RESEARCH ARTICLE

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*]

*Correspondence:
iishii@hiroshima-u.ac.jp
[1]Graduate School of Engineering, Hiroshima University, Higashi-Hiroshima, Japan
Full list of author information is available at the end of the article
†deepak@robotics.hiroshima-u.ac.jp, sushil@hiroshima-u.ac.jp, simasaki@hiroshima-u.ac.jp, taku-senoo@hiroshima-u.ac.jp

Abstract

An informative object pointing method using a spatiotemporal-modulated pattern projection is proposed to recognize and localize pointed objects by using a distantly located high-frame-rate vision system. We developed a prototype for projection-mapping-based object pointing that consists of an AI-camera-enabled projection (AiCP) system used as a transmitter, for informative projection mapping, and an HFR vision system operated as a receiver. The AiCP system detects multiple objects in real time at 30 fps with a CNN-based object detector, and simultaneously encodes and projects the recognition results of the detector as 480-Hz-modulated light patterns on to the objects to be pointed. The multiple 480-fps cameras can directly recognize and track the objects pointed at by the AiCP system without camera calibration or complex recognition methods by decoding the brightness signals of pixels in the images. To demonstrate the effectiveness of our proposed method, several desktop experiments using miniature objects and scenes were conducted under various conditions.

Keywords: High-Speed Camera-Projector; Informative Projection Mapping; Real-Time Pattern Recognition; Visible Light Communication

Introduction

Projection mapping is an augmented reality (AR) approach that is used in the entertainment and gaming industries as a surface-oriented video projection system. AR approaches are used in projecting realistic videos onto desired surfaces to accommodate boundaries for augmenting artistic effects [1, 2, 3]; they are widely used for visual augmentation in places such as buildings, rooms, and parks [4, 5, 6]. Projection-mapping-based AR systems are mainly classified as static and dynamic projection mapping. Static projection mapping (SPM) is usually preferred in industries and scientific research for shape analysis by projecting patterns as structured light [7, 8]. It involves the projection of light patterns by manually aligning the objects and projector source [9, 10, 11, 12]. In the case of dynamic projection mapping (DPM), a system tracks the position and shape of the desired surfaces using a marker [13, 14, 15] and model [16, 17, 18] tracking methods to project videos onto the moving surfaces. Asayama et al. [19] proposed visual markers for the projection of dynamic spatial AR on fabricated objects. DPM requires heavy computation to acquire dynamic, realistic effects in real time [20, 21]. Narita et al. [**?**] explained the use of 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 also reported [23] for developing an interactive system of surface reconstruction.

Several approaches to projection mapping have been implemented using a non-intrusive and imperceptible pattern for human visual systems (HVSs). Lee et al. [24] 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 frames per second (fps)[25, 26, 27, 28]. These have been applied in structured-light-based three-dimensional (3D) sensing, interactive projection mapping, and other geometric and photometric applications [29, 30, 31, 32]. Daniel et al. [33] presented a simultaneous acquisition and display method that can embed imperceptible patterns in projected images. In their system, high-speed switching between the projected pattern and its complementary pattern with DLP is used; this results in imperceptibility to the HVS. However, the resultant projection leads to lower brightness and requires modifications to the hardware [34]. High-speed projection systems that can emit light at a higher frequency than the HVS have been used in AR applications to maintain imperceptibility. The projection patterns and their complementarity at 120 Hz are sufficient to generate undetected projections to the HVS [35, 36, 37]. Color-wheel filter-based 3D projectors with the DLP principle can emit 120-Hz color-plane patterns [38, 39].

Visible light communication (VLC) was used to establish a wireless link between projection and sensing systems to transmit anticipated information [40, 41, 42]. Kodama et al. [43] designed a VLC position detection system embedded in one-colored light using a DMD projector. They used photodiodes as sensors to decode the projected area location for an IoT application. 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; this leads to a severe loss of temporal information. Hence, an HFR vision system to sense temporal alterations in high-speed projection data is required.

HFR vision systems with millisecond-level accuracy operate at hundreds or thousands of frames per second and have been developed for various industrial applications [44, 45, 46, 47]. For example, a saccade mirror and HFR cameras have been used to add visual information in real time for projection-based mixed reality of dynamic objects[48]. An HFR camera–projector depth vision system has been used for simultaneous projection mapping in which the RGB light patterns are augmented on 3D objects by computing the depth using an HFR camera projector system [49]. Temporal dithering of high-speed illumination for fast active vision was reported [50]. An HFR camera–projector-based structured-light range finding was used to eliminate common problems. Because of their high sampling rate, HFR vision systems have been used as sensing devices in many applications such as optical flow [51], color histogram-based cam-shift tracking [52, 53], face tracking [54], image mosaicking, and stabilization [55, 56]. HFR vision systems have also been used in dynamic sensing applications including the scratching behavior analysis of laboratory mice [57, 58], cell deformation analysis [59, 60, 61], vision-assisted structure analysis [62, 63], motion-blur-free video shooting [64, 65], robot manipulation [66, 67], and multicopter tracking [68, 69].

In this study, we propose an active projection mapping system that can point smart (i.e., object-oriented informative) markers onto objects. Such markers can only be detected and further decoded by a remotely located HFR vision system while maintaining the imperceptibility for HVS. The rest of the paper is constructed as follows. The concept of imperceptible AI-camera-enabled projection (AiCP) and HFR vision-based object pointing by VLC are discussed in the concept section. In the algorithm section, the mathematical expressions are stated for parts 1 and 2 of the proposed system. The resources and their specifications are reported in the system configuration section. We demonstrate the functionality of the implemented system and its robustness in the experiment section. Finally, in the conclusion section, we discuss achievements and future work.

Concept

The projections are temporally periodic throughout the projection area; however, generating the unique spatiotemporally modulated imperceptible patterns for each dynamic object is a challenging task. Dynamic projection is usually guided by a vision system to identify the type of the objects or their geometry. As illustrated in Figure 1, we propose an active projection mapping system that consists of three parts, (1) a smart object pointing system, (2) an HFR vision-based object recognition system, and (3) encoding and decoding communication protocol.

Smart Object Pointing System

The AiCP is a smart object pointing system assisted by an Ai-enabled vision system; it detects the objects in its view area and generates an object pointing code (OPC) for the high-speed DLP projector, as shown in Figure 1. We applied a CNN-based object detection method for efficient detection and localization of the environmental objects. The object detection system outputs the classes of objects of interest and their region of interest (ROI) in real-time. Then, a unique spatiotemporal-modulated color mask is generated based on OPC for each detected object.

Various studies reported that under most conditions, HVS could not resolve rapid visual changes beyond the critical frequency F_{cf} of 60 Hz. 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 pulses of light at variable time intervals to obtain the desired light intensity. The smart object pointing system transmits informative light using pulse-width modulation by the projector based on OPC, which is unique for every input intensity, known as temporal dithering of the illumination.

