

Machine vision for light measurement and human-centric light management systems

Theodore Tsesmelis, Irtiza Hasan, Marco Cristani, Alessio Del Bue[†], Fabio Galasso[†]

Abstract—Lighting needs to be measured and customized to the people using it. This is the case in offices, retail shops and industrial warehouses, whereby a proper lighting ensures higher productivity and safety. However measuring light is still today time-demanding (*e.g.* needing sensor measures with luxmeters everywhere in the space) or resource-demanding (*e.g.* acquiring detailed CAD models for processing-intensive simulations). Both approaches are offline and need to be repeated entirely if the space structure changes. In addition neither of the procedures considers the live presence of people in the environment. Here we propose the automation of real-time lighting measurement, the online detection of the people presence in the environment and the integration of both aspects into a smart light management system. Our system uses an RGBD sensor and a radiosity-based light propagation model. Thanks to the integration of light distribution and perception curves into the radiosity model, we outperform a commercial software (Relux) on a newly introduced dataset, proposed to quantify the task. The computational complexity of our model allows for real-time performance and it considers that people in the environment perceive light according to their visual frustum. When integrated into a light management system, this allows to change lighting adaptively based on the presence, attention and activity of people. While the lighting quality is preserved, the energy consumption can be reduced by up to 66%, as we experimentally quantify.

Index Terms—machine vision, light modeling, radiosity, illumination map, light management system, human-centric lighting.

I. INTRODUCTION

LIGHT is indispensable in our perception of the world and it affects our emotional and physiological responses [1], [2], [3]. Good-quality lighting can improve well-being, visual performance and interpersonal communications. Well-lit workplaces provide visual comfort and improvement of intellectual productivity and effectiveness [4], [5], [6]. However, lighting may reach 15% of the overall building electricity consumption, with peaks of 25% or more [7], [8], [9]. In line with the above, the stated mission of lighting installations according to the International Association of Lighting Designers [10] should be the *optimal balance among human needs, architectural considerations and energy efficiency*.

T.Tsesmelis (t.tsemelis@osram.com) has a triple affiliation with Osram GmbH, the Italiano Istituto di Tecnologia (IIT), Genova, Italy, and the Department of Computer Science, University of Verona, Verona, Italy.

I. Hasan, is affiliated with the Inception Institute of Artificial Intelligence (IIAI), Abu Dhabi, UAE. This work was done while working at Osram GmbH and University of Verona.

M. Cristani is affiliated with the Department of Computer Science, University of Verona, Verona, Italy.

A. Del Bue is affiliated with the Italiano Istituto di Tecnologia (IIT), Genova, Italy.

F. Galasso is affiliated with Osram GmbH, Munich, Germany.

[†]These two authors contributed equally to the work.

Current lighting systems lack the capability of real-time dense spatial light measurement. Light measurement is only accomplished via manually operated instruments such as luxmeters (point-to-point measurements) or via offline CAD-based simulations. It requires therefore either an operator to visit the scene and perform the measurements, or a detailed CAD model of the 3D scene any time this changes. But offices are nowadays flexible open-plan spaces, which should adapt to the daily work plan.

Any modern light management system (LMS) lacks the real-time estimation of human light perception and the capability to control light according to it, which limits its performance and energy efficiency, especially in large open-plan spaces. Current LMSs limit energy waste by time switching, daylight harvesting and via an occupancy sensing system (OSS). The first stands for scheduling off-times outside working hours. The second adjusts the luminaires to maximize the use of daylight, when available [11]. Only an OSS switches lights on/off according to the people presence, but it only leverages a coarse thermal- or radar-based motion detector [12], [13]. The OSS reacts therefore to presence anywhere in the office-space, switching-on all luminaires, no matter the size of the open-plan office [14]. In fact, no lighting system estimates how much light each person in the scene receives, as a real-time function of their position and visual attention (*aka.* visual frustum of attention – VFOA – or gaze). But this is needed to realize a real-time end-to-end LMS, which only illuminates the perceived office space.

Here we introduce a new LMS which estimates both the scene lighting and the human perception of it. The first leverages an RGBD camera and a radiosity model for light propagation, and distinguishes the scene 3D structure, the object reflectance and light positions. The second locates people in the environment and estimates their visual attention, by a variant of the Mask R-CNN model [15]. We integrate both aspects to estimate how much light each person perceives, depending on their position and gaze.

We introduce a new labelled dataset, featuring a number of lit rooms with and without human activity based on RGBD images and 3D meshes which are labelled with the material reflectance properties, the luminaire positions, characteristics and dimming level, the person locations and VFOA. The actual light intensity at multiple scene location and from the people viewpoint is additionally measured with luxmeters, to benchmark light estimation and perception.

Finally, we introduce a new end-to-end system architecture to bring together the light estimation and human perception with the light management, by creating a novel autonomous system that we call the “*invisible light switch*” (ILS). The

system encompasses an RGBD camera, a processor, a light controller, a communication bus and the luminaires. Since *ILS* estimates how much light each person receives, it may switch off or dim those luminaires which are not visible, *e.g.* on the other side of large open spaces or behind cubicle panels. This removes the need for manual switches and provides a boost in energy efficiency, saving up to 66% without compromising the light quality.

Our main contributions are: *i*. we propose real-time light estimation from an RGBD sensor as well as its perception by the scene occupants; *ii*. we collect new data and metrics for a quantitative evaluation; *iii*. we propose a new end-to-end LMS architecture, to leverage the light and human perception estimation for autonomous smart lighting and energy saving. This manuscript brings together two previous conference publications on the topics of light estimation and the invisible light switch [16], [17], and extends the work with *a*) more novel datasets and annotations, *b*) more experiments, and *c*) the definition of the overall end-to-end system architecture.

Next we review the related work; we present the proposed end-to-end system and modelling in Section III; we introduce the new dataset, comparative experiments and ablation studies in Section IV; and we conclude the manuscript in Section V.

II. RELATED WORK

Light measurement and management encompass different fields in science and engineering. First we review prior work on light measurement from image processing, computer graphic and lighting design and simulation. Then, we focus on LMSs and their relationship to power efficiency.

