the color filter wavelength emitted at time dt. The mirror angle α_t of all DMD mirrors for P_{ipp} is always complementary to P_{fpp} for each information packet. The angle of the DMD mirror at position (u, v) is determined by image plane decided by the OPC as listed in Table 4.

Thus, the spatiotemporal packet of information is generated based on the following condition,

$$D(x, y, T) = \begin{cases} P_{fpp}(x, y, t) + P_{ipp}(x, y, t), & \text{if } T \leq \frac{1}{F_{cf}} \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

In this way, the AiCP system points the informative color mask on each detected object at high-speed while imperceptible to the HVS.

High-speed vision based recognition and localization
The HFR camera receiver's frame rate is set at 480 fps (same as P_f of HSP) for frame-by-frame decoding. The HFR camera acquires each plane of the projected packets in sequence and computes the changes in intensities of projection within the packet leading to decoding. Hence, HFR camera focuses on high-speed projector photometric properties rather than geometric calibration [13]. The nature of projected light based on the projection device principle and light reflectance from the projected surface is estimated and decoded using HFR camera.

Image acquisition
Pixel position of the projected header data is manually assigned to assist HFR camera in understanding the color filter cycle and projection sequence cycle. Images are acquired simultaneously corresponding to the projected

Kumar *et al. Robomech J*   (2021) 8:8

Page 7 of 21

image. Initially, the HFR camera looks for blue filter data with the first projection sequence emitted by the projector. The HFR camera is interfaced with a function generator to acquire images uniformly.

$$I_k(x, y, \delta_t) = D(x, y, \delta_t)s(x, y, \lambda_{\delta_t}), \tag{4}$$

where k and δ_t are the frame number and time interval of monochrome HFR camera at 480 fps. The x- and y- coordinate systems corresponds to the HFR pixel position. Note that for acquiring all the color filter segment data, δ_t is $\frac{1}{480} = 2.083\, ms$. As mentioned previously $D(m, n, \delta_t)$ is the projected information observed at duration δ_t. $s(x, y, \lambda_{\delta_t})$ is spectral reflectance from the surface of the object in the projected area at time δ_t.

Sequential thresholding
The acquired frames are sequentially thresholded by binarizing the projected and non-projected areas in the scene. The binarization of image $I_k(x, y, \delta_t))$ at time δ_t with the threshold θ is represented as,

$$B_k(x, y, \delta_t) = \begin{cases} 1, & \text{if } I_k(x, y, \delta_t) \geq \theta \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

Sequential weighting and accumulation
White pixels in the thresholded image plane are weighted based on the status of the projection and color filter sequence and accumulated a packet of information. The decoding plane $I_{dec}(x, y, T)$ at overall decoding time T is represented as,

$$I_{dec}(x, y, T) = \sum_{P_s=1}^{N_p} \sum_{C_s=0}^{C_f-1} 2^{(P_s C_s)} \int_0^T B_k(x, y, \delta_t)dt, \tag{6}$$

where N_p is the total number of projection sequences for an accumulation time T, C_s is the color sequence, and I_{dec} represents the labeled decoded information of each nonzero pixel.

Pixels of the same values are segregated based on the recognition identity; hence, the HFR vision system can decode spatiotemporally transmitted information mapped on the objects pointed by the AiCP system.

Localization of pointed objects
To determine the trajectory of each labeled object in the projection area, we calculate the zeroth and first-order moments of the I_{dec} as,

$$M_{pq}(I_{dec}(T)) = \sum_{(x,y)\epsilon I_{dec}(T)} (x^p y^q I_{dec}(x, y, T)). \tag{7}$$

The zeroth and first-order moments were used to calculate the decoded area (O_{area}) and centroid (O_{xy}) of the decoding plane ($I_{dec}(T)$) that corresponds to each object after accumulated time T,

$$O_{area}(I_{dec}(T)) = M_{00}(I_{dec}(T)), \tag{8a}$$

$$O_{xy}(I_{dec}(T)) = \left(\frac{M_{10}(I_{dec}(T))}{M_{00}(I_{dec}(T))}, \frac{M_{01}(I_{dec}(T))}{M_{00}(I_{dec}(T))} \right), \tag{8b}$$

where M_{00}, M_{01} and M_{10} are the summations of decoded pixels, x-position and y-position, respectively of the decoded regions in $I_{dec}(T)$. The decoded regions are labeled on the basis of OPC. In this way, the HFR vision system decodes visible light information and localizes the objects as $O_{xy}(I_{dec}(T))$; it determines their trajectories pointed by bounding boxes $B(I_{yolo})$ of each detected object in the AiCP encoded system.

System configuration
In this study, we used visible light as a medium to transmit information using the phenomenon of temporal dithering. The specifications of AiCP system as a transmitter and the HFR vision system as a receiver are explained as follows,

AiCP system as transmitter
As shown in Fig. 5, the prototype of AiCP system consists of USB3.0 (XIMEA MG003CGCM) VGA-resolution (640 × 480-pixels) RGB-camera head with 8.5 mm C-mount lens and high-speed DLP projector (Optoma EH503). The RGB-camera captures images at 30 fps without interfering with the flickering frequency of the DLP projector. DLP projector's frame size is 1024 × 768 with 120 Hz refresh rate for projecting color planes. It has a color wheel with equal segments of blue (B, 460 nm), red (R, 623 nm), green (G, 525 nm), and blank transparent filters. It rotates at 120 rps to generate various combinations of RGB-planes for each image as modulated and emitted color plane slices of blue, red, and green patterns. The blank transparent filter adds the overall brightness in the projected area. Thus, each color-filter plane projects at 480 Hz, which is higher than $F_c f$ of HVS. A PC with an Intel Core i7-960 CPU, 16 GB RAM running on a Windows-7 (64-bit) operating system (OS) is used for interfacing two GPUs in dual-channel 16x PCIe slots on the motherboard. The NVIDIA GTX 1080Ti (GPU-1) is used for accelerating CNN-based YOLOv3 object detection algorithm and an NVIDIA Quadro P400 (GPU-2) for accelerating video projection. The refresh rate, video synchronization (Vsync), and projection rate are synchronized with the rotating color wheel's frequency at 120 Hz. The focal-length and throw ratio of the projector-lens are set to 28.5 m and 2.0, respectively, with a maximum luminance of 5200 lux.

Fig. 5 AiCP system for informative object pointing
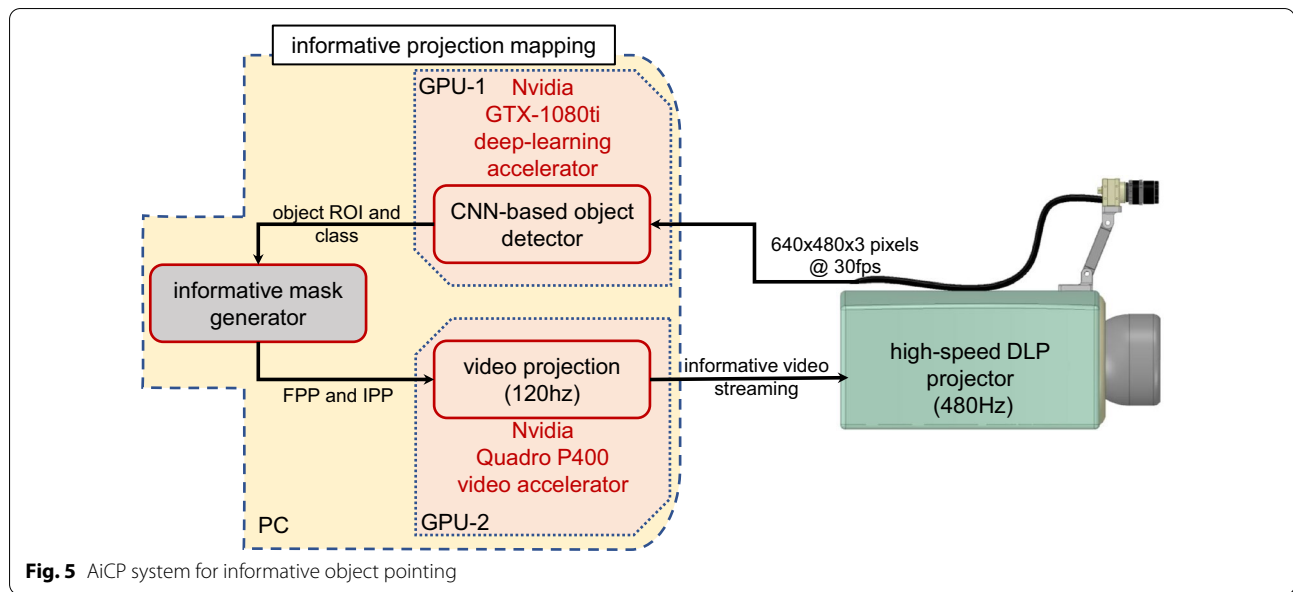
Table 1 Execution times on AiCP System (unit: ms)

	Time
(1) Image acquisition and CNN-based object detector	32.3
(2) Informative mask generator	1.02
(3) Video projection (FPP + IPP)	16.15
Total (1)–(2)	33.32

We used a CNN-based YOLO [58] algorithm to detect and localize environmental objects. GPU-1 accelerates the detection algorithm to detect and localize the objects from pre-learned models in real-time. The inference process outputs the class and ROI of detected objects based on the pre-learned weights for YOLOv3 from the 80-class COCO dataset. The ROIs are segregated based on the classification to prepare informative color masks.