HFR Vision-based Recognition

An HFR vision system of an equivalent frame rate as the AiCP system can perceive temporal changes (i.e., photometric properties) in the projection area. Hence, in the proposed VLC system, HFR vision was used for recognizing the rapidly transmitted information and track the pointed objects accordingly. The OPC is a packet of information in the form of a time-varying color mask on the pointed objects. Both

systems should be synchronized to know the start and end of a packet of information projected on each pointed object and to understand the temporal changes in the sequences of an informative color mask on each desired object. In this study, an HFR vision system is distantly located without any wired synchronization and acts as a receiver of informative light. We synchronized both systems optically by projecting a spatial header with the status of the projection plane. Hence, any number of remotely located HFR vision systems can recognize the pointed objects using computationally efficient methods.

Encoding and decoding protocol

The functional blocks of the VLC-based transmitter and receiver are shown in Figure 2. The AiCP system as a transmitter consists of an AI-enabled camera and a spatiotemporal encoding system. As receiver, the HFR vision system is utilized as a spatiotemporal decoding system. A color-wheel-based high-speed single-chip DLP projector has a light passing-color wheel consisting of spectral distributions with segments of blue filter (B, 460 nm), red filter (R, 623 nm), green filter (G, 525 nm) and 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. In this proposed study, the factor of imperceptibility is possible when a packet of N-light planes has spatiotemporal $N/2$ color planes and the remaining $N/2$ are its inverse planes in sequential order. This phenomenon can be implemented to transmit various imperceptible information within the visible light spectrum. Thus, owing to temporal dithering, the mapped informative color masks as a pointer should be visualized as a uniform and non-flickering light source to the HVS as well as to the conventional 25 to 30 fps vision systems. As illustrated in Figure 3, we encode the information in two phases of the projection as 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 complimentary color mask onto the objects to be pointed. The combination of color masks in FPP and IPP are chosen in such way that the accumulated light is visualized as a uniform bright gray level light within the projection area. At the other end, the information encoded by a smart object pointing system is only decoded when each projected plane of a DLP projector is captured at same rate of projection sequentially. An HFR vision system with equivalent frame rate is utilized for decoding. Each pixel in an image corresponds to the projection area is observed temporally by the HFR vision system. The amplitude determined by variation in the brightness of planes of the informative color mask represent the information of pointed objects. All patterns in the projection area have the same high projection frequency; however, variation in phase values are based on OPC. Pixels with same phase are segregated as a single object; the remaining pixels correspond to the non-projection area is referred to as zero-pixel value. In this way, HFR vision systems can accumulate frame-by-frame data, decode the embedded information, recognize globally pointed objects, and track them simultaneously.

The communication protocol and data transaction from the VLC-based transmitter to the receiver are depicted in Figure 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 in a sequence. 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 are imperceptible for HVS. In this way, we achieve spatiotemporal encoding at VLC-based transmitter. At receiver side, an HFR vision system captures all the transmitted planes frame-by-frame in a same sequence of projection. 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 objection that is further used for object trajectory estimation. Thus, spatiotemporal encoding and decoding can be achieved using the VLC system of the HFR projector and camera, respectively.

Smart projection mapping 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) [70]. 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 of the top-scored candidate class of all N-objects detected in the acquired image I_{yolo} and bb^n. There are 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 its bounding box $B(I_{yolo}(x,y,t))$; they are then passed through the DLP projector and later projected as modulated colored planes on the respective objects. This is expressed as,

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

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

$P_{fpp}(x,y,t)$ 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 the DMD mirror angle in position (u, v) and (λ_t) is the color filter wavelength emitted at time dt. The mirror angle α_t of 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,

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

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

High-speed vision based recognition and trajectory estimation algorithm

To decode the projected visible light data, the HFR camera recipients frame rate is set at 480 fps(same as P_f of HSP). 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 photometric calibration of high-speed projector rather than geometric calibration [12]. That is, the geometric parameters of objects where projection occurs is not considered. Rather, the nature of projected light based on projection device principle and light reflectance from the surface of projected area 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 projected image. Initially, the HFR looks for blue filter data with first projection sequence emitted by the projector. The HFR camera is interfaced with a function generator in order 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 averaged 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,

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

Sequential weighting and accumulation

The white pixels in the thresholded image plane are weighted based on the status of the projection and color filter sequence and accumulated to 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 non-zero pixel.

Pixels of the same values are segregated based on the recognition identity in the database; 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,

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

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

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 further 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 AiCP system as a transmitter and the HFR vision system as a receiver are explained as follows,

AiCP system as transmitter

As shown in Figure 5, the AiCP system comprises of AI-enabled 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. A DLP projector with a frame size 1024 × 768 with a refresh rate of 120 Hz is used 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 a 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 operating at 3.20 GHz clock speed, 16 GB RAM with 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) and NVIDIA Quadro P400 (GPU-2) were used for CNN-based YOLOv3 object detection algorithm and for accelerating video projection, respectively. The refresh rate, video synchronization (Vsync), and projection rate are synchronized with the frequency of the rotating color wheel 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 luminescence of 5200 lux.

We used a CNN-based YOLO [70] algorithm to detect and localize objects on the image planes. GPU-1 accelerates the detection algorithm to detect and localize the objects from pre-learned models in real-time. The inference process outputs the ROI of detected objects based on the originally learned weights for YOLOv3 from the 80-class COCO dataset. The ROIs are segregated based on classification to prepare informative color masks.

As shown in Figure 4, to observe the informative color mask as a uniform gray-level brightness by the HVS, we used FPP and IPP at 120 Hz by the DLP projector which is 60 Hz each phase. 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, which can be pointed by the AiCP system. The PC-based software generates FPP and IPP 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 Figure 6, the HFR vision system consists of a monochrome USB3.0 high-speed camera head (Baumer VCXU-02M) with resolution of 640 × 480 pixels. The specification of the PC to implement 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 (approx. 255-bits); 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, then it is denoted as 1 (255-bits); otherwise, it is denoted as 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 which are pointed by the AiCP module simultaneously.

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 interval 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 and comparison

To confirm the functionalities of our algorithms, we quantify the robustness of proposed system. We then compare our method with conventional object trackers to verify their compatibility with AiCP system.

Robustness of the proposed system

First, we quantify the robustness of he system by confirming the functionalities of HFR vision-based recognition 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 characterize 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 Figure 7, three miniature cars of different sizes with a complex background were placed on the linear slider. The AiCP module was placed in front of the experimental scene at 1 m, and an HFR camera with a RICOH C-mount 8-48mm f/1.0 manual zoom lens was placed at distance of 2 m. The linear slider moved the objects of interest horizontally at a velocity 50 mm/s within the projection area. As shown in Figure 8 (a) the aperture is varied from f/4.0 to f/16.0, the total brightness is reduced exponentially, but the decoded area is not significantly affected. In the case of a lens-focus blur, as shown in Figure 8 (b), even with a varying focus depth from 1 m to 7 m, the blur index is reduced, yet the decoded area is substantially unaffected. In both cases, the focal length and focal depth were set to 16 mm and 1.5 m, respectively. However, the decoded area increases with an

increase in focal length (zoom in) from 8 mm to 36 mm because the number of pixels having informative light is increased, as shown in Figure 8 (c) with a fixed aperture of f/4 and focus depth of 1.5 m. It is evident that if the decoded area is visible to the HFR vision system, the recognition and trajectory estimation are always efficient.