A. Light Measurement

1) Image processing: It is stated that lighting technology is shifting from the measurement and setup of lighting in the scene to the assessment and design of the light arriving at the eyes of the observers [18]. Towards this goal, camera-aided light measurement solutions have recently been proposed by [19], [20], [21]. Specifically, the white paper by Hiscocks *et al.* [19] illustrates luminance measurement with a digital camera. [20] makes applies it to small simulated environments, made of a carton box, a cheap web camera and a processing unit. On the other hand, [21] proposed to use high-dynamic-range (HDR) computational photography, requiring more costly equipment. However, the light intensity illuminating a surface is not reliably recoverable from a pixel-like array of radiance values as dependent to other physical variables such as irradiance, surface reflectance, and complex inter-reflections between all the surfaces in the 3D scene [22]. Untangling them is challenging in the general case, as well explained in pioneering works on “*inverse rendering*” [23].

2) Computer graphics: Light modeling and understanding have also been studied in computer graphics for the creation of photo-realistic renderings [24], [25]. The forward and inverse light transport theory simulates the path of transmitted light in a 3D environment and models the image as an integration process. This requires knowledge of the 3D mesh, the material surface reflectance, and many other physical properties [26],

[27]. Several light models have been proposed to render scenes under seemingly life-like illuminations. The well-known radiosity model [28] is popular for its simplicity and efficiency. Other recent approaches are the instant radiosity [29], based on bouncing energy and virtual point lights, photon mapping [30] and progressive photon-tracing [31], given by tracing photons from a light source and recording their hits on diffuse surfaces.

3) Light design and simulation software: Relax [32], DIALux [33] and AGi32 [34] are commercial CAD-design modelling software products that are commonly used in the lighting design field, for measurement and evaluation of lighting solutions. These require the luminaire manufacturer specifications, the material and photometric properties of objects in the scene (generally from online libraries) and the CAD model of the indoor environment. All of the above perform light simulation by the radiosity model [35] or variants of it. Most recently, research efforts [36], [37] addressed the combination of a physically-based real-time rendering with an interactive lighting simulation for complex architectural environments. In particular, HILITE [36] provides a many-light global-illumination solution provided as light maps. This considers glossy materials [38] and proposes virtual point lights, to represent the scene compactly by clusters of virtual polygon lights. On the other hand, LiteMaker [37] combines multi-resolution image filtering with an interactive, progressive photon-tracing [31].

B. Light Management Systems (LMSs)

LMSs play a crucial role in reducing energy consumption in offices [39], [40] and prior work mainly addressed daylight harvesting [11], [41], [42] and occupancy sensing systems (OSS) [12], [13]. In particular, [39] emphasized the potential of OSSs for energy saving, surveying available technology and the performance compromise; [12] additionally focused on the implementation of occupancy sensors; [13] specifically considered an OSS based on ultrasonic and proposed improvements in user localization based on time-difference-of-arrival and transmission over multiple time slots. However, none of the above addressed energy saving in open-plan offices, nor distinguished among private, single-user and open-plan offices as one would intuitively argue for.

Among related work considering open-plan offices, [43] determined three layout categories: (a) cubicle layout with high partitions ($\sim 1.5\text{m}$ or higher), (b) cubicle layout with low partitions (up to $\sim 1.5\text{m}$) and (c) open layout with no or limited partitions. [44] and [45] proposed and demonstrated by simulation an optimization algorithm to compromise energy efficiency and user-comfort, by considering user presence and daylight conditions. They did not address any occupancy sensing algorithm, nor a real-time light propagation model, which we propose here for a real end-to-end light management system. Finally, the lack of a proper review study on LMSs in open-plan offices has been recently stressed by [14]. They envisioned modern LMSs, capable of dimming light according to how much of it is perceived by each person in the scene, by means of their visual frustum of attention. We realize this vision here with our “*invisible light switch*”.

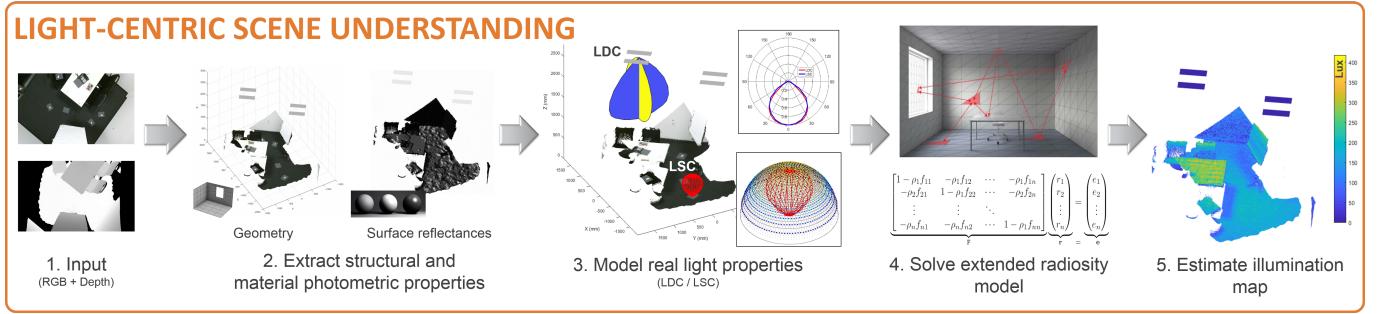


Fig. 1: Pipeline of the proposed light propagation and illumination map estimation. First *ILS* acquires the RGBD images (*step 1*). Then it extracts the albedo values of the surfaces by photometric stereo, by leveraging the RGB and depth image pairs (*step 2*). Afterwards it models the actual light sources (luminaires) using distribution curves, and it similarly does so with light-perceiving devices (luxmeters or the viewing people) (*step 3*). Both aspects are introduced into an extended radiosity model (*Step 4*). Finally, *ILS* estimates the illumination over the visible 3D scene (*Step 5*).