A single filter of the color wheel represents a single bit of information that can be transmitted at 480 Hz. Hence, using a combination of three-color filters of the DLP projector, we can send 2^3-bits sequentially that resemble the information of a maximum of eight objects. The PC-based software generates FPP and IPP at 120 Hz for each CNN-detection frame based on the predefined OPC. The DLP projector emits corresponding color filter combinations as an informative color mask based on the fetched FPP and IPP.

The execution times of the AiCP system are listed in Table 1. The image acquisition, CNN-based object detection, and informative mask generation steps are executed in 33.32 ms. However, video projection was conducted in a separate thread to maintain the projection rate at constant intervals synchronized with a frame rate of the DLP projector for each projection-phase that is 60 fps (approx. 16 ms).

HFR vision system as a receiver

As shown in Fig. 6, the HFR vision system consists of a monochrome USB3.0 high-speed camera head (Baumer VCXU-02M) with a resolution of 640×480 pixels. The PC's specification to implement HFR image processing is, Intel Core i7-3520M CPU, 12 GB RAM with windows-7 (64-bits) OS for processing acquired images.

Initially, the header information projected by the DLP projector is inferred by the HFR vision system to synchronize with the projection sequence. Visible light decoding starts when the first and fourth blocks of the header are read as 1; that is, the blue color plane of the first sequence is acquired. Once the decoding starts, each acquired image plane is sequentially thresholded based on the presence or absence of informative masked pixels in the image. If the pixel is part of the colored mask, it is denoted as 1; otherwise, it is 0. Each thresholded image plane is then weighted based on the corresponding color and projection phases. All weighted images are accumulated by summing the 8 planes of both the projection phases. The confirmed pixels are informative; they are segregated based on the clusters and labels in the database and later localized on the image plane. Hence, the HFR vision system can sense the temporally dithered imperceptible information and decode correctly by recognizing the same objects pointed by the AiCP module simultaneously.
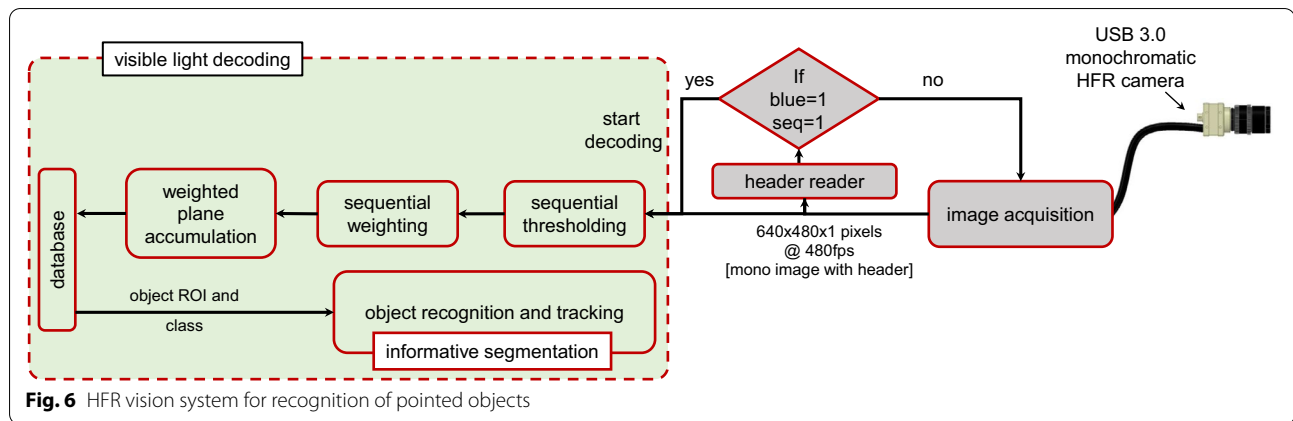
Fig. 6 HFR vision system for recognition of pointed objects

Table 2 Execution times on HFR recognition and tracking system (unit: ms)

	Time
(1) Image acquisition	0.3
(2) Header reader	0.001
(3) Sequential thresholding	0.11
(4) Sequential weighting	0.27
(5) Weighted plane accumulation	0.9
(6) Cluster and label [database]	0.057
(7) Object recognition and localization	0.002
(8) Image display	16.66
Total [(1–5) x 8]+(6)+(7)	12.707

The execution times for HFR recognition and trajectory estimation are shown in Table 2, Steps (1)–(5) are repeated eight times to buffer a packet of information that consumes 12.648 ms; followed by steps (6) and (7). The total computation time for decoding a frame of pointed objects is 12.707 ms, which is less than the information projection cycle of the AiCP system. Hence, the proposed system can recognize and estimate the trajectory of objects in real-time. The simultaneous video display is implemented in a separate thread for real-time monitoring of the object recognition.

System characterization, comparison, and confirmations
To confirm the functionalities of our algorithm, we quantify the robustness in decoding with varying lens parameters, compare the real-time execution with conventional object trackers, demonstrate the effectiveness in pointing and recognizing multiple objects at indoor and outdoor scenarios as proof of concept.

Robustness in decoding with varying lens parameters
Firstly, we quantify the robustness of our system by confirming the object recognition ability of HFR vision in terms of varying lens-aperture, lens-zoom, and lens-focus blur when three objects of the same class are pointed by AiCP system. We characterized the capability of the proposed system to observe the projection area and decode the informative color mask on each object at varying conditions in five steps. As shown in Fig. 7, three miniature cars of different sizes were placed on the linear slider with a complex background. The AiCP module put 1 m in front of the experimental scene, whereas the HFR camera at 2 m away from the scene. The HFR camera head was mounted with a C-mount 8-48 mm f/1.0 manual zoom lens. We used a linear slider to move the miniature objects to and fro in the experimental scene. Since the AiCP system has a latency due to the CNN-object detection and projection system, it may affect the mapped light onto the object like a tail behind the rapidly-moving objects. To avoid this unpleasant effect and map the IOP properly onto the objects, the linear slider was moved at 50 mm/s velocity.

As shown in Fig. 8a, the total brightness was reducing exponentially without significantly affecting the decoded area while the aperture varied from f/4.0 to f/16.0, at 16 mm focal length and 1.5 m focal depth. A variance of the Laplacian method-based lens-focus blur index was used to determine the de-focal blurring level occurring in the images. As the de-focal blur increases, the blur index reduces considerably. In this experiment, the focal depth was set to 1.5 m to acquire sharp images. As shown in Fig. 8b, the decoded area was not substantially affected at 16 mm focal length with the decreasing blur index by varying the focal depth from 1 to 7 m as marked on the lens. However, as shown in Fig. 8c, the decoded area was increasing with focal length (zoom in) from 8 mm to 36 mm due to increasing number of pixels
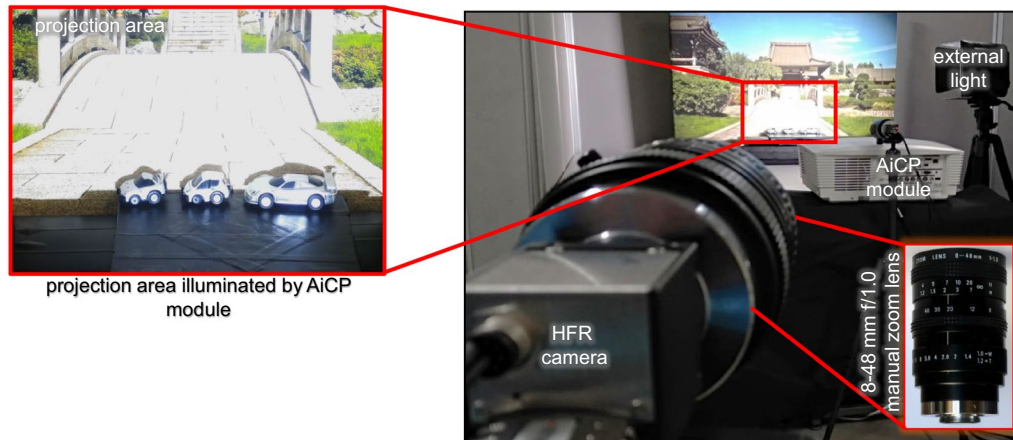
Fig. 7 Overview of the experimental setup for quantification of robustness

having informative light, at f/4 aperture and 1.5 m focus depth. It is evident that if the decoded area is visible to the HFR vision system, the recognition is always efficiently working.

Real-time execution comparison

Secondly, we compare our HFR vision-based object recognition approach with conventional object trackers to confirm computational efficiency required for real-time execution. We considered conventionally used object trackers for the real-time execution comparison. Trackers such as MOSSE [59], KCF [60], Boosting [61], and Median Flow [62] which are distributed in the OpenCV [63] standard library. Table 3 indicates that the computational cost for synthesizing 640×480 images in other methods is higher than our proposed method except for the MOSSE and Boosting object trackers. We observed that apart from the MOSSE tracker, other trackers lose the pointed objects when occlusion arises. However, our method has advantages in computational efficiency and recognition accuracy as long as the decoded area is visible to the HFR vision system. We fill up bounding boxes obtained from the YOLO object detector with the same size of SMP color masks for IOP in the projection area, affecting the object's decoded area and position based on the HFR vision observation viewpoint. Hence, the object localization may not be on the pointed object, but it is always inside the decoded area.