Comparison with object trackers

Second, we compare our HFR vision-based trajectory estimation method with conventional object trackers to confirm its compatibility with the AiCP system and computational efficiency using MOSSE [71], KCF [72], Boosting [73], and Median Flow [74] which are distributed in the OpenCV [75] 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. It is apparent that apart from the MOSSE tracker, other trackers lose the pointed objects when occlusion arises. These methods can track any object pointed by the user are highly incompatible with AiCP-based object pointing. Hence, our method has advantages over other methods in terms of computational cost; it is also compatible in estimating the trajectory of objects pointed by the AiCP system.

Multi-object pointing and HFR vision-based recognition and trajectory estimation

Next, we demonstrate the effectiveness of this system in multi-object pointing using the AiCP system and their recognition as well as the trajectory estimation by a distantly placed HFR vision system. As shown in Figure 9, seven miniature objects were placed in a scene with a complex background. Among these objects, the traffic signal and clock were immovable, 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. To demonstrate multiple view-point recognition and trajectory estimation, we used two HFR vision systems placed at 1.5 m away from the scene. Both the HFR cameras were mounted with 4.5 mm C-mount lenses and operated at 480 fps with 2 ms exposure. The projection area was set to 510 mm widthwise and 370 mm heightwise with a luminescence of 805 lux. Both linear sliders moved the corresponding objects to and fro horizontally to 200 mm at a velocity of 50 mm/s within the projection area.

The informative color mask of each object for AiCP system and their corresponding recognition ID for the HFR vision system are 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 is based on a combination of FPP and IPP for a particular object. Once the single bit information transmitted by AiCP system sequentially, the HFR vision system has acquired them in the same sequence frame by frame. Each HFR vision system has a database of predefined recognition ID of desired objects.

The experiment was conducted for 6 s; during this period, both linear sliders moved the objects one time to and fro horizontally. As shown in Figure 10, the

CNN-detection frames were captured at 1 s interval, contain the detected objects marked with bounding boxes. The AiCP system-generated informative color masks for each bounding box (that is, rectangular marker) were projected onto the respective objects by AiCP system. Figures 11 and 12 show a color map of the decoded informative masks (left frames) along with recognition results (right 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 Figures 12 (a) and 12 (b), respectively, whereas, those of camera-2 of HFR vision systems are shown in Figures 14 (a) and 14 (b). It is evident that the decoded areas on moving objects vary in each frame.By constrast, in the case of static objects such as a clock and traffic light, there is no significant change. The decoded areas are often affected by the light absorbent properties of the material and curvatures of the objects. Sometimes, due to the low reflective parts of the objects, the FPP and IPP are indifferentiable. This affects the decoded area and the object is not detected by the HFR vision system.

We quantified the amount of communication that occurred between 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; thus, it could detect those 7 objects approximately 180 times; the DLP projector required 2880 planes to project the information in 6 s. In the case of the HFR vision system, all the 2880 planes were acquired at 480 fps; our system could detect the pointed objects 360 times (that is, a packet of 8 frames for per recognition). The recognition rate is double the detection rate; however, it is the same as the projection rate of AiCP system.

Conclusion

In this study, we developed an AiCP system for object pointing and HFR vision systems that can recognize and estimate the pointed object trajectory simultaneously using VLC technology. The AiCP module is a smart object system that can transmit an informative visible light by active and an informative projection mapping. It can generate unique spatiotemporal-modulated imperceptible patterns locally for objects to be pointed globally. We quantified the robustness of HFR vision system in terms of varying lens-aperture, lens-zoom, and lens-focus blur and confirmed the compatibility of recognition and trajectory estimation with AiCP by comparing with conventional trackers. In addition, we demonstrated multi-object pointing and simultaneous recognition and trajectory estimation of moving as well as static objects, where we quantified the amount of data transmission during the experiment time. Hence, our proposed system applies to real-world scenarios such as security and surveillance in vast areas, SLAM for mobile robots and automatic driving systems. Currently, the system is limited for static and slowly moving objects owing to the projection-latency in commercially available projectors and the reflectance properties of the objects. We intend to improve the proposed system using bright and low latency high-speed projectors to recognize high-speed moving multi-objects in 3-D scenes from a long distance.

Acknowledgements
Not applicable.

Author's 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.

Funding
Not applicable.

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. 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
2. Rekimoto J, Katashi N (1995) The World Through the Computer: Computer Augmented Interaction with Real World Environments. Proceedings of UIST. pp 29-36 . https://doi.org/10.1145/215585.215639
3. Caudell, Thomas P (1994) Introduction to augmented and virtual reality. SPIE Proceedings: Telemanipulator and Telepresence Technologies. 2351:272–281. https://doi.org/10.1117/12.197320
4. Bimber O, Raskar R (2005) Spatial Augmented Reality, Merging Real and Virtual Worlds.
5. Mine M, Rose D, Yang B, Van Baar J, Grundhofer A (2012) Projection-Based Augmented Reality in Disney Theme Parks. IEEE Computer. 45(7):32–40. https://doi.org/10.1109/MC.2012.154
6. 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
7. 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. Lecture Notes in Computer Science 5414:410–422. https://doi.org//10.1007/978-3-540-92957-4 _36
8. Jason G, Structured-light 3D surface imaging: a tutorial (2011). Adv. Opt. Photon. 3:128-160. https://doi.org//10.1364/AOP.3.000128
9. Amit B, Philipp B, Anselm G, Daisuke I, Bernd B, Markus G (2013) Augmenting physical avatars using projector-based illumination. ACM Trans Graph, Article 189. https://doi.org/10.1145/2508363.2508416
10. 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. pp: 321-322. https://doi.org/10.1109/ISMAR.2003.1240740
11. Yamamoto G, Sato K (2007) PALMbit: A Body Interface Utilizing Light Projection onto Palms. Inst. Image Information and Television Engineers. 61(6): 797–804. https://doi.org/10.3169/itej.61.797
12. Grundhöfer A, Iwai D (2018) Recent Advances in Projection Mapping Algorithms, Hardware and Applications. Computer Graphics Forum. 37: 653-675. https://doi.org//10.1111/cgf.13387
13. 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. pp:85–94. https://doi.org/10.1109/IWAR.1999.803809
14. Zhang H, Fronz S, Navab N (2002) Visual Marker Detection and Decoding in AR Systems: A Comparative Study. Proc. Int. Symp. Mixed and Augmented Reality. pp 97–106. https://doi.org/10.1109/ISMAR.2002.1115078
15. 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/10.1109/TVCG.2009.99
16. Chandaria J, Thomas G A, 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
17. Hanhoon P, Jong-Il P (2005) Invisible marker based augmented reality system. Proc. SPIE Visual Communications and Image Processing 5960:501–508. https://doi.org/10.1117/12.631416
18. Lima J, Roberto R, Francisco S, Mozart A, Lucas A, João M T, Veronica T (2017) Markerless tracking system for augmented reality in the automotive industry. Expert Systems with Applications 82(C):100 - 114. https://doi.org/10.1016/j.eswa.2017.03.060
19. 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