III. PROPOSED END-TO-END SYSTEM

The proposed end-to-end system encompasses three functional tasks: the *light-centric scene understanding*, the *human-centric scene understanding*, and the *light management*. These are depicted in Figs. 1, 2 and 4 respectively.

Light-centric scene understanding is responsible to estimate the scene 3D geometry and the light propagation within the scene. For the light propagation, we adopt and extend the radiosity model to additionally model the actual light source (office luminaires) as well as the receiving sensor (luxmeters or viewing people).

Human-centric scene understanding stands for detecting the people in the scene and estimating their view frustum of attention (VFOA). We use the VFOA to quantify how much light a single person perceives from each light source.

Light-centric scene understanding and human-centric scene understanding are necessary modules when it comes to open-plan offices, where multiple light sources create complex light patterns which act on moving employees.

Finally, light management stands for automatically adjusting lighting in response to the people position and attention. This includes dimming or switching the luminaires off when people do not see them, as in our *ILS* system.

The first two are sensing tasks, performed in first place by the sole use of an RGBD camera. The third task is about light control and leverages an established light-communication BUS (DALI), to read statuses and send commands.

A. Light-centric Scene Understanding

ILS targets the estimation of the scene 6-DoF illumination, *i.e.* anywhere in the 3D space and from any 3D direction, by the sole use of RGBD images. Towards this goal, we define a procedure to retrieve the scene 3D surface from the point cloud, as well as the surface reflectance. We then use and extend radiosity to model the light interaction among 3D scene parts by introducing the distribution characteristics of the actual lights (luminaires) and sensing elements (the luxmeters or the people in the scene) *c.f.* illustration in Fig. 1.

From RGBD images to surface and reflectance. Light propagation requires a scene described by patches, *i.e.* 3D facets characterized by a surface and an orientation. But an RGBD camera only provides a sparse and noisy point cloud.

We set to recover the scene surfaces by first denoising the point cloud by means of the color and depth image pairs [46], [47]. Then we reconstruct the surface as a 3D mesh of multiple scattered patches by using the Open3D library [48].

Given the surface normals, *ILS* estimates the albedo ρ of each surface element by the first-order spherical-harmonics method of [49]. This requires multiple images of each surface patch under different illumination conditions (*e.g.* alternating the lights from each of the scene luminaires). We attain the image-set by time-lapse recordings and by selecting / synthesizing single-light-source-lit images with the technique of [50].

The Radiosity Model. Given the scene 3D surface and the corresponding reflectance, *ILS* estimates the illuminance of each surface patch with the radiosity model. Radiosity is adopted by most commercial light simulation software (Relux, DIALux, AGi32) because it describes the physical light propagation phenomenon and provides light estimations which are close to those measured, in absolute terms. By contrast, computer graphics techniques target realistic effects and only the relative plausibility of patch illumination estimation, *e.g.* with respect to neighboring ones.

Furthermore, radiosity is a *simple* linear model:

$$r_i = e_i + \rho_i \sum_{j=1}^{n-1} f_{ij} r_j, \quad (1)$$

where i and j index the n patches of the room 3D mesh, e_i is a scalar for the self-emittance of patch i ($e_i = 1$ for light source patches in the unit measure, 0 otherwise), ρ_i is the isotropic reflectivity of patch i , f_{ij} is the form factor between patch i and j and r_i is the estimated radiance value on surface i related with the actual light intensity accordingly.

The form factors f_{ij} are defined as the fraction of energy reaching one patch j over the emitted by patch i . These describe the scene geometry and encode two main aspects:

- **Visibility:** whether two patches are visible from each other, being 0 if there is no line of sight between them;
- **Distance and orientation:** how well two patches *see each other*. Small values correspond to far away patches with an oblique line of sight, while large values mean close, fronto-parallel patches.



Fig. 2: Pipeline for the detection of people and the estimation of their VFOA and perceived illumination. ILS first acquires RGBD images of the scene and the occupants (*Step 1*). Then it detect people (the depicted bounding boxes and segmentation provided by Mask R-CNN [15]) and it estimates the VFOA of each of them (the yellow arrow, by means of a head pose estimation algorithm [51]) (*Step 2*). Finally, we integrate the amount of illumination which reaches the people sight, by means of the LSC (*Step 3*). See Section III-B for details.

Here we define the form factors by the “*Isocell*” [52] technique (*c.f.* Fig. 3), due its higher precision at a given computation time (compared to analytic alternatives):

$$f_{ij} = \frac{\mu_j}{\mu_i}, \quad (2)$$

where μ_j is the number of rays reaching patch j over the total number of rays μ_i emitted by facet i .

Light distribution and sensitivity curves. The radiosity model in Eq. (1) has two main limitations: it only models point light sources and it disregards the light perception. The latter limitation affects all current commercial light planning software.

We address both aspects by extending the radiosity model and introducing the LDC (for considering any *non-ideal* light source type) and the LSC (to model the light observer / sensor perception) distributions respectively.

a) Light distribution curve (LDC): The radiosity formulation 1 assumes isotropic light sources, *i.e.* active patches radiate with the same intensity in all directions. However, isotropy is hardly the case for real light sources which are normally represented by a radiation map, such as in Figure 3a (*red*). This distribution is in general non-linear and encoded as a curve which provides the radiant intensity with respect to the distance and angle from the emitting source (*i.e.* how much light is emitted in each direction).

We include the LDC curve into the radiosity model by encoding the non-linear radiation in the form factors f_{ij} . In more detail, we associate scalar values (*i.e.* weights) to each cast ray leaving a light source patch, proportionally to the angle of emission, as illustrated in Fig. 3b. These are then used to re-write Eq. (2) as a weighted mean. By encoding the non-linearity of the LDC curve into the form factors, our reformulation of radiosity preserves the linearity thus having a minimal impact on the computation.

b) Light sensitivity curve (LSC): In a similar fashion as with light sources, the light-measuring sensors have different sensitivity to lighting, depending on the lighting angle and distance as well as to manufacturing characteristics. The LSC plot in Figure 3a (*blue*) illustrates the perception characteristic of the luxmeter sensor, by which we measure the spatial and human perceived illumination in the gathered data.