Indoor multi-object pointing and HFR vision-based recognition

Next, we demonstrate the effectiveness of the proposed method in multi-object pointing using the AiCP system and their recognition by distantly placed HFR vision

systems at the indoor scenario. As shown in Fig. 9, seven miniature objects were placed in a scene with a complex background. The traffic signal and clock were immovable among these objects, whereas two humans and three cars were placed on two separate horizontal linear sliders parallel to the projection plane. The AiCP module was placed at 1.5 m in front of the experimental scene. We used two HFR vision systems set at 1.5 m away from the scene to demonstrate multiple viewpoint recognition. Both the HFR cameras were mounted with 4.5 mm C-mount lenses and operated at 480 fps with 2 ms exposure time. The projection area was set to $510 \, \text{mm} \times 370 \, \text{mm}$, with 805 lux acquired luminance. The approximate length of a miniature car and the height of a miniature person was 10 cm. They were moved 200 mm by corresponding linear sliders to and fro horizontally at a speed of 50 mm/s within the projection area, which is equivalent to 7.92 km/h and 3.25 km/h for a real car and person when located 66 m and 27 m away from the AiCP system, respectively.

OPC of Seven objects for the AiCP system and recognition ID for the HFR vision system is tabulated in Table 4. We used the seven combinations of FPP and IPP for seven objects of different appearances and a single combination as a background of the projection plane. The active or inactive status of light passing through the color wheel filter based on a combination of FPP and IPP for a particular object. The experiment was conducted for 6 s; both linear sliders moved the objects one time to and fro horizontally during this period. As shown in Fig. 10, the CNN-detection frames were captured at 1 s interval, contain the detected objects marked with bounding boxes. The SMP color masks for each bounding box (that is, rectangular
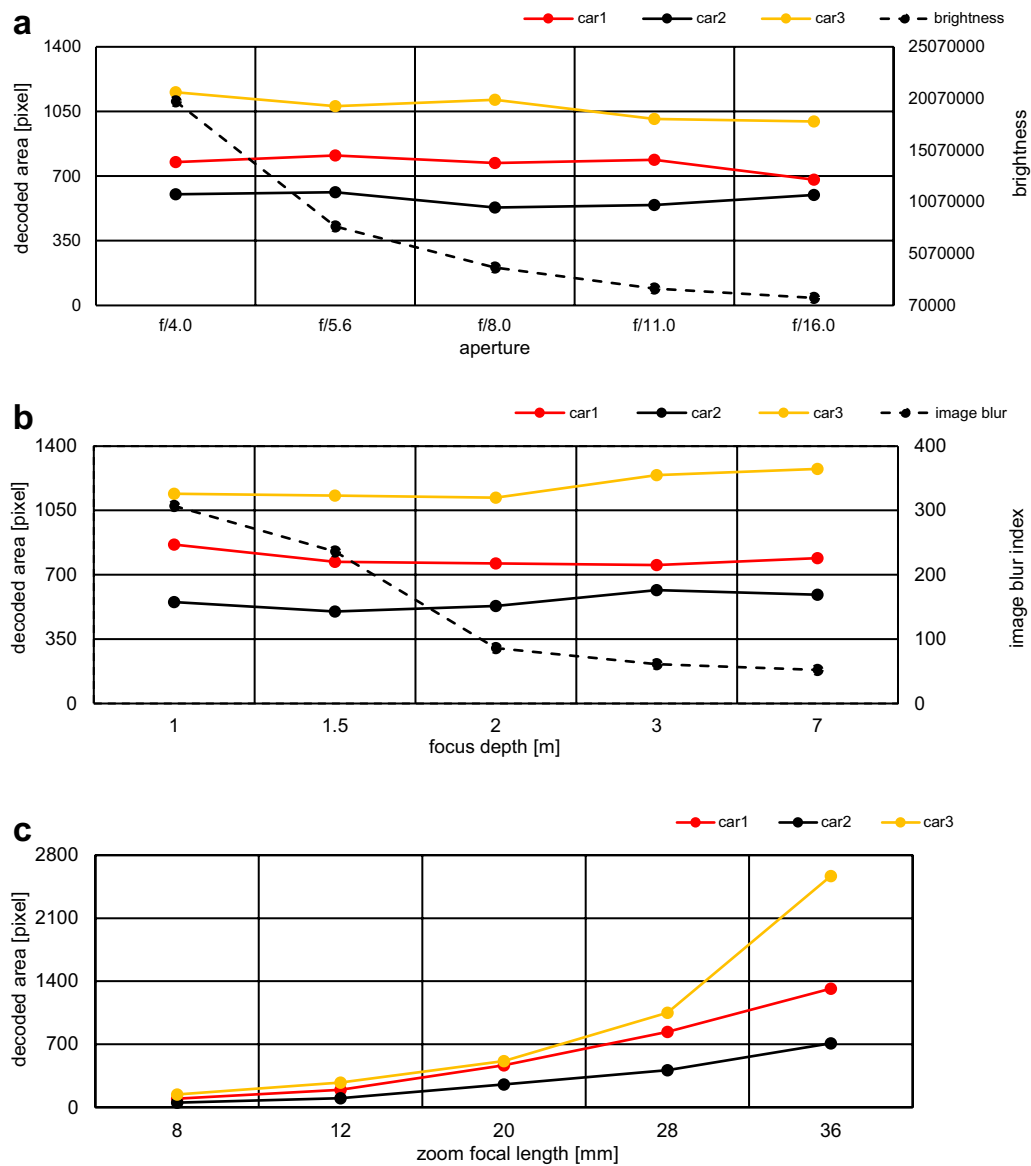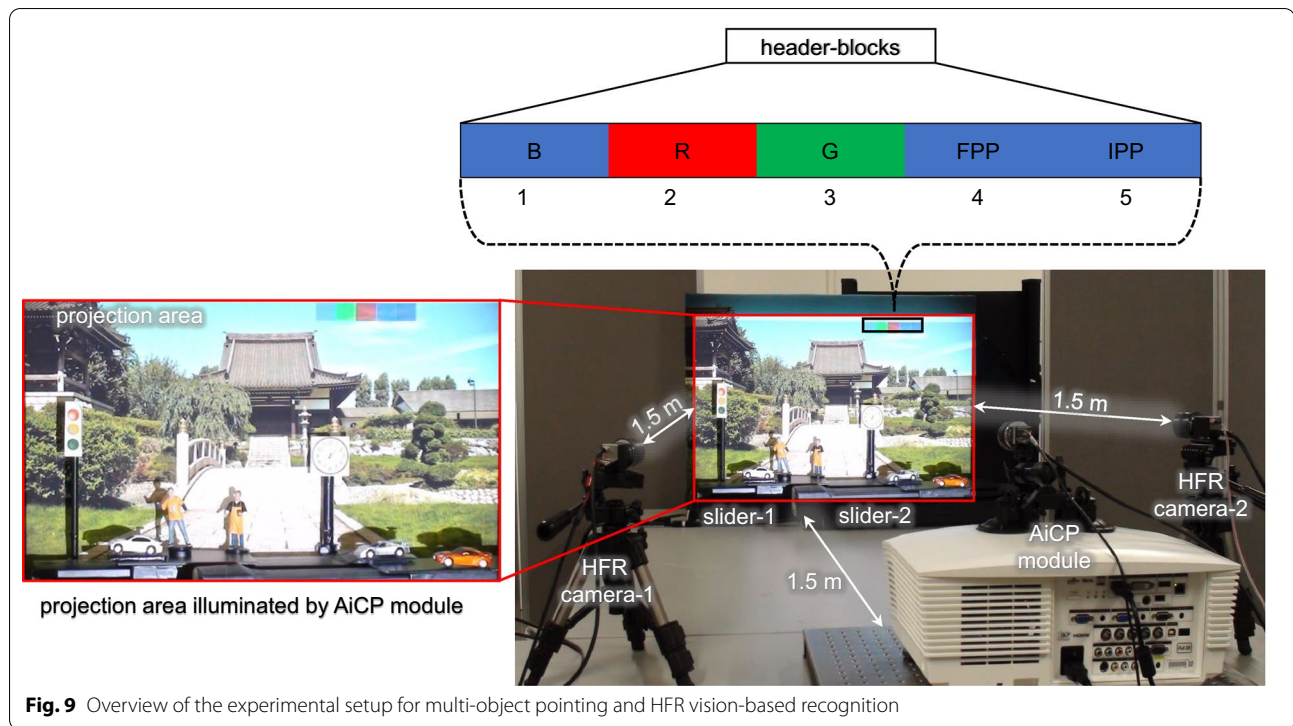
Fig. 8 Robustness in the decoded area when **a** varying aperture **b** varying focus depth, and **c** varying focal length (zoom in)

Table 3 Execution time comparison between proposed method and conventional object trackers (unit: ms)

Method	Time
(1) MOSSE	0.5
(2) KCF	19.45
(3) Boosting	8.01
(4) Median flow	40.26
(5) Our method	12.707

pointer) were projected onto the respective objects by AiCP system. Figures 11 and 12 show a color map of the decoded informative masks (left side frames) along with recognition results (right side frames) of camera-1 and camera-2 of HFR vision systems, respectively. The decoded area and displacements of each object acquired in camera-1 are plotted in Fig. 13a, b, respectively, whereas, those of camera-2 are shown in Fig. 14a, b. The decoded areas on moving objects vary in each frame. Whereas, in the case of a clock and traffic light, there is no significant change in the decoded area.