20. Punpongsanon P, Iwai D, Sato K (2015) Projection-based visualization of tangential deformation of nonrigid surface by deformation estimation using infrared texture. Virtual Reality 19:45–56. https://doi.org/10.1007/s10055-014-0256-y
21. Punpongsanon P, Iwai D, Sato K (2015) SoftAR: Visually Manipulating Haptic Softness Perception in Spatial Augmented Reality. IEEE Transactions on Visualization and Computer Graphics 21(11):1279-1288. https://doi.org/10.1109/tvcg.2015.2459792
22. Narita G, Watanabe Y, Ishikawa M (2017) Dynamic Projection Mapping onto Deforming Non-Rigid Surface Using Deformable Dot Cluster Marker.IEEE Transactions on Visualization and Computer Graphics. 23(3):1235-1248 https://doi.org//10.1109/TVCG.2016.2592910.
23. 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. International Journal of Distributed Sensor Networks 14: 155014771879085. https://doi.org/10.1177/1550147718790853
24. 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
25. Hornbeck L J (1995) Digital Light Processing and MEMS: Timely Convergence for a Bright Future. Plenary Session, SPIE Micromachining and Microfabrication.
26. Younse J M (1995) Projection Display Systems Based on the Digital Micromirror Device (DMD). SPIE Conference on Microelectronic Structures and Microelectromechanical Devices for Optical Processing and Multimedia Applications. 2641:64 – 75. https://doi.org/10.1117/12.220943
27. Heinz M, Brunnett G, Kowerko D (2018) Camera-based color measurement of DLP projectors using a semi-synchronized projector camera system. Photonics Europe 10679: 178–188. https://doi.org/10.1117/12.2307119
28. Dudley D, Walter M, John S (2003) Emerging digital micromirror device (DMD) applications. Proc. SPIE , MOEMS Display and Imaging Systems. 4985: 14–25. https://doi.org/10.1117/12.480761
29. Oliver B, Iwai D, Wetzstein G, Grundhöfer A (2008) The Visual Computing of Projector-Camera Systems. In ACM SIGGRAPH 2008 classes. 84:1–25. https://doi.org/10.1145/1401132.1401239
30. 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. pp 3211-3216. https://doi.org/10.1109/IROS.2007.4399626
31. 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
32. 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.
33. 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
34. 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
35. Watson A B (1986) Temporal Sensitivity. Handbook of perception and human performance. 1:6.1-6.43
36. Hewlett G, Pettitt G (2001) DLP Cinema™ projection: A hybrid frame-rate technique for flicker-free performance. Journal of the Society for Information Display. 9:221-226. https://doi.org/10.1889/1.1828795
37. Fofi D, Sliwa T, Voisin Y (2004) A Comparative Survey on Invisible Structured Light. Machine Vision Applications in Industrial Inspection XII. 5303:90 – 98. https://doi.org/10.1117/12.525369
38. Siriborvornratanakul T, Sugimoto M (2010) ipProjector: Designs and Techniques for Geometry-Based Interactive Applications Using a Portable Projector International Journal of Digital Multimedia Broadcasting. 2010(352060):1–12. https://doi.org/10.1155/2010/352060
39. Siriborvornratanakul T (2018) Enhancing User Experiences of Mobile-Based Augmented Reality via Spatial Augmented Reality: Designs and Architectures of Projector-Camera Devices. Advances in Multimedia. 2018:1-17. https://doi.org/10.1155/2018/8194726.
40. 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.
41. 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
42. 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
43. Kodama M, Haruyama S (2017) A Fine-Grained Visible Light Communication Position Detection System Embedded in One-Colored Light Using DMD Projector. Mobile Information Systems, 2017, [9708154]. https://doi.org/10.1155/2017/970815
44. 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
45. 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
46. 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.
47. 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
48. 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
49. 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
50. 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
51. 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
52. Ishii I, Tatebe T, Gu Q, Takaki T (2012) Color-histogram-based Tracking at 2000 fps. J. Electronic Imaging. 21(1):1 – 14. https://doi.org/10.1117/1.JEI.21.1.013010
53. Ishii I, Tatebe I, Gu Q, Takaki T (2011) 2000 fps real-time target tracking vision system based on color histogram. Proceedings of the SPIE. 7871:21-28. https://doi.org/10.1117/12.871936
54. 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
55. 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. Journal of Robotic and Mechatronics. 27(1):12-23. https://doi.org/10.20965/jrm.2015.p0012
56. Raut S, 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
57. Ishii I, Kurozumi S, Orito K, Matsuda H (2008) Automatic scratching pattern detection for laboratory mice using high-speed video images. IEEE Trans. Autom. Sci. Eng.. 5(1):176–182. https://doi.org/10.1109/TASE.2007.902868
58. Nie Y, Ishii I, Yamamoto K, Orito K, Matsuda H (2009) Real-time scratching behavior quantification system for laboratory mice using high-speed vision. J. Real-Time Image Process.. 4:181–190. https://doi.org/10.1007/s11554-009-0111-7
59. Gu Q, Kawahara T, Aoyama T, Takaki T, Ishii I, Takemoto A, Sakamoto N (2015) LOC-based high-throughput cell morphology analysis system IEEE Trans Autom Sci Eng. 12(4):1346–1356. https://doi.org/10.1109/TASE.2015.2462118
60. Sakuma S, Kuroda K, Tsai C, Fukui W, Arai F, Kaneko M (2014) Red blood cell fatigue evaluation based on the close-encountering point between extensibility and recoverability.Lab on a Chip. 14: 1135-1-141. https://doi.org/10.1039/c3lc51003d
61. Gu Q, Aoyama T, Takaki T, Ishii I (2015) Simultaneous vision-based shape and motion analysis of cells fast-flowing in a microchannel IEEE Trans Autom Sci Eng. 12(1):204–215. https://doi.org/10.1109/TASE.2013.2292583.
62. Yang H, Gu Q, Aoyama T, Takaki T, Ishii I (2013) Dynamics-based stereo visual inspection using multidimensional modal analysis. IEEE Sensors Journal. 13(12)4831–4843. https://doi.org/10.1109/JSEN.2013.2276620
63. Aoyama T, Li L, Jiang M, Inoue K, Takaki T, Ishii I, Yang H, Umemoto C, Matsuda H, Chikaraishi M, Fujiwara A (2017) Vibration sensing of a bridge model using a multithread active vision system. IEEE/ASME Trans. Mechatronics. 23(1): 179-189. https://doi.org/10.1109/TMECH.2017.2764504
64. Hayakawa T, Watanabe T, Ishikawa M (2015) Real-time high-speed motion blur compensation system based on back-and-forth motion control of galvanometer mirror. Opt. Express. 23(25):31648–31661. https://doi.org/10.1364/OE.23.031648
65. Ueno T, Gu Q, Aoyama T, Takaki T, Ishii I, Kawahara T (2015) Motion-blur-free microscopic video shooting based on frame-by-frame intermittent tracking. Proceedings of the IEEE Conference on Automation Science and Engineering. pp 837–842. https://doi.org/10.1109/CoASE.2015.7294185.
66. Namiki A, Hashimoto K, Ishikawa M (2003) A hierarchical control architecture for high-speed visual servoing. Int. J. Robot. Res. 2003. 22(10–11):873-–888. https://doi.org/10.1177/027836490302210006
67. Senoo T, Namiki A, Ishikawa M (2006) Ball control in high-speed batting motion using hybrid trajectory generator. In Proceedings of the IEEE International Conference on Robotics and Automation. pp 1762–1767. https://doi.org/10.1109/ROBOT.2006.1641961
68. Jiang M, Aoyama T, Takaki T, Ishii I (2016)Pixel-level and robust vibration source sensing in high-frame-rate video analysis . Sensors. 16(11):842-849. https://doi.org/10.3390/s16111842
69. Jiang M, Gu Q, Aoyama T, Takaki T, Ishii I (2017) Real-time vibration source tracking using high-speed vision. IEEE Sensors Journal.17(5):1513-1527. https://doi.org/10.1109/JSEN.2016.2647690
70. Redmon J, Farhadi A (2018) YOLOv3: An Incremental Improvement. CoRR. abs/1804.02767. http://arxiv.org/abs/1804.02767.
71. 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.
72. Henriques J F, 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
73. Grabner H, Grabner M, Bischof H (2006) Real-Time Tracking via on-Line Boosting. In Proceedings of the British Machine Vision Conference. 1:47-56. https://doi.org/10.5244/C.20.6
74. 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.
75. https://opencv.org/