As for LDC, we integrate the angular weighting of the cast rays by altering the cast rays, *c.f.* 3b. Also this extension maintains the model linearity and computation time.

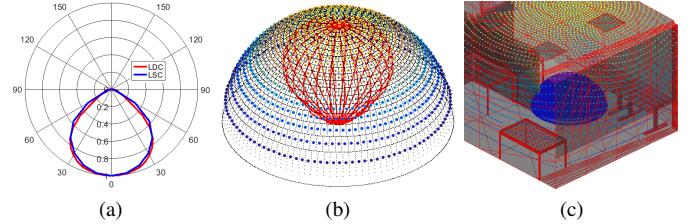


Fig. 3: (a) Radial representation of LDC and LSC curves, illustrating how both quantities are attenuated (radius) with the growing light incident angle (radial angle); (b) weighted ray distribution of the Isocell unit sphere for LDC, heatmap-color-coded; (c) Isocell unit sphere (blue hemisphere) for a patch on the floor of the office CAD. The color-dots in (c) are taken from (b) but placed at the intersection of the cast ray with the other scene patches.

B. Human-centric Scene Understanding

The main human factors of a light management system are the localization of people, the estimation of their visual attention and the consequent estimation of their lighting perception in the scene. We describe here algorithms to address the three aspects, given an RGBD ceiling-mounted camera.

We cast the localization of people as a human detection task from top-view imagery based on the state-of-the-art Mask R-CNN model [15] with the ResNet-101 [53] backbone.

We denote the visual attention of people by their visual frustum of attention (VFOA), which we estimate with the model of [51]. This consists of a Faster R-CNN architecture with VGG16 backbone, extended by a VFOA branch for gaze estimation, taking as input the detected person bounding-box. As noted in [51], the whole-person bounding box provides an important contextual cue, to complement the tiny heads from the top-view imagery.

The VFOA is cast as classifying the person viewing angle into quantized direction bins. We experimented with 4-quantized viewing directions (*North, West, South, East*), as well as with 8 (yielding a granularity of 45°). Moreover, attempting to estimate the VFOA by regression, underperformed the classification approach, based on the fact that it is easier to estimate a class label than the exact angle.

Once we extract the detected 2D people positions and viewing angles we map them onto the 3D space by means of depth to 3D mapping.

Finally, we model the light perceived by the people as the illumination reaching the person’s field-of-view, as described by the 6-DoF head position and VFOA orientation in the scene.

Each field-of-view is assumed to be a conic reception field. The arriving light follows the radiosity model of Sec. III-A, the ray-casting simulation as described for the light sensitivity curve (LSC, *c.f.* 3), and the consequent integration across the human field-of-view. Overall, each person's light perception is approximated as a light sensor alike the luxmeter, positioned between the eyes.

C. Light Management

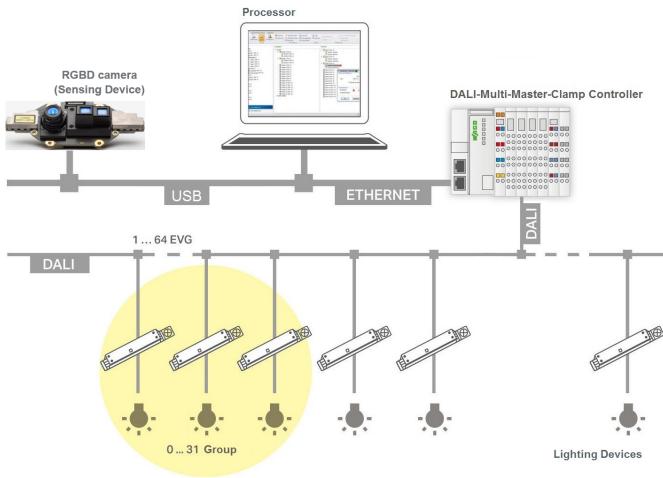


Fig. 4: The end-to-end *ILS* architecture, as described in Sec. III-C. Note the RGBD sensor connected to the processor (exemplified as a laptop) via USB. The processor runs the algorithms to estimate the scene illumination and how much of it each person perceives. It also runs the WAGO DALI Configurator program to encode luminaire commands to a DALI Master Light Controller via TCP/IP. The controller relays the commands to the luminaire and reads out their statuses via the DALI BUS.

ILS is a camera-aided smart light management system (LMS) to control lighting in response to the currently spatial and human perceived estimated light as well as the people position and attention. The *ILS* reads the luminaire status and switches them on and off based on the people presence. Additionally it dims luminaires down when partially visible by the people, therefore saving energy “*in the invisible*”, while maintaining the desirable scene illumination.

As illustrated in Fig. 4, the proposed LMS consists of the sensing RGBD camera, a computing device to estimate the light and human factors in the scene (*c.f.* Secs. III-B,III-A), a Master Light Controller to communicate commands to luminaires, and the luminaires themselves, interconnected via a suitable protocol BUS. *ILS* is implemented as a computer program running on the computing device, based on the people presence and attention, and on the luminaire status readouts. Overall, the proposed LMS system is autonomous, end-to-end and real-time.

We have adopted and report results on images acquired via a Kinect v2 RGBD camera. The computing device is a laptop, running the sensing algorithms of Secs. III-B,III-A and the controlling invisible light switch program. The Kinect v2 is connected to the laptop via a USB2 port.

The Master Light Controller interfaces the computer with the luminaires, by forwarding switching and dimming commands and reporting the luminaire statuses (on, off, dimming

level, *etc.*). We adopt the WAGO-I/O-SYSTEM 750 through a DALI (*Digital Addressable Lighting Interface*) Multi-Master Module 753-647 [54]. As with most such controllers, it connects to the computer via Ethernet TCP/IP, enabling therefore IoT and cloud-based intelligence. The connection is accommodated via proprietary software, *i.e.* the WAGO DALI Configurator [55] running on the computer. The WAGO DALI Configurator allows easy commissioning of the devices connected on the DALI network. This includes the offline configuration of the entire DALI network, including the electromagnetic control gears (ECGs), the sensors and the saving/repeating device configurations.