Kumar *et al. Robomech J*   (2021) 8:8

Page 12 of 21

Fig. 9 Overview of the experimental setup for multi-object pointing and HFR vision-based recognition

Table 4 Seven objects OPC for AiCP system and the recognition ID for HFR vision system

Object of interest	Informative color masks		DLP color wheel filter ('1' active '0' inactive)								Recognition ID (HFR Vision)
	FPP	IPP	Color filter for FPP				Color filter for IPP				
			blue	Red	Green	Blank	Blue	Red	Green	Blank	
Background	Black	White	0	0	0	0	1	1	1	0	14
Person 1	Green	Pink	0	0	1	0	1	1	0	0	70
Person 2	Pink	Green	1	0	0	0	0	1	1	0	16
Car 1	Red	Cyan	0	1	0	0	1	0	1	0	26
Car 2	Blue	Yellow	1	0	0	0	0	1	1	0	28
Car 3	Cyan	Red	1	0	1	0	0	1	0	0	72
Clock	Yellow	Blue	0	1	1	0	1	0	0	0	82
Traffic light	White	Black	1	1	1	0	0	0	0	0	84

They are often affected by the light absorbent properties of the material and curvatures of the objects. We observed that sometimes, due to the objects' low reflective parts, the FPP and IPP are indifferent, which affects the decoded area, and the HFR vision system does not recognize the object. We quantified the amount of communication between the AiCP system and the HFR vision system, in terms of the amount of data transmitted and received during the 6 s period. The CNN object-detector detected 7 objects per frame in 0.033 s;
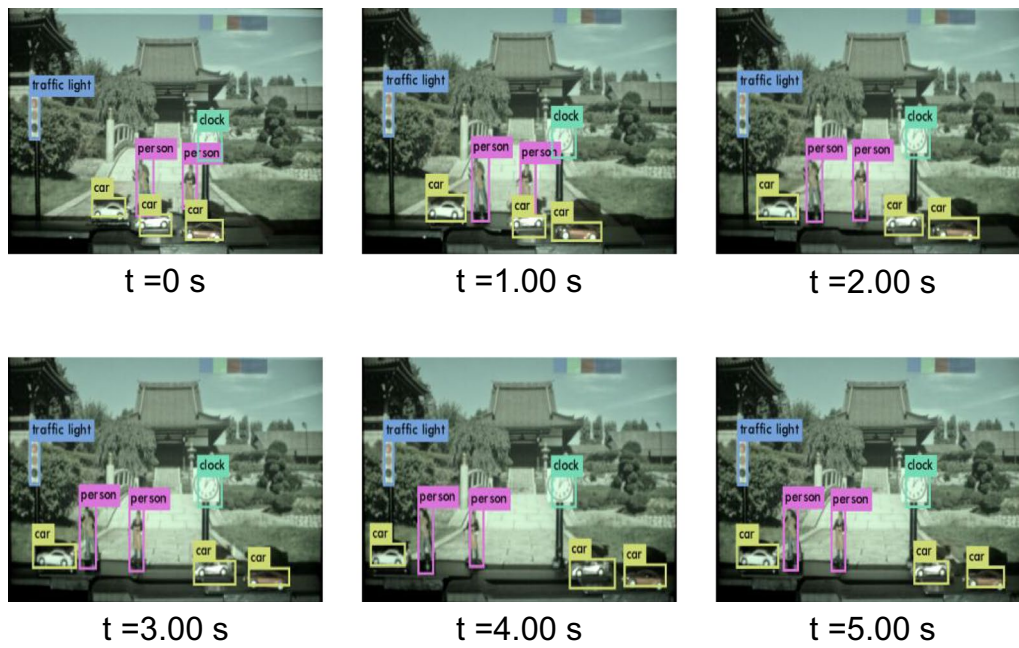
Kumar *et al. Robomech J* (2021) 8:8

Page 13 of 21

Fig. 10 CNN-object detection based multi-objects pointing by AiCP system

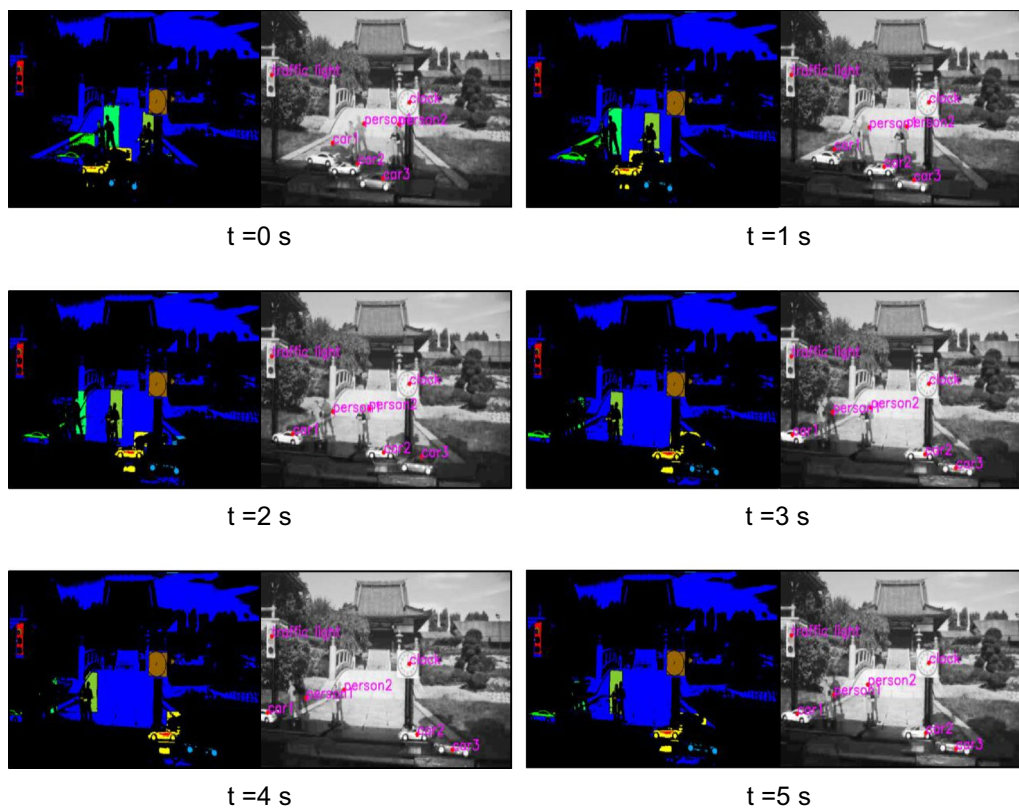**Fig. 11** Color map of decoded SMP color mask (left) and corresponding recognition (right) acquired in camera-1 of HFR vision system
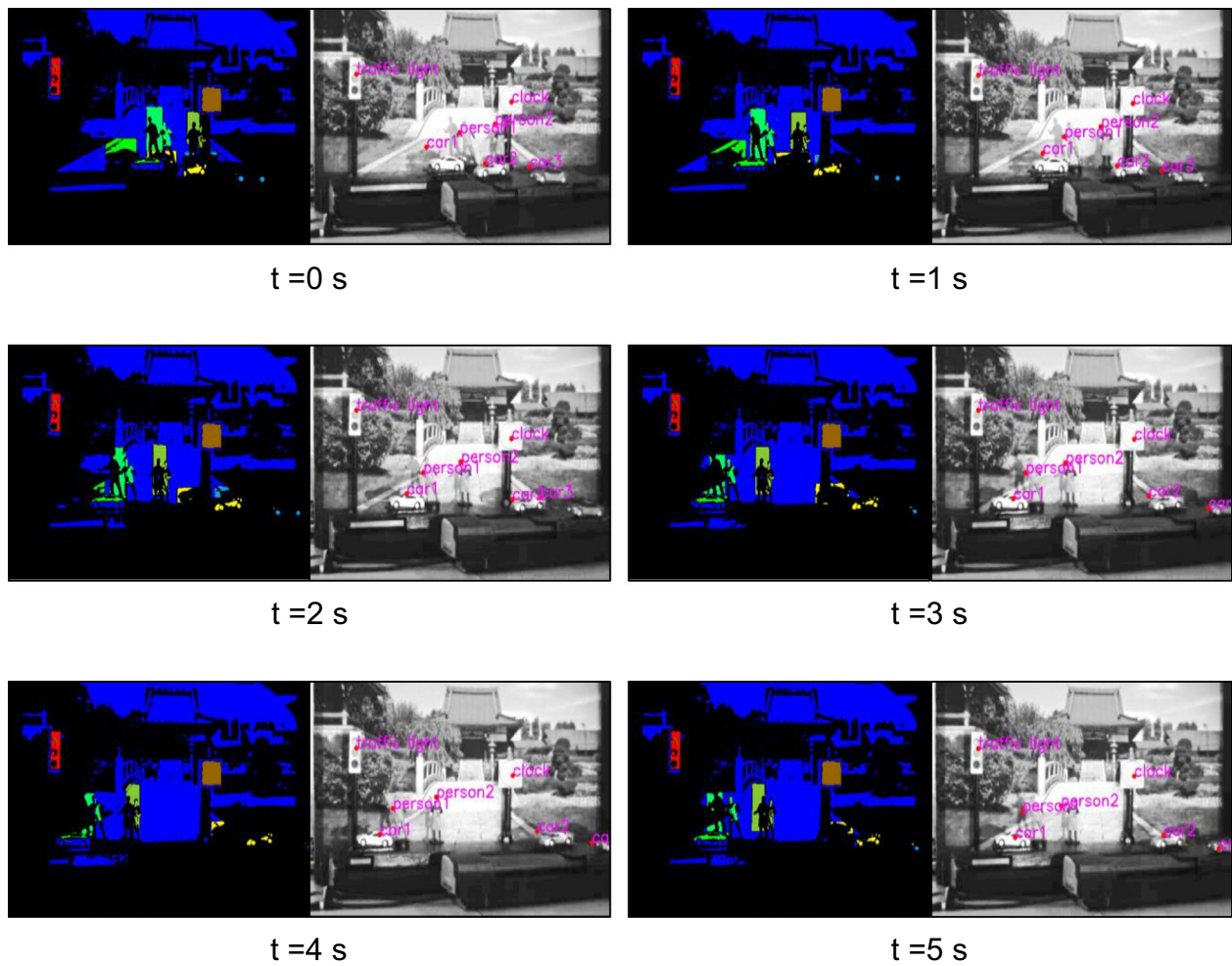
Kumar *et al. Robomech J* (2021) 8:8