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 (image plane + inverse plane)	16.15
Total (1)–(2)	33.32

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 Tracking	0.002
(8) Image Display	16.66
Total [(1-5) x 8]+(6)+(7)	12.707

Table 3: Execution times on HFR recognition and tracking system (unit: ms).

	time	compatiblity with AiCP based VLC	remark
(1) MOSSE	0.5	no	any object
(2) KCF	19.45	no	any object
(3)Boosting	8.01	no	any object
(4) MEDIAN FLOW	40.26	no	any object
(5) Our Method	12.707	yes	recognize and track only pointed objects by AiCP System

Table 4: Relation between objects under AiCP Projection and IVLInfo Planes

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

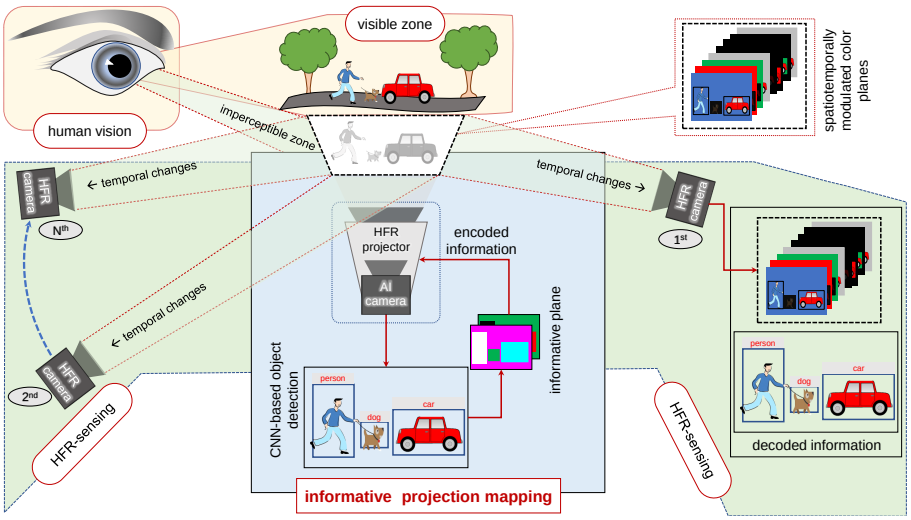

Figure 1: Concept of AiCP-based object pointing and HFR vision-based recognition

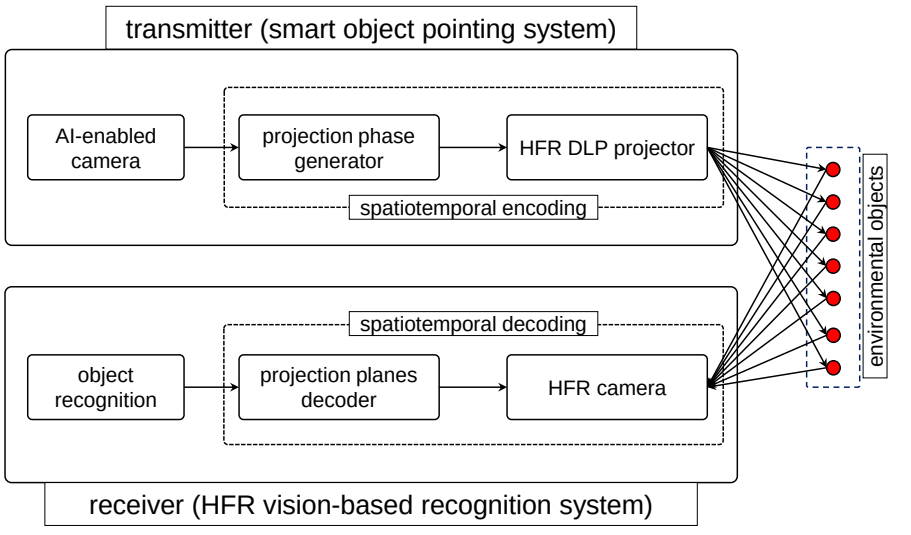

Figure 2: VLC-based transmitter and receiver

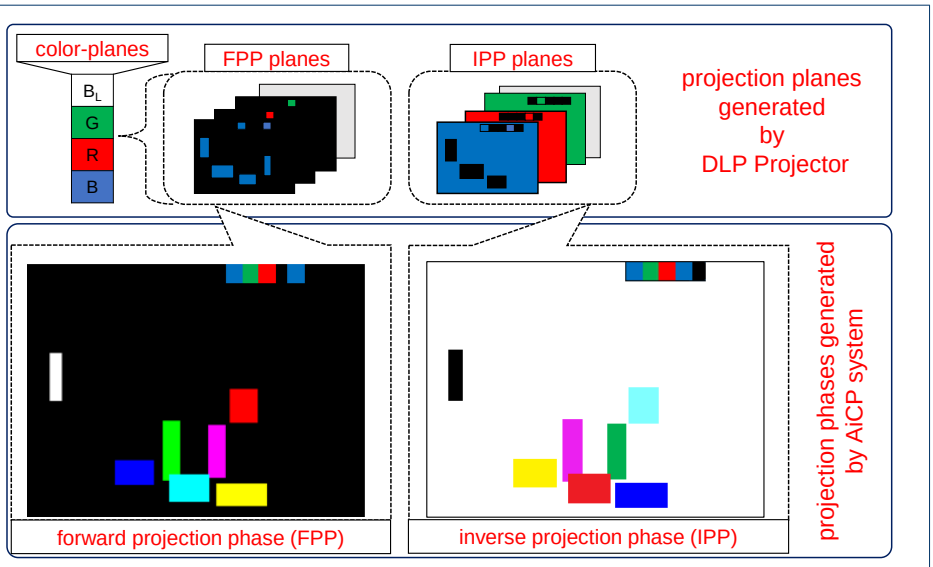

Figure 3: Projection phases and corresponding projection planes of AiCP system

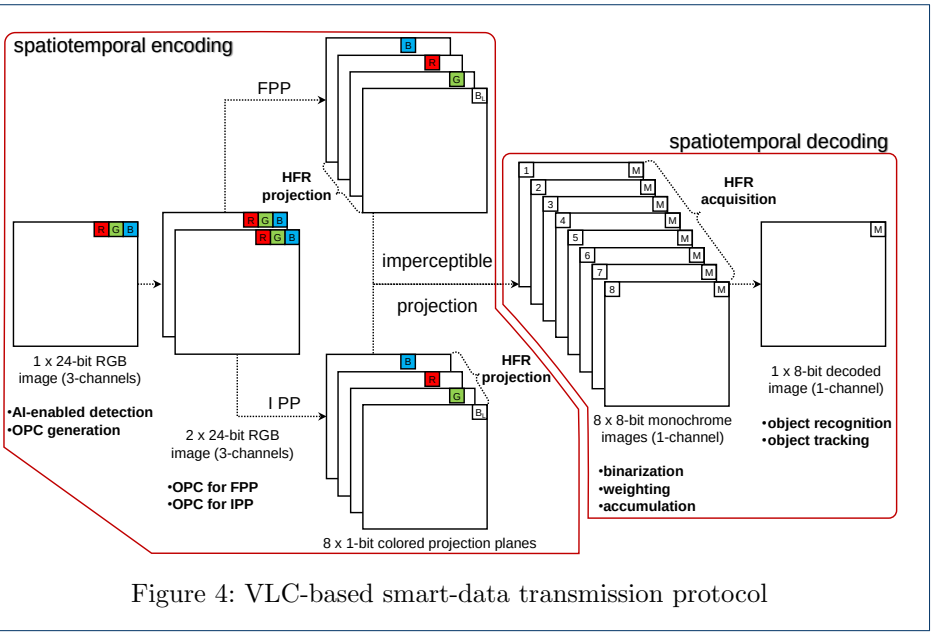

Figure 4: VLC-based smart-data transmission protocol

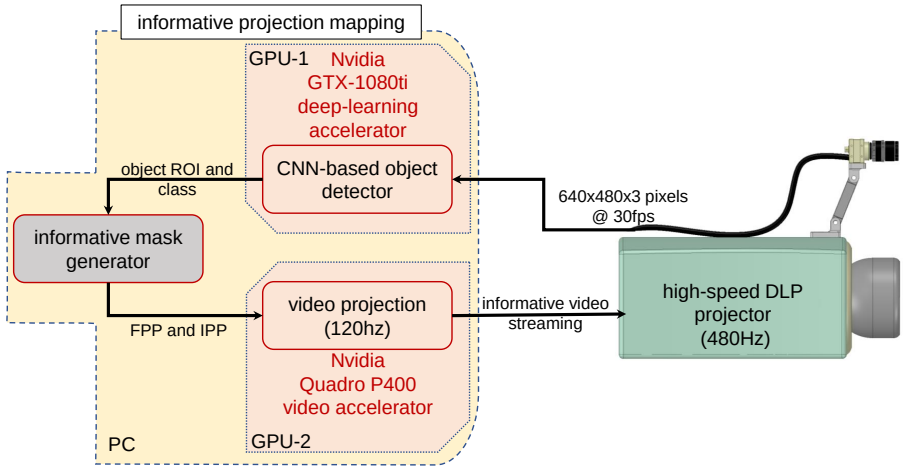

Figure 5: AiCP system for active projection mapping

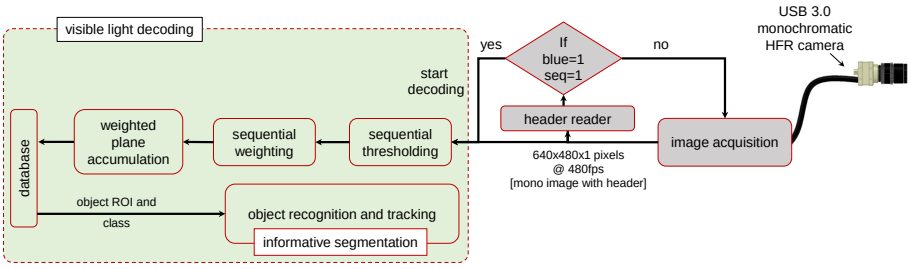

Figure 6: HFR vision system for recognition and tracking

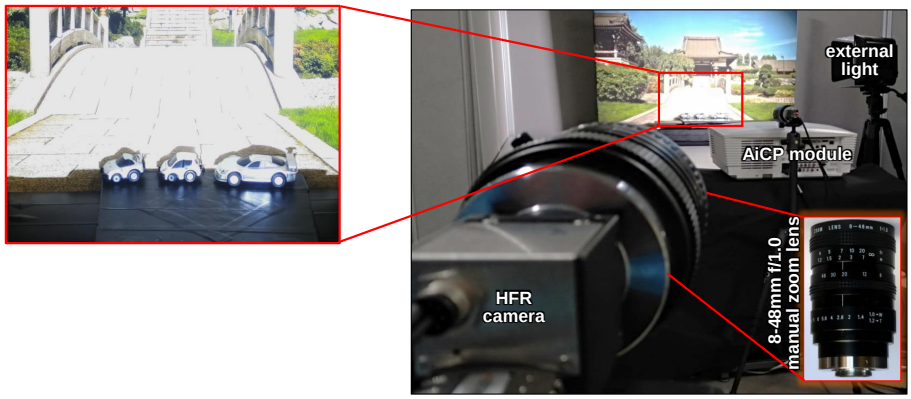