As communication BUS between the Master Light Controller and the luminaires *ILS* adopts the established DALI BUS. DALI is one of the simplest duplex-communication protocols, which allows to flexibly connect up to 64 devices in series, grouping them into up to 16 clusters.

Our proposed “*invisible light switch*” turns the luminaires on and off automatically, based on the people position. Additionally, it uses the scene light estimation and human VFOA to simulate lighting perception when changing the luminaire dimming level. Experiments in the next section illustrate how to save up to 66% energy with minimally-perceived illumination changes.

IV. EXPERIMENTS

In this section we evaluate the performance of the system. First, we benchmark the accuracy of the scene illumination map (Sec. IV-B) and of the detector and head-pose estimator (Sec. IV-C). Then we evaluate the effectiveness of *ILS* (Sec. IV-D). Next we illustrate the dataset and metrics.

A. Dataset and Metrics

There is currently no ground-truthed dataset for benchmarking light measurements and relating it to human perception. We define therefore the first such benchmark, gathering RGBD images and light measurements in 5 diverse scenes, with different illumination scenarios (combinations of on/off light sources for each scene) and defining a suitable metric. Scenarios include offices, meeting rooms and resting areas as depicted in Fig. 5. For two scenarios (*c.f.* Figs. 5b,c) we provide detailed 3D CAD models, as well as an accurate labelling of the object textures and reflectivities. This allows the comparison to ReLux, a state-of-the-art commercial software for light modelling (*c.f.* next subsection). Moreover, for two other scenarios, we provide annotated human activity (watching TV, working at the desk, chatting, *etc.*), detailing the people position and their visual frustum of attention (VFOA) – *c.f.* Figs. 5g,h – as well as the illumination reaching their sight (*i.e.* within in their VFOA). This allows benchmarking lighting efficiency, in relation to human perception.

The new dataset consists of 130 registered RGB + depth image pairs, acquired by a Kinect v2 camera installed in the ceiling of the 5 rooms at a height of 2.8m. For each frame, the illumination maps are ground-truthed by a number of synchronized luxmeters (depicted in Fig. 5 as red boxes), to measure the light intensity reaching the specific spot, in

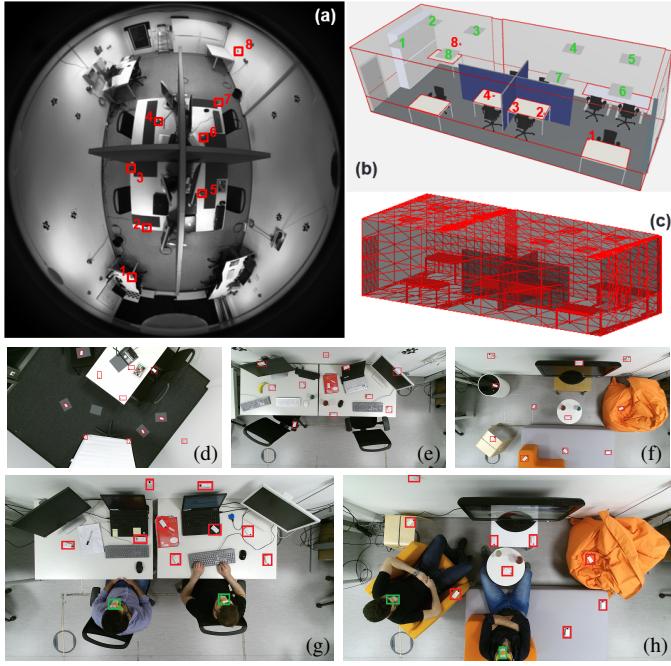


Fig. 5: Illustration of the 5 dataset scenarios (5 rooms). (a) shows Room 1, from the top-view camera. (b) and (c) are detailed CAD models of Room 1, illustrating the luminaire positions (green boxes) and the scene subdivision into patches (c.f. III-A) respectively. Room 2 (not shown) is similar to Room 1, but it has no panels across the central four desks. (d), (e) and (f) depict Rooms 3, 4, 5, also used to evaluate the illumination estimation quality (note the red boxes, highlighting the positions of the luxmeters for light measurement). (g) and (h) illustrate Rooms 4 and 5 additionally featuring human activity (note the green boxes on the occupants forehead, indicating the worn luxmeters).

lux. Depending on the scene, we installed between 8 and 11 luxmeters. We set to measure the light estimation quality by the absolute error in *lux*, compared to the luxmeter readings.

To assess the quality of the human-centric system, on the one hand we benchmark people detection and (quantized) pose estimation accuracy by the established metrics of mean average precision (mAP) and classification accuracy. Both measures are quantified on the external larger dataset of [56]. This includes 123900 images (89180 train + 34720 test), mainly featuring a single person from a top-view camera. All frames are labelled with bounding boxes of people. Here we additionally label all frames with the people head pose.

On the other hand, we assess the accuracy in estimating the perceived illumination by each person by additional luxmeters, worn by the users on their forehead (see green boxes in Figs. 5g,h). This task, consisting of 54 labelled frames, is also quantified by the absolute errors in *lux*.

B. Ambient Light Estimation

Table I, compares our proposed light estimation approach (“Ours w/o CAD”) to the commercial software Relux, within the scenes of Rooms 1 and 2. Since Relux requires the manual input CAD design, we also test our model when provided it (“Ours w/ CAD”), as well as a number of ablation studies.

Given the CAD model, our method (“Ours w/ CAD”) outperforms the Relux software (“Relux w/ CAD”) on both rooms. We achieve an average error of 36 and 70 *lux* for

Rooms 1 and 2 across the installed luxmeter sensors (“Avg. 1-8”), vs. the Relux errors of 63 and 84 *lux* respectively. Furthermore, the computational complexity of our method allows real-time performance, while each Relux simulation requires ~ 20 minutes. In general, one may note that Room 2 is more difficult than Room 1, because the removed panels among the central desks (c.f. caption of Fig. 5) result in light rays reaching each scene part from nearly each luminaire. Note also that these errors are to be intended over an illumination range of [0,2000] *lux*. In both cases, these are good estimates for commissioning, since errors below 200 *lux* are generally acceptable in the lighting industry.