Page 14 of 21

Fig. 12 Color map of decoded SMP color mask (left) and corresponding recognition (right) acquired in camera-2 HFR vision system

thus, it could detect those 7 objects approximately 180 times; the DLP projector required 2880 planes to project the information in 6 s.

Outdoor real-world object pointing and HFR vision-based recognition

We also confirmed the usefulness of the proposed method in pointing and recognizing the real-world objects for security and surveillance as an application of CPS. The experimental scene comprised the classroom as a background, a person moving at average speed

of 1000 mm/s , an umbrella, and a chair, as shown in Fig. 15. The AiCP module placed 8.5 m in front of the classroom door with 230 lux of acquired luminance and 5.75 m × 2.25 m projection area. An HFR vision system operated at 480 fps, 2 ms exposure with an 8.5 mm C-mount lens set 18.5 m away from the scene. To acquire adequate projection brightness required for decoding frame-by-frame at HFR vision system, we selected the pixel binning function with 320 × 240-pixels image resolution in HFR camera. To avoid daylight's influence on SMP projection mask projected on the pointed objects,

Kumar *et al. Robomech J*   (2021) 8:8
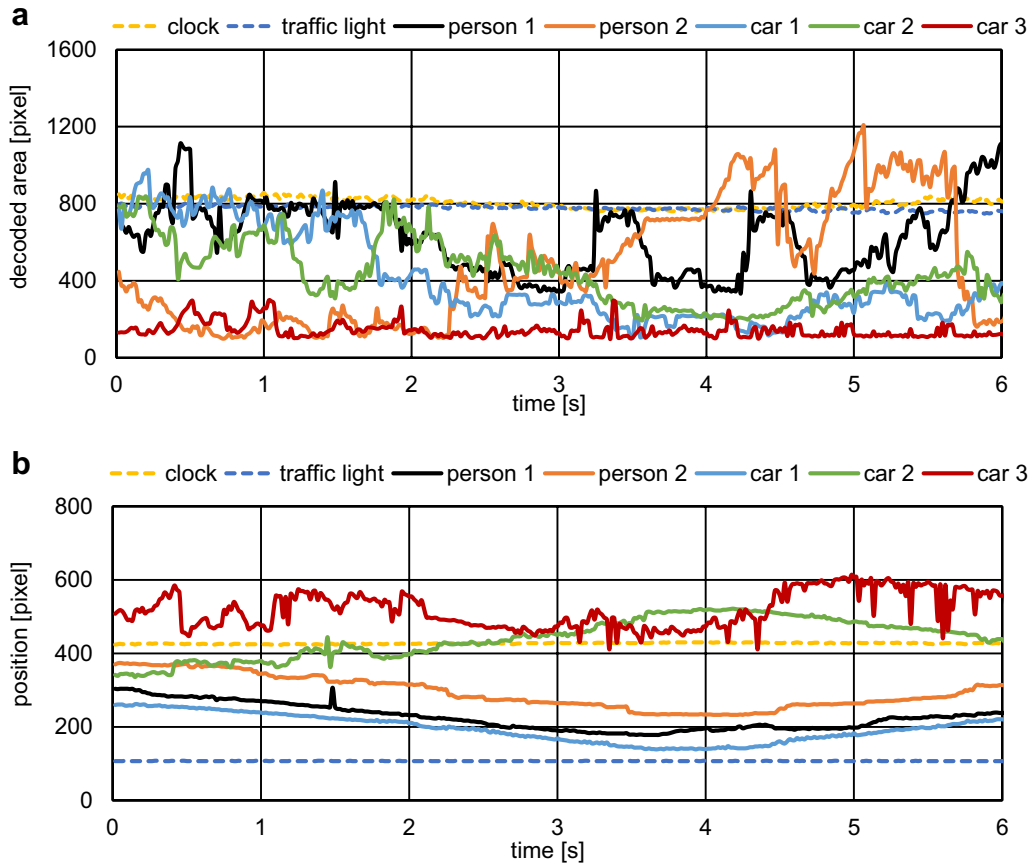
Page 15 of 21

Fig. 13 Graphs of **a** decoded area acquired, and **b** corresponding displacement of each recognized objects in camera-1 HFR vision system

we experimented using the corridor light illumination in the evening time. OPC of the multi-class object for the AiCP system and recognition ID for the HFR vision system is outlined in Table 5. During the 6 s experiment, the person was walking throughout the experiment scene and sometimes occluded chair and umbrella. The CNN-detection results are shown in Fig. 16 captured at 1 s interval.

The HFR vision system efficiently recognized the pointed person, umbrella and chair in real-time despite the limitations of the projector. We involved the background subtraction method by considering every blank sequence as a reference frame and subtracted every color sequence from it to enhance the thresholding and weighing process in HFR vision-based recognition. Thus, Eq. (4) is replaced as,

$$B_k(x,y,\delta_t) = \begin{cases} 1, & \text{if} \quad |I_k(x,y,\delta_t) - I_{ref}(x,y,\delta_{t-1}))| \geq \theta \\ 0, & \text{otherwise} \end{cases}$$

(9)

where $I_{ref}(x,y,\delta_{t-1})$ is the reference frame, that is the frame acquired from the black sequence filter from previous packet, to subtract from subsequent frames
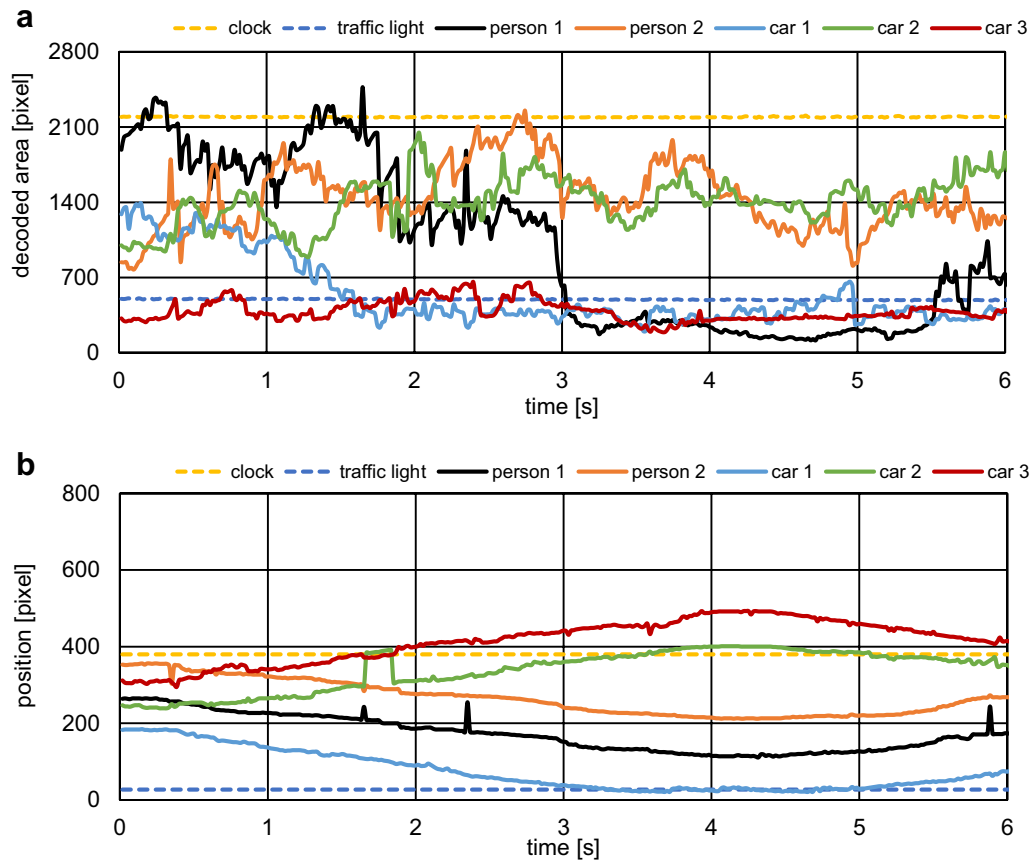
Fig. 14 Graphs of **a** decoded area acquired, and **b** corresponding displacement of each recognized objects in camera-2 HFR vision system
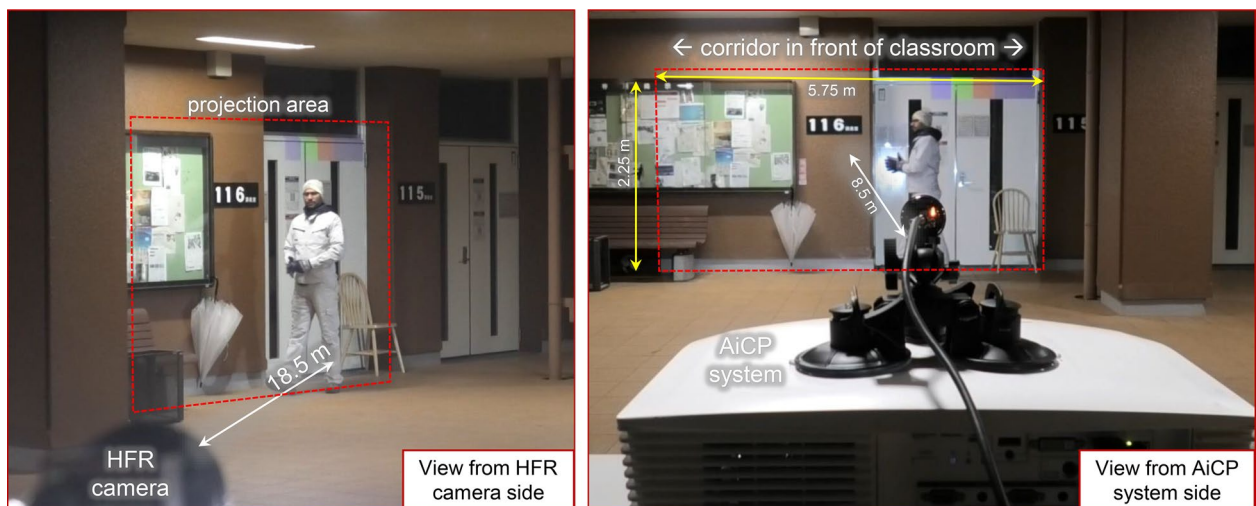