Figure 7: Overview of the experimental setup for quantification of robustness

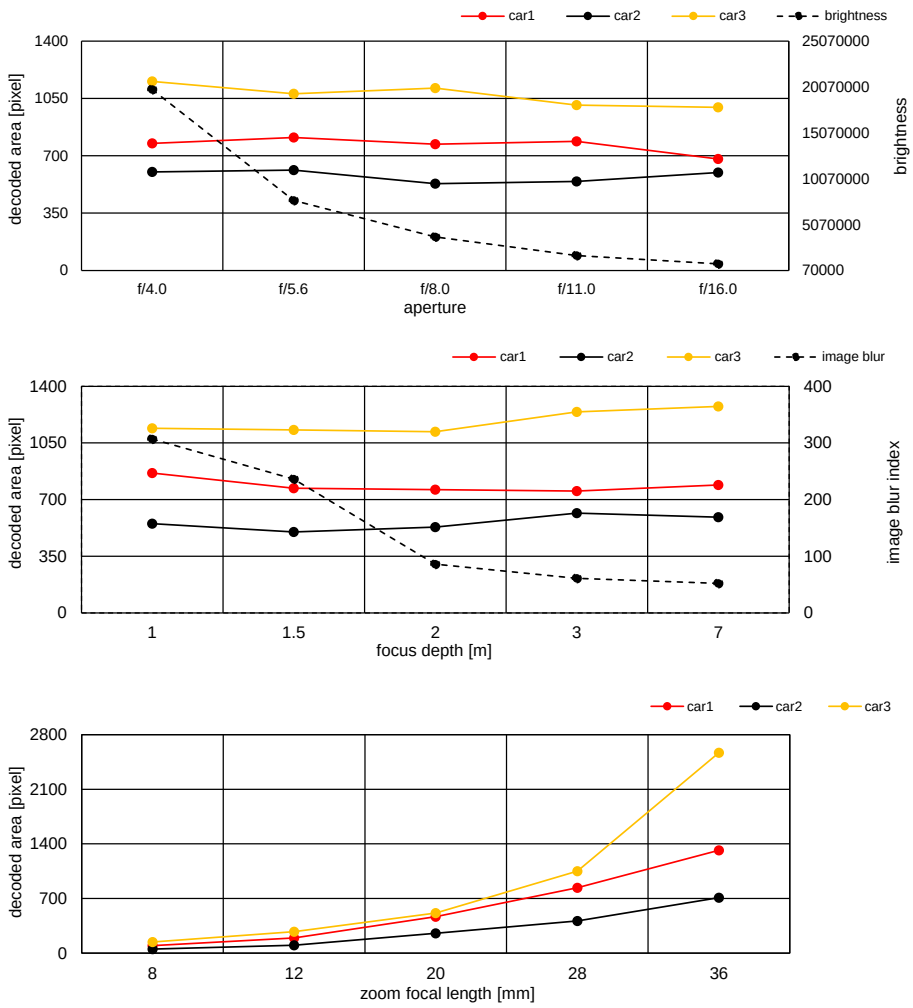

Figure 8: Robustness in the decoded area when (a) varying aperture (b) varying focus depth, and (c) varying focal length (zoom in)

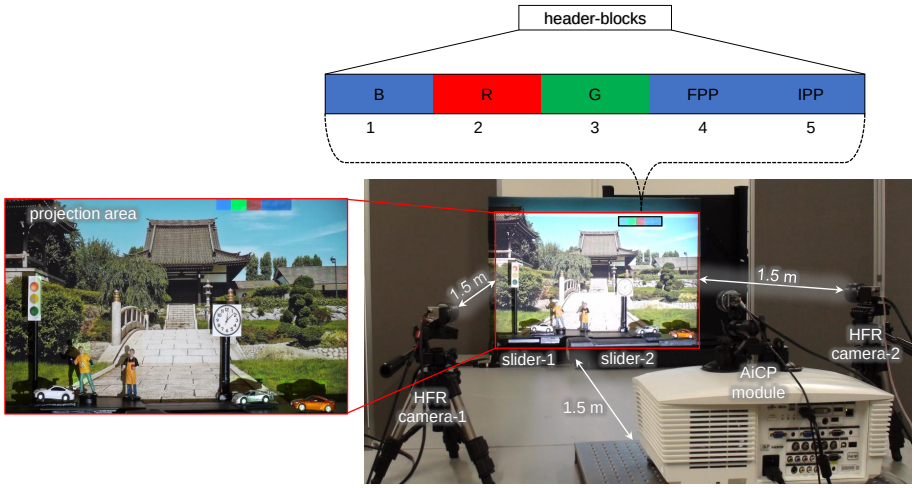

Figure 9: Overview of the experimental setup for multi-object pointing and HFR vision-based tracking

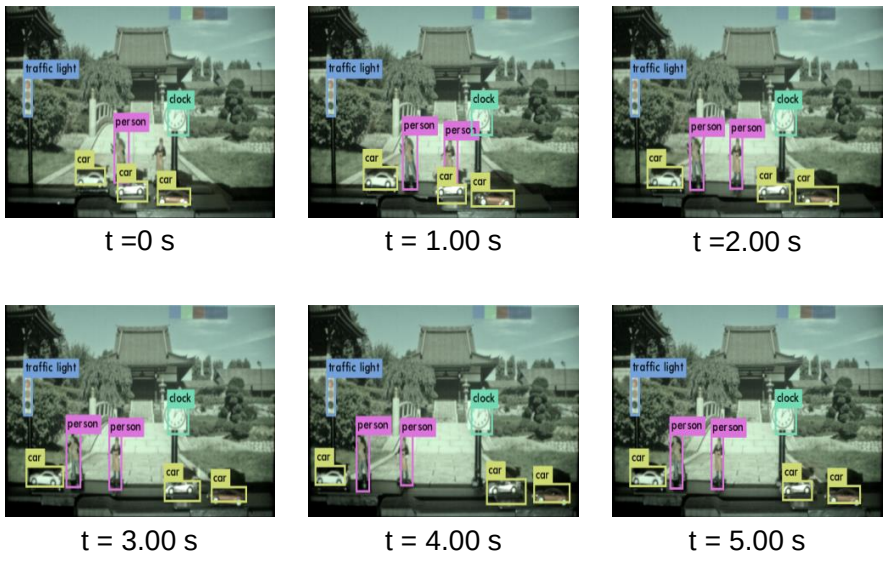

Figure 10: Multiple objects pointed by AiCP system

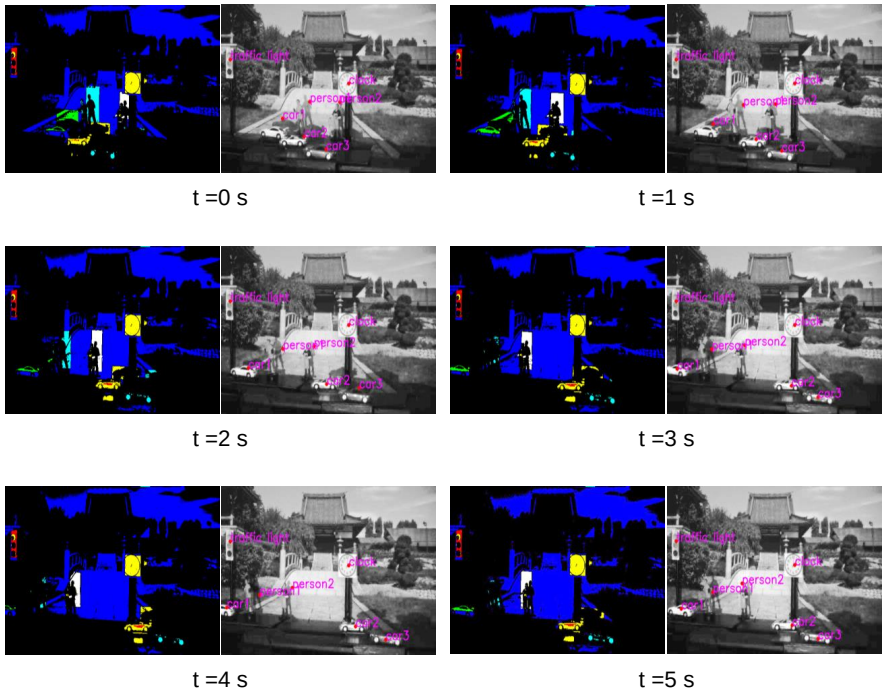

Figure 11: Color map of decoded informative color mask (left) and corresponding recognition (right) acquired in Camera-1 of HFR vision system