Table I also contains ablation studies “w/ CAD”, given by removing the LSC and LDC distribution curves (“no_LDC_LSC”, “no_LDC”, “no_LSC”). Results confirm that both LDC and LSC are key to best performance. It is of interest that LSC yields a larger error reduction than the LDC for both rooms. In our view, this happens because LSC properly considers the angle of impacting rays, especially down-weighting the rays coming to the sensor from the side, reflecting the sensor absorption characteristic. This benefits less in the case of LDC, as those rays emitted laterally by the luminaires are only reaching the high portion of walls (e.g. above 1.7m), where we did not measure lighting.

The last row of Table I reports performance of our full light estimation approach, which does not require any manual input CAD (“Ours w/o CAD”). When only using the RGBD image, we may only estimate light propagation to those scene parts which are within the camera field-of-view (FOV). Since the camera is placed around the room center, we leave luxmeters 1 and 8 at the scene corners out (c.f. Fig. 5a), and only provide the average errors for luxmeters 2-7 (61 and 99 *lux* for room 1 and 2 respectively).

Dealing with the estimated room geometry instead of the manual CAD is clearly more challenging, partly because of the camera FOV, partly because the room reconstruction is effectively a noisy 2.5D. In more detail, the camera FOV only covers 40% of the actual scene, as it misses the ceiling and most of the walls. On the other hand, the camera-based scene reconstruction only allows a noisy estimation of those surfaces which are visible from the camera viewpoint. We set to discern which factor matters the most by the ablation method “Ours w/ CAD (within camera FOV)”, which considers a manual input CAD, cropped according to the camera FOV. As shown in Table I, this performs closely to “Ours w/ CAD” (the slightly lower error for room 2 seems due to the challenging scene with entangled lights, where removing walls reduces some of the possibly misestimated light rays). We conclude that most of the added challenge comes from the noise in the point-cloud and in the effective 2.5D estimate, resulting in a geometry with holes. However, in contrast to all other techniques, the proposed “Ours w/o CAD” is the sole fully-automatic.

Table II illustrates similar performance of our full automatic technique on the rooms 3, 4 and 5, depicted in Fig. 5. Additionally, we introduce two sets of *dynamic* experiments in rooms 4 and 5. Here *dynamic* means that scenes contain people engaged in activities with different VFOAs and interacting with scenes, e.g. moving chairs, monitors and laptops. As it

		Room 1										Room 2									
		Avg. Light Estimation Error in lux								Avg. Light Estimation Error in lux											
		Luxmeters								Luxmeters											
		1	2	3	4	5	6	7	8	Avg. 1-8	Avg. 2-7	1	2	3	4	5	6	7	8	Avg. 1-8	Avg. 2-7
Relux w/ CAD		167	96	27	26	43	10	96	39	63	50	206	97	27	80	97	49	73	44	84	71
Ours w/ CAD		69	24	22	38	28	28	38	41	36	30	70	57	76	106	75	69	55	53	70	73
Ours w/ CAD (no_LDC_LSC)		188	150	33	45	43	34	91	65	81	66	207	114	99	148	105	117	93	81	120	112
Ours w/ CAD (no_LSC)		199	152	29	41	40	33	95	57	81	65	213	117	82	125	97	97	86	63	110	100
Ours w/ CAD (no_LDC)		73	45	24	32	40	34	46	52	43	37	69	80	98	136	70	84	56	62	82	87
Ours w/ CAD (within camera FOV)		-	64	28	20	17	22	52	-	-	34	-	54	36	59	101	69	54	-	-	62
Ours w/o CAD (within camera FOV)		-	53	41	67	68	40	98	-	-	61	-	98	90	85	136	108	77	-	-	99

TABLE I: Comparative evaluation of our approach vs. the Relux commercial software, alongside ablation studies. Relux is only tested in the case of manual input CAD model (“Relux w/ CAD”). We test it against ours, which does not need the CAD (*last row*, “Ours w/o CAD”), and against ours with the provided CAD (*second row*, “Ours w/ CAD”) for fairness. “no_LSC” and “no_LDC” stand for removing the light emitter and sensor distribution curves. “within camera FOV” in the last two rows means cropping the 3D scene to the camera field-of-view (FOV). In the table, only “Ours w/o CAD (within camera FOV)” uses an estimated 3D scene geometry. Please refer to Sec. IV-B for more details.

can be seen results are in line with previous experiments.

		Avg. Light Estimation Error in lux											
		Luxmeters								Luxmeters			
		1	2	3	4	5	6	7	8	9	10	11	Avg.
Static scene	Room 3	70	93	69	23	25	28	59	28	49	71	82	54
	Room 4	18	23	41	26	76	23	35	34	40	69	31	38
	Room 5	35	38	23	27	29	49	23	40	33	25	34	32
Dynamic scene	Room 4 (dynamic)	62	26	68	65	48	57	44	30	28	-	-	48
	Room 5 (dynamic)	35	34	44	20	32	40	24	28	27	-	-	31

TABLE II: Light estimation errors across rooms 3-5, obtained with the proposed approach (w/o CAD). *Dynamic* scenes are additional image sets, including people, activities and scene interactions. See Sec. IV-B for details.

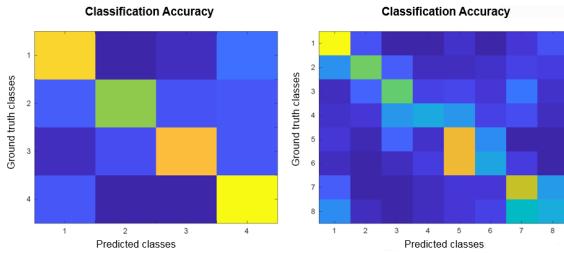


Fig. 6: Confusion matrices of the head pose estimator. From left to right, the 4 and 8 classes confusion matrix respectively.