Fig. 15 Overview of the experimental setup for real-world multi-objects' pointing using AiCP module and recognition by HFR vision system

Table 5 Multi-object OPC for AiCP system and the recognition ID for HFR vision system at outdoor conditions

Object of interest	Informative color masks		DLP color wheel filter ('1' active '0' inactive)								Recognition ID (HFR Vision)
	FPP	IPP	Color filter for FPP				Color filter for IPP				
			Blue	Red	Green	Blank	Blue	Red	Green	Blank	
Background	Black	White	0	0	0	0	1	1	1	0	14
Chair	Pink	Green	1	1	0	0	0	0	1	0	16
Umbrella	Red	Cyan	0	1	0	0	1	0	1	0	26
Person	White	Black	1	1	1	0	0	0	0	0	84
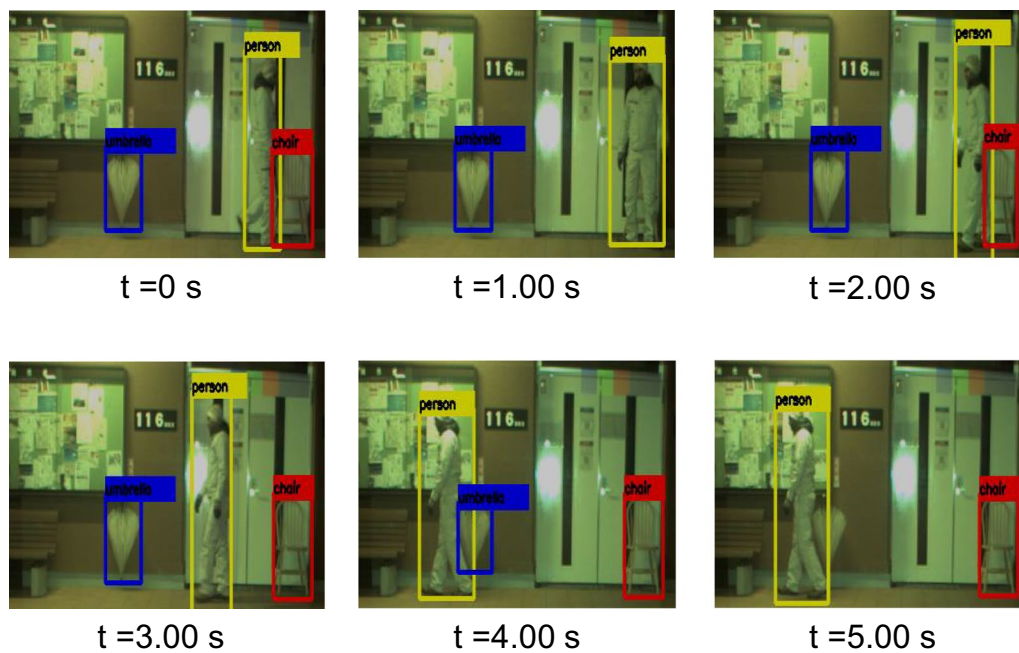
Fig. 16 Multi-objects pointed by AiCP system

corresponding to blue, red and green color filter. θ is the threshold to enhance the subtraction process. Figure 17 shows a color map of the decoded SMP masks (left side frames) and corresponding recognition results (right side frames). The decoded area and displacements of the objects obtained from the receiver are plotted in Fig. 18a and b, respectively. The decoded area is significantly larger than the miniature objects in a previous experiment, as it varies with the viewpoint of HFR vision system. In this way, we confirmed the usefulness of our proposed system for real-world scenarios.
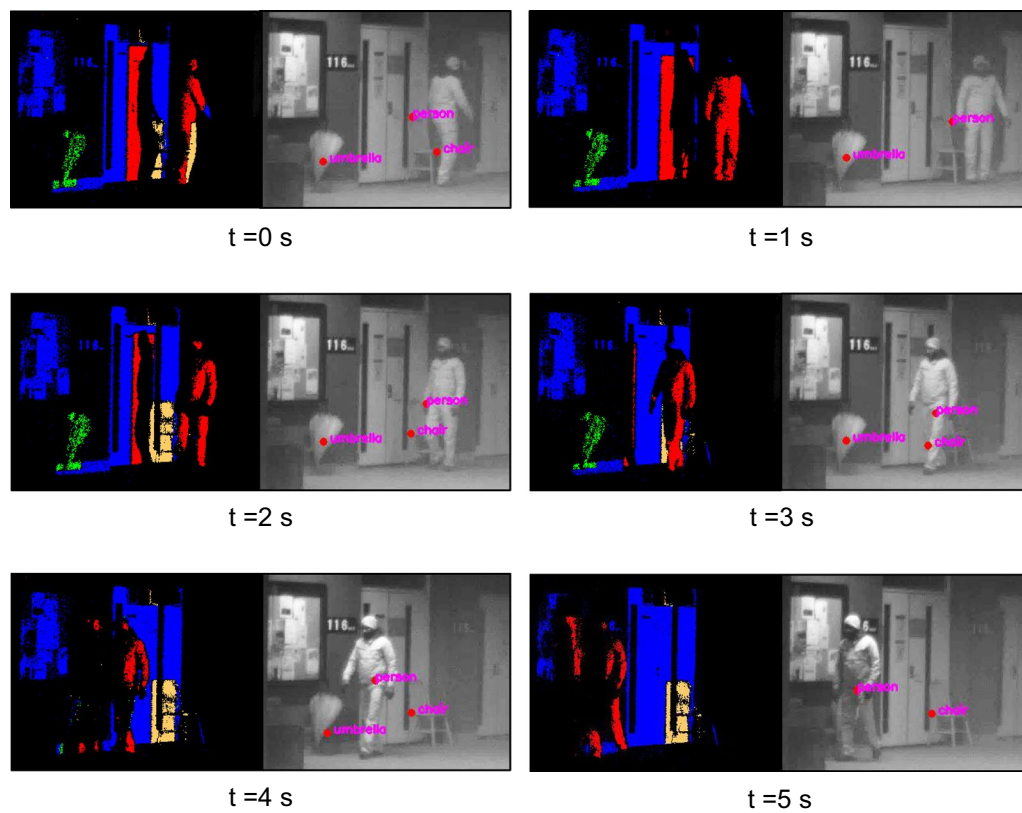
Fig. 17 Color map of decoded SMP color mask (left) and corresponding recognition (right) acquired in HFR vision system