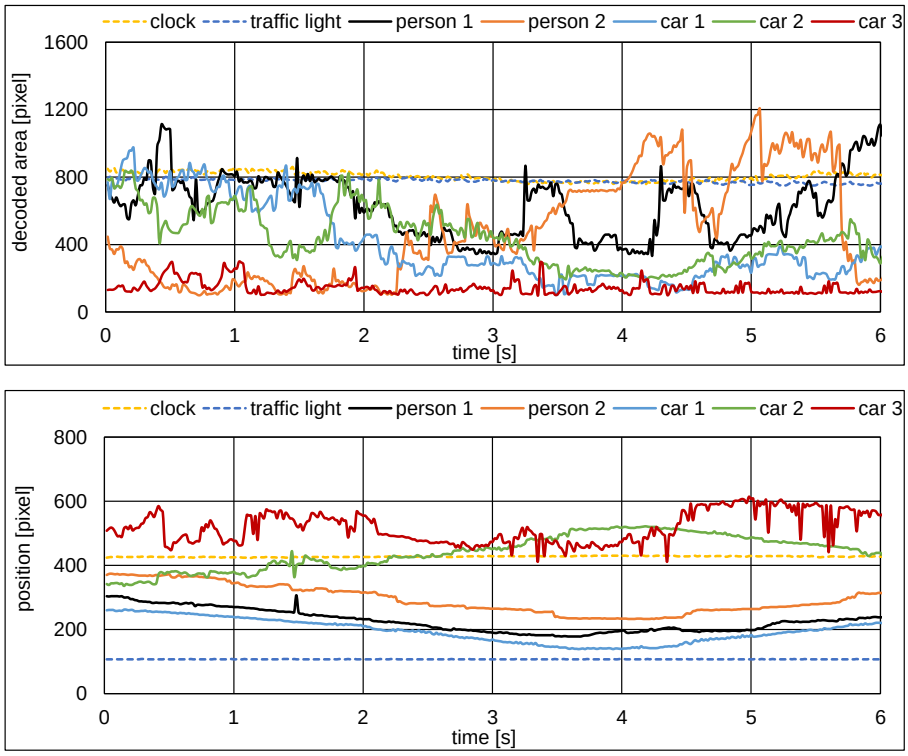

Figure 12: Graphs of (a) decoded area acquired, and (b) corresponding displacement of each recognized objects in Camera-1 of HFR vision system

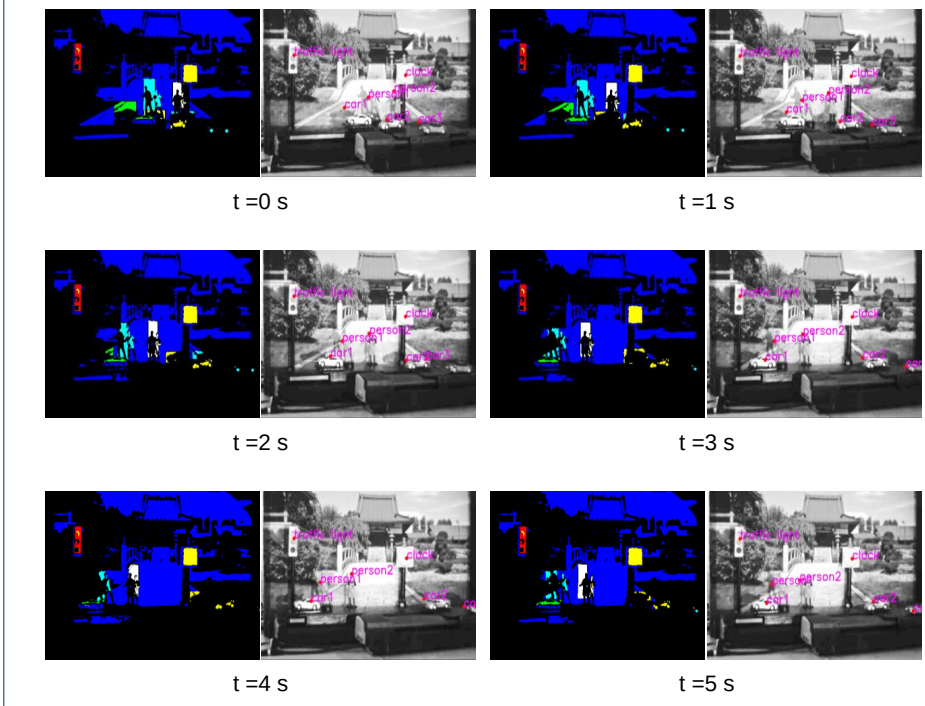

Figure 13: Color map of decoded informative color mask (left) and corresponding recognition (right) acquired in Camera-2 of HFR vision system

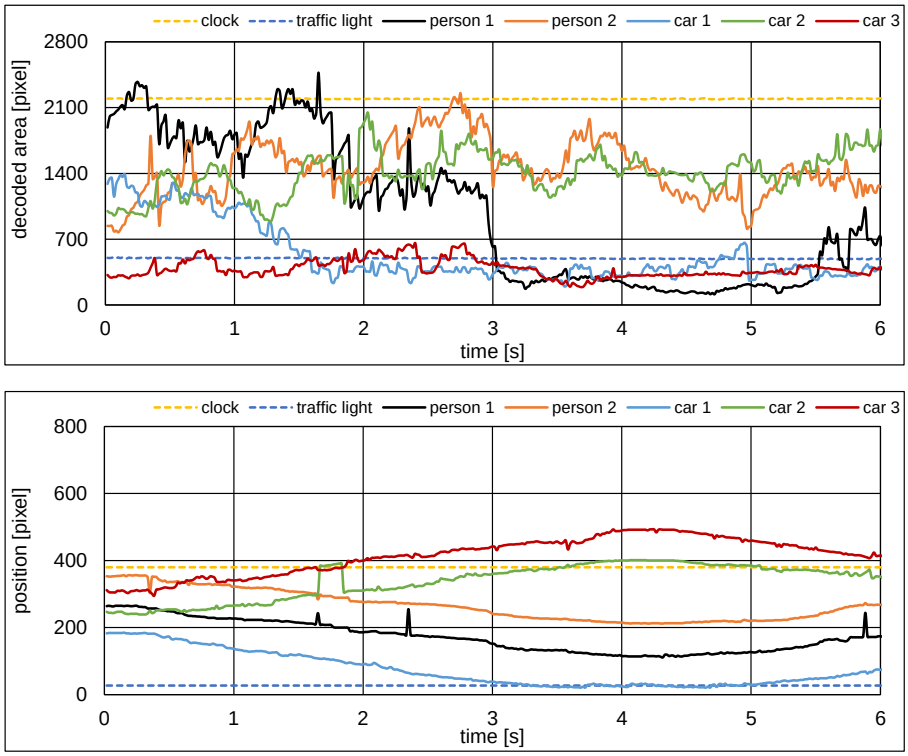

Figure 14: Graphs of (a) decoded area acquired, and (b) corresponding displacement of each recognized objects in Camera-2 of HFR vision system