C. People Detection and Head-pose Estimation

We attain best person detection performance by pre-training the detector on the large MS-COCO dataset [57] (80k training + 35k validation images) and then fine-tuning on a selection of diverse top-view images from [56]. Following [56], we select diverse frames by sampling one every 20 frames, which yields 4459 training + 1736 test images. On this test set, we achieve 98% mAP (IoU=0.5).

For estimating the VFOA, we train the tiny head-pose estimator of [51] on the above-named selection of training images from [56], after having labelled all dataset. We test on head-pose angles quantized into 4 and 8 classes and obtained accuracy of 70.7% and 43.2% respectively on our test set. Out of visual inspection of the corresponding confusion matrices

		Avg. Human Light Perception Error in lux									
		Luxmeters (head-worn)					Luxmeters (head-worn)				
		1	2	3	4	Avg.	1	2	3	4	Avg.
Room 4 (dynamic)	Estimated head pose	216	166			191					
	Oracle head pose	98	92			95					
Room 5 (dynamic)	Estimated head pose	55	152			104					
	Oracle head pose	42	69			55					

TABLE III: Evaluation of error in estimating the light perception of people with our proposed method, compared to the ground-truthed values provided by head-worn luxmeters. We also report error values for the cases where the head pose is provided (*oracle*). Overall the error accounted for by the head pose estimator is significant, compared to the quality of the light estimation.

in Fig. 6, it shows that the degraded performance in the 8-class case is due to adjacent viewing angle confusion. This can be further explained by the *tiny* head region of people in the images, of just 40 × 50 pixels. Thus, we considered the 4-quantized-angle head pose estimator.

D. Human-centric Light Estimation

In Table III, we illustrate the error of our method in estimating the amount of illumination arriving at the people sight. We assume as ground truth the illumination estimation of a luxmeter which the occupants wear on their forehead. Average errors for rooms 4 and 5 are 191 and 104 lux respectively. The figures are relatively large, compared to the average light estimation errors reported in Tables I,II. These are due to challenges in the light propagation as well as in the head pose estimation. We set therefore to estimate errors when the head pose is given by an oracle.

Light estimation errors for the case of oracle head-pose are also reported in Table III and are significantly lower (~ 50%). This shows that there is much progress needed in head pose estimation. Still, the residual error remains larger in average, compared to the scene illumination estimates (95 vs. 48 lux for Room 4; 55 vs. 31 lux for Room 5). We explain the discrepancy by two main factors: **a**) the reconstructed 3D mesh is less accurate for the people heads than on desks; and **b**) the light estimation arriving at the people suffers from the limited

		Room 4 (dynamic)			Room 5 (dynamic)			
Active luminaires (out of 8 available, 1 2 3 4 5 6 7 8)		3 4 7 8	2 3 4 5	3 4	1 2 3 4 5 6	2 3 4 5	1 3 4 6	3 4
Luxmeters (head-worn)	1	Δ_{watt} (w.r.t. full-lit)	387.2	387.2	580.8	193.6	387.2	387.2
	1	Δ_{lux} (w.r.t. full-lit)	116.15	123.77	189.01	106.52	148.12	157.07
	1	Light est. error (w.r.t. GT)	167.2	144.09	102.73	22.94	12.97	13.59
	2	Δ_{lux} (w.r.t. full-lit)	97.68	125.15	169.72	99.17	154.28	167.93
		Light est. error (w.r.t. GT)	194.63	171.74	131.55	9.4	241.12	2.81
								203.69

TABLE IV: Quantitative evaluation of how much power may be saved by *ILS*, Δ_{watt} , by switching off luminaires that are not directly affecting the human light perception of the occupants, in rooms 4 and 5. The table additionally reports the difference in the perceived illumination in lux Δ_{lux} , measured by luxmeters worn on the occupants' forehead. Finally, the table reports the error in lux when estimating the illumination in sight of each person, which results in an end-to-end solution. See also Fig. 7 for illustration.

FOV of the depth sensor, since it excludes parts of walls, an important factor for the head-worn luxmeter facing them.

E. Human-centric Light Management System

Table IV evaluates whether we can effectively adopt our proposed end-to-end system, *i.e.* *ILS*, for light management, implementing the “*invisible light switch*” principle. The table reports experiments for rooms 4 and 5, whereby occupants engage in activities, while we change the room illumination and switch off some of the 8 luminaires (setup 1|2|3|4|5|6|7|8 refers to “full-lit” where all luminaires are on, while 3|4 means keeping only luminaires 3 and 4 on). Fig. 7 illustrates the setup of room 4 and the resulting illumination maps, under different lighting setups.

The main performance measure in Table IV is Δ_{watt} , which quantifies how much power can be saved when turning some luminaires off. For example, when keeping only luminaires 3 and 4 on (setup 3|4), one may save up to 580.8 *watt*. Over a full working day (8 hours), this setup allows to save up to 99 *KWh* (including the power consumption of the processing unit), meaning 66% of energy efficiency.

The second important measure in the table is Δ_{lux} , which quantifies how much the illumination in sight changes. We measure this by means of luxmeters which the occupants wear on their forehead, as also previously described in Sec. IV-D. For example, in room 4 and setup 3|4, the person on the right (*c.f.* Fig. 7) wearing the head-worn luxmeter 1 perceives a difference of 189.01 lux (over the ambient lighting of 1200 lux), the largest in the most-energy saving scenario.

Finally, we report the light estimation error. In the above example (room 4, occupant 1, setup 3|4) the error is 102.73 lux . This remains comparable to the actual light variation Δ_{lux} .

Fig. 7 supports the results quantitatively. In more detail, in Fig. 7d we illustrate the estimated illumination maps across the four lighting scenarios, from the full lit (1|2|3|4|5|6|7|8) to the most energy-saving (3|4). The little change over the space in front of the people (desks, monitors) substantiates adopting the proposed invisible light switch in light management systems.