Conclusion

In this study, we emphasize prototyping VLC based physical security application using AiCP system for object pointing and an HFR vision system for pointed object recognition. The AiCP system broadcasts CNN-object detection results by the IOP method using a high-speed projector. The distantly located HFR vision systems can perceive the information and recognize the pointed objects by decoding projected SMP from observation viewpoints. We also explained the imperceptibility to HVS using SMP based projection mapping to maintain the confidentiality of the information. We quantify the robustness in HFR vision-based recognition by varying lens-aperture, lens-zoom, and lens-focus blur and confirm the computational efficiency by comparing it with conventional object tracking methods. As a proof of concept, we demonstrated the efficiency of our approach in communicating information of multiple objects in the indoor scene using miniature objects and its usefulness in the outdoor scenario for real-world objects despite projector limitation. The localization accuracy can be improved drastically in the future using pixel-based IOP as per the contour of an object instead of bounding box-based IOP. Hence, our proposed system can be applied to real-world scenarios such as security and surveillance in vast areas, SLAM for mobile robots, and automatic driving systems. The system is currently limited for static and slowly moving objects due to the projection-latency in commercially available projectors and the reflectance properties. We intend to improve the proposed system
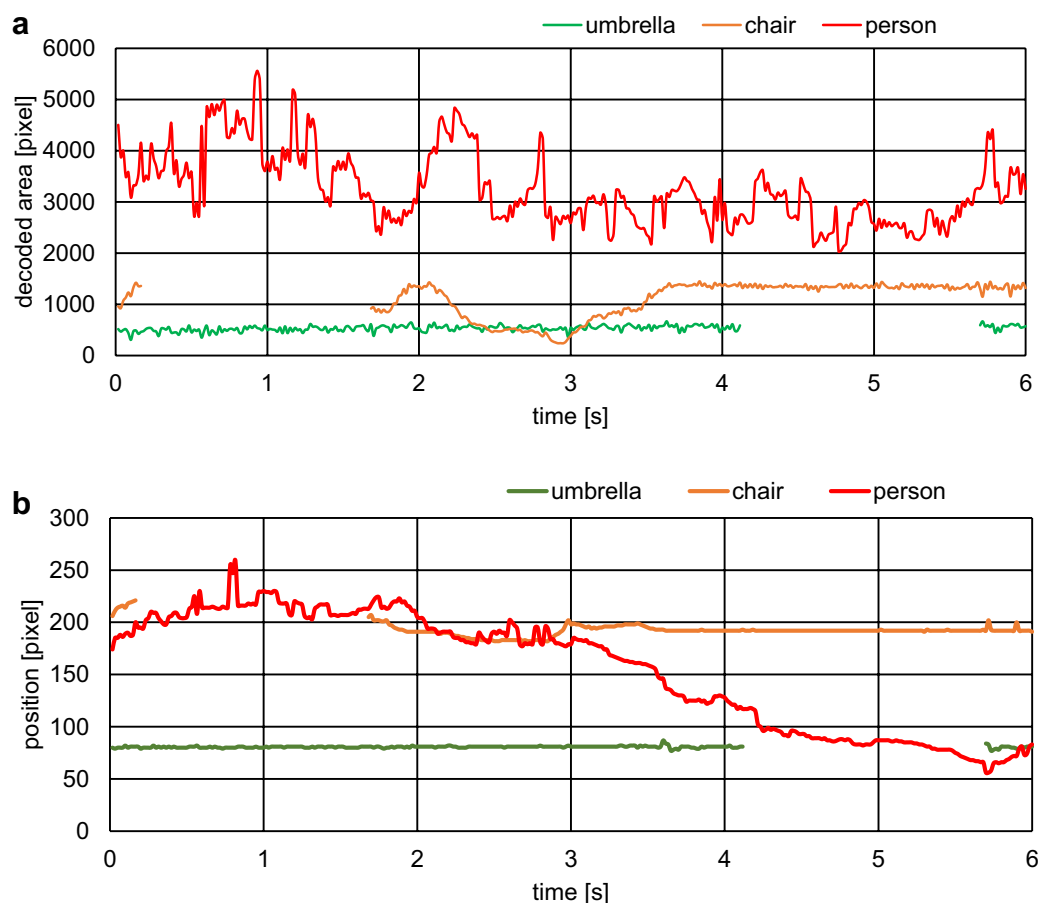
Fig. 18 Graphs of **a** decoded area acquired, and **b** corresponding displacement of single object recognized by HFR vision system

using bright and low latency high-speed projectors to recognize high-speed moving multi-objects in 3-D scenes from a long-distance during daylight.

Authors' contributions
DK carried out the main part of this study and drafted the manuscript. DK and SR set up the experimental system of this study. KS, TS, and II contributed concepts for this study and revised the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
Not applicable.

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1] Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, Japan. [2] Digital Monozukuri (Manufacturing) Education Research Center, Hiroshima University, Higashi-Hiroshima, Japan. [3] Graduate School of Advanced Science and Engineering, Hiroshima University, Higashi-Hiroshima, Japan.

References
1. Lee J, Bagheri B, Kao HA (2015) A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manuf Lett 3:18–23. https://doi.org/10.1016/j.mfglet.2014.12.001
2. Azuma T (1997) A survey of augmented reality. Presence Teleoper Virtual Environ 6(4):355–385. https://doi.org/10.1162/pres.1997.6.4.355
3. Rekimoto J, Katashi N (1995) The world through the computer: computer augmented interaction with real world environments. Proceedings of UIST.: 29-36 . https://doi.org/10.1145/215585.215639
4. Caudell Thomas P (1994) Introduction to augmented and virtual reality. SPIE Proc Telemanip Telepresence Technol. 2351:272–281. https://doi.org/10.1117/12.197320
5. Bimber O, Raskar R (2005) Spatial augmented reality, merging real and virtual worlds. CRC Press, Boca Raton

6. Mine M, Rose D, Yang B, Van Baar J, Grundhofer A (2012) Projection-based augmented reality in Disney theme parks. IEEE Comput. 45(7):32–40. https://doi.org/10.1109/MC.2012.154

7. Grossberg M D, Peri H, Nayar S K, Belhumeur P N (2004) Making one object look like another: controlling appearance using a projector-camera system. Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 1:I-I. https://doi.org/10.1109/CVPR.2004.1315067

8. Okazaki T, Okatani T, Deguchi K (2009) Shape reconstruction by combination of structured-light projection and photometric stereo using a projector-camera system. Advances in image and video technology. Lect Notes Comput Sci 5414:410–422. https://doi.org/10.1007/978-3-540-92957-4_36

9. Jason G (2011) Structured-light 3D surface imaging: a tutorial. Adv Opt Photon. 3:128–160. https://doi.org/10.1364/AOP.3.000128

10. Amit B, Philipp B, Anselm G, Anselm G, Daisuke I, Bernd B, Markus G (2013) Augmenting physical avatars using projector-based illumination. ACM Trans Graph. https://doi.org/10.1145/2508363.2508416

11. Toshikazu K, Kosuke S (2003) A Wearable Mixed Reality with an On-Board Projector. Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality. 321-322. https://doi.org/10.1109/ISMAR.2003.1240740

12. Yamamoto G, Sato K (2007) PALMbit: a body interface utilizing light projection onto palms. Inst Image Inf Telev Eng. 61(6):797–804. https://doi.org/10.3169/itej.61.797

13. Grundhöfer A, Iwai D (2018) Recent advances in projection mapping algorithms, hardware and applications. Comput Graph Forum. 37:653–675. https://doi.org/10.1111/cgf.13387

14. Kato H, Billinghurst M (1999) Marker tracking and HMD calibration for a video-based augmented reality conferencing system. Proc. IEEE/ACM Int. Workshop on Augmented Reality. 85–94. https://doi.org/10.1109/IWAR.1999.803809

15. Zhang H, Fronz S, Navab N (2002) Visual marker detection and decoding in AR systems: a comparative study. Proc Int Symp Mixed Augment Real. https://doi.org/10.1109/ISMAR.2002.1115078

16. Wagner D, Reitmayr G, Mulloni A, Drummond T, Schmalstieg D (2010) Real-time detection and tracking for augmented reality on mobile phones. IEEE Trans Vis Comput Graph 16(3):355–368. https://doi.org/10.1109/TVCG.2009.99

17. Chandaria J, Thomas GA, Stricker D (2007) The MATRIS Project: real-time markerless camera tracking for augmented reality and broadcast applications. J Real Time Image Process 2:69–79. https://doi.org/10.1007/s11554-007-0043-z

18. Hanhoon P, Jong-Il P (2005) Invisible marker based augmented reality system. Proc SPIE Visual Commun Image Process 5960:501–508. https://doi.org/10.1117/12.631416

19. Lima J, Roberto R, Francisco S, Mozart A, Lucas A, João MT, Veronica T (2017) Markerless tracking system for augmented reality in the automotive industry. Expert Syst Appl 82(C):100–114. https://doi.org/10.1016/j.eswa.2017.03.060

20. Hirotaka A, Daisuke I, Kosuke S (2015) Diminishable visual markers on fabricated projection object for dynamic spatial augmented reality. SIGGRAPH Asia 2015 Emerging Technologies, Article 7. https://doi.org/10.1145/2818466.2818477

21. Punpongsanon P, Iwai D, Sato K (2015) Projection-based visualization of tangential deformation of nonrigid surface by deformation estimation using infrared texture. Virtual Real 19:45–56. https://doi.org/10.1007/s10055-014-0256-y

22. Punpongsanon P, Iwai D, Sato K (2015) SoftAR: visually manipulating haptic softness perception in spatial augmented reality. IEEE Trans Vis Comput Graph 21(11):1279–1288. https://doi.org/10.1109/tvcg.2015.2459792

23. Narita G, Watanabe Y, Ishikawa M (2017) Dynamic projection mapping onto deforming non-rigid surface using deformable dot cluster marker. IEEE Trans Vis Comput Graph. 23(3):1235–1248. https://doi.org/10.1109/TVCG.2016.2592910

24. Guo Y, Chu S, Liu Z, Qiu C, Luo H, Tan J (2018) A real-time interactive system of surface reconstruction and dynamic projection mapping with RGB-depth sensor and projector. Int J Distrib Sensor Netw 14:155014771879085. https://doi.org/10.1177/1550147718790853

25. Lee J C, Hudson S, Dietz P (2007) Hybrid infrared and visible light projection for location tracking. In: Proceedings of the Annual ACM Symposium on User Interface Software and Technology. pp 57–60. https://doi.org/10.1145/1294211.1294222

26. Hornbeck L J (1995) Digital Light Processing and MEMS: Timely Convergence for a Bright Future. Plenary Session, SPIE Micromachining and Microfabrication

27. Younse JM (1995) Projection display systems based on the digital micromirror device (DMD). SPIE Conf Microelectr Struct Microelectromech Dev Opt Process Multimedia Appl. 2641:64–75. https://doi.org/10.1117/12.220943

28. Heinz M, Brunnett G, Kowerko D (2018) Camera-based color measurement of DLP projectors using a semi-synchronized projector camera system. Photo Eur 10679:178–188. https://doi.org/10.1117/12.2307119

29. Dudley D, Walter M, John S (2003) Emerging digital micromirror device (DMD) applications. Proc SPIE, MOEMS Disp Imaging Syst. 4985:14–25. https://doi.org/10.1117/12.480761

30. Oliver B, Iwai D, Wetzstein G, Grundhöfer A (2008) The visual computing of projector-camera systems. In ACM SIGGRAPH 2018 classes 84:1–25. https://doi.org/10.1145/1401132.1401239

31. Takei J, Kagami S, Hashimoto K (2007) 3,000-fps 3-D Shape measurement using a high-speed camera-projector system. Proc. 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems. 3211-3216. https://doi.org/10.1109/IROS.2007.4399626

32. Kagami S (2010) High-speed vision systems and projectors for real-time perception of the world.2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops. pp 100-107. https://doi.org/10.1109/CVPRW.2010.5543776

33. Gao H, Aoyama T, Takaki T, Ishii I (2013) Self-Projected Structured Light Method for Fast Depth-Based Image Inspection. Proc. of the International Conference on Quality Control by Artificial Vision pp 175-180

34. Cotting D, Näf M, Gross M, Fuchs H (2004) Embedding Imperceptible Patterns into Projected Images for Simultaneous Acquisition and Display. Third IEEE and ACM International Symposium on Mixed and Augmented Reality. pp 100-109. https://doi.org/10.1109/ISMAR.2004.30

35. Raskar R, Welch G, Cutts M, Lake A, Stesin L, Fuchs H (1998) The office of the future: a unified approach to image-based modeling and spatially immersive displays. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques (SIGGRAPH '98). Association for Computing Machinery, pp 179–188. https://doi.org/10.1145/280814.280861

36. Watson AB (1986) Temporal sensitivity. Handbook Percept Hum Perform. 1:6.1–6.43

37. Hewlett G, Pettitt G (2001) DLP Cinema™ projection: a hybrid frame-rate technique for flicker-free performance. J Soc Inf Disp 9:221–226. https://doi.org/10.1889/1.1828795

38. Fofi D, Sliwa T, Voisin Y (2004) A comparative survey on invisible structured light. Mach Vision Appl Ind Inspect XII. 5303:90–98. https://doi.org/10.1117/12.525369

39. Siriborvornratanakul T, Sugimoto M (2010) ipProjector: designs and techniques for geometry-based interactive applications using a portable projector. Int J Dig Multimed Broadcast 2010(352060):1–12. https://doi.org/10.1155/2010/352060

40. Siriborvornratanakul T (2018) Enhancing user experiences of mobile-based augmented reality via spatial augmented reality: designs and architectures of projector-camera devices. Adv Multimed. 2018:1–17. https://doi.org/10.1155/2018/8194726

41. Chinthaka H, Premachandra N, Yendo T, Panahpour M , Yamazato T, Okada H, Fujii T, Tanimoto M (2011) Road-to-vehicle Visible Light Communication Using High-speed Camera in Driving Situation. pp 13-18. Forum on Information Technology. https://doi.org/10.1109/IVS.2008.4621155

42. Kodama M, Haruyama S (2016) Visible light communication using two different polarized DMD projectors for seamless location services. In: Proceedings of the Fifth International Conference on Network, Communication and Computing. pp 272–276. https://doi.org/10.1145/3033288.3033336

43. Zhou L, Fukushima S, Naemura T (2014) Dynamically Reconfigurable Framework for Pixel-level Visible Light Communication Projector. 8979:126–139. https://doi.org/10.1117/12.2041936

44. Kodama M, Haruyama S (2017) A fine-grained visible light communication position detection system embedded in one-colored light using DMD projector. Mob Inf Syst. https://doi.org/10.1155/2017/970815

Kumar *et al. Robomech J* (2021) 8:8

Page 21 of 21

45. Ishii I, Taniguchi T, Sukenobe R, Yamamoto K (2009) Development of high-speed and real-time vision platform, H3 vision. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. pp 3671-3678. https://doi.org/10.1109/IROS.2009.5354718

46. Ishii I, Tatebe T, Gu Q, Moriue Y, Takaki T, Tajima K (2010) 2000 fps Real-time Vision System with High-frame-rate Video Recording. Proc. IEEE Int. Conf. Robot. Autom., pp 1536–1541. https://doi.org/10.1109/ROBOT.2010.5509731

47. Yamazaki T, Katayama H, Uehara S, Nose A, Kobayashi M, Shida S, Odahara M, Takamiya K, Hisamatsu Y, Matsumoto S, Miyashita L, Watanabe Y, Izawa T, Muramatsu Y, Ishikawa M (2017) A 1ms high-speed vision chip with 3D-stacked 140GOPS column-parallel PEs for spatio-temporal image processing. IEEE International Solid-State Circuits Conference (ISSCC), pp 82–83

48. Sharma A, Shimasaki K, Gu Q, Chen J, Aoyama T, Takaki T, Ishii I, Tamura K, Tajima K (2016) Super High-Speed Vision Platform That Can Process 1024x1024 Images in Real Time at 12500 Fps. Proc. IEEE/SICE International Symposium on System Integration. pp 544–549. https://doi.org/10.1109/SII.2016.7844055

49. Okumura K, Oku H, Ishikawa M (2012) Lumipen: Projection-Based Mixed Reality for Dynamic Objects. IEEE International Conference on Multimedia and Expo. pp 699–704. https://doi.org/10.1109/ICME.2012.34

50. Chen J, Yamamoto T, Aoyama T, Takaki T, Ishii I (2014) Simultaneous projection mapping using high-frame-rate depth vision. IEEE International Conference on Robotics and Automation (ICRA). pp 4506-4511. https://doi.org/10.1109/ICRA.2014.6907517

51. Narasimhan S, Koppal, Sanjeev J, Yamazaki S (2008) Temporal Dithering of Illumination for Fast Active Vision Computer Vision – ECCV. pp 830–844. https://doi.org/10.1007/978-3-540-88693-8_61

52. Chen L, Yang H, Takaki T, Ishii I (2012) Real-time optical flow estimation using multiple frame-straddling intervals. J Robot Mechatron 24(4):686–698. https://doi.org/10.20965/jrm.2012.p0686

53. Ishii I, Tatebe T, Gu Q, Takaki T (2012) Color-histogram-based Tracking at 2000 fps. J Electr Imaging. 21(1):1–14. https://doi.org/10.1117/1.JEI.21.1.013010

54. Ishii I, Tatebe I, Gu Q, Takaki T (2011) 2000 fps real-time target tracking vision system based on color histogram. Proc SPIE. 7871:21–28. https://doi.org/10.1117/12.871936

55. Ishii I, Ichida T, Gu Q, Takaki T (2013) 500-fps face tracking system. J Real-Time Image Proc. 8(4):379–388. https://doi.org/10.1007/s11554-012-0255-8

56. Gu Q, Raut S, Okumura K, Aoyama T, Takaki T, Ishii I (2015) Real-time image Mosaicing system using a high-frame-rate video sequence. J Robot Mechatron. 27(1):12–23. https://doi.org/10.20965/jrm.2015.p0012

57. S Raut, Shimasaki K, Singh S, Takaki T, Ishii I (2019) Real-time high-resolution video stabilization using high-frame-rate jitter sensing. Robomech J. https://doi.org/10.1186/s40648-019-0144-z

58. Redmon J, Farhadi A (2018) YOLOv3: An incremental improvement. CoRR. arXiv:1804.02767

59. Bolme D S, Beveridge J R, Draper B A, Lui Y M (2010) Visual object tracking using adaptive correlation filters. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp 2544-2550. https://doi.org/10.1109/CVPR.2010.5539960

60. Henriques JF, Caseiro R, Martins P, Batista J (2015) High-speed tracking with kernelized correlation filters. IEEE Trans. Pattern Anal Mach Intell. 37:583–596. https://doi.org/10.1109/TPAMI.2014.2345390

61. Grabner H, Grabner M, Bischof H (2006) Real-time tracking via on-line boosting. Proc Br Mach Vision Confer. 1:47–56. https://doi.org/10.5244/C.20.6

62. Kalal Z, Mikolajczyk K, Matas J (2010) Forward-backward error: automatic detection of tracking failures. In: Proceedings of the International Conference on Pattern Recognition. pp 2756–2759. https://doi.org/10.1109/ICPR.2010.675

63. https://opencv.org/

Publisher's Note