V. CONCLUSION

We have motivated, introduced and benchmarked the first real-time, end-to-end and human-centric light management system. *ILS* is based on models for estimating the scene

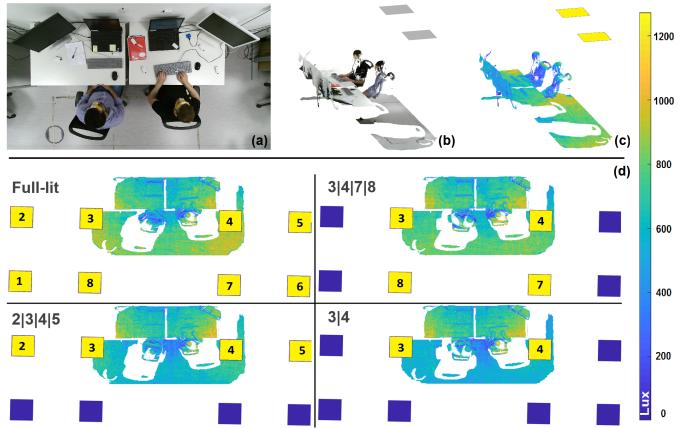


Fig. 7: Illustration of the scene, person activity and illumination map for room 4. (a) shows the top-view image, corresponding to the 3D in (b) with mapped textures. (c) and (d) present illumination maps, from 3D- and top-view respectively. In more detail, (d) shows four lighting setups, from the full lit (1|2|3|4|5|6|7|8) to the most energy-saving (3|4). Note how the illumination map over the space in front of the people (desks, monitors) is only minimally affected across the setups, which justifies proposing the invisible light switch.

lighting as well as the illumination in sight of the scene occupants. Based on these, we have defined an invisible light switch light management algorithm, to switch off or dim luminaires which are partially visible. Key aspects of this proposition are the relatively small light estimation errors, compared to lighting industry standards, which support the proposed “*invisible light switch*”.

ACKNOWLEDGMENTS

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 676455 and has been partially supported by the project of the Italian Ministry of Education, Universities and Research (MIUR) ”Dipartimenti di Eccellenza 2018-2022”.

REFERENCES

- [1] T. Partonen and J. Lönnqvist, “Bright light improves vitality and alleviates distress in healthy people,” *J. of Affective Disorders*, 2000.
- [2] J. Phipps-Nelson, J. R. Redman, D.-J. Dijk, and S. Rajaratnam, “Daytime exposure to bright light, as compared to dim light, decreases sleepiness and improves psychomotor vigilance performance.” *Sleep*, 2003.
- [3] D. Serghides, C. Chatzinikola, and M. Katafygiotou, “Comparative studies of the occupants behaviour in a university building during winter and summer time,” *International Journal of Sustainable Energy*, 2015.
- [4] S. Tanaka, M. Yoshikata, M. Miki, and T. Hiroyasu, “An evolutionary optimization algorithm to provide individual illuminance in workplaces,” in *Conference on Systems, Man, and Cybernetics*, Oct 2009.
- [5] Y. Taniguchi, M. Miki, T. Hiroyasu, and M. Yoshimi, “Preferred illuminance and color temperature in creative works,” in *Conference on Systems, Man, and Cybernetics*, Oct 2011.
- [6] R. Kralikova, M. Pinsova, and B. Hricova, “Lighting quality and its effects on productivity and human health,” *Int. J. Interdiscip. Theory Pract*, 2016.
- [7] R. Kralikova, M. Andrejiova, and E. Wessely, “Energy saving techniques and strategies for illumination in industry,” *Procedia Engineering*, 2015.
- [8] X. Zhou, D. Yan, T. Hong, and X. Ren, “Data analysis and stochastic modeling of lighting energy use in large office buildings in China,” *Energy and Buildings*, 2015.
- [9] M. S. REA, “The IESNA Lighting Handbook: Reference & application.” in *The IESNA Lighting Handbook: Reference & application*. Illuminating Engineering Society of North America, NY, 2000.

- [10] IALD. (2018) International Association of Lighting Designers (IALD). <https://www.iald.org/Advocacy/Advocacy/Quality-of-Light>.
- [11] A. Kaminska and A. Oadowicz, "Lighting control including daylight and energy efficiency improvements analysis," *Energies*, 2018.
- [12] X. Guo, D. Tiller, G. Henze, and C. Waters, "The performance of occupancy-based lighting control systems: A review," *Lighting Research & Technology*, 2010.
- [13] A. Pandharipande and D. Caicedo, "User localization using ultrasonic presence sensing systems," in *Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2012.
- [14] C. de Bakker, M. Aries, H. Kort, and A. Rosemann, "Occupancy-based lighting control in open-plan office spaces: A state-of-the-art review," *Building and Environment*, 2017.
- [15] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *ICCV*, 2017.
- [16] T. Tsesmelis, I. Hasan, M. Cristani, F. Galasso, and A. Del Bue, "RGBD2lux: dense light intensity estimation with an rgbd sensor" *WACV*, 2019.
- [17] T. Tsesmelis, I. Hasan, M. Cristani, A. Del Bue, and F. Galasso, "Human-centric light sensing and estimation from rgbd images: The invisible light switch," *WACV*, 2019.
- [18] C. Cuttle, "Towards the third stage of the lighting profession," *Lighting Research & Technology*, 2010.
- [19] P. D. Hiscocks and P. Eng, "Measuring luminance with a digital camera," Syscomp Electronic Design Limited, Tech. Rep., 2014.
- [20] T. L. Choo and O. H. See, *Illuminance Calibration of A Digital Camera based on Image Colour*. NatGrad2015, 2015.
- [21] H. Cai, "Luminance gradient for evaluating lighting," *Lighting Research & Technology*, 2016.
- [22] R. Deeb, D. Muselet, M. Hebert, and A. Tremeau, "Interreflections in computer vision: A survey and an introduction to spectral infinite-bounce model," *Journal of Mathematical Imaging and Vision*, 2018.
- [23] R. Ramamoorthi and P. Hanrahan, "A signal-processing framework for inverse rendering," in *SIGGRAPH*. ACM, 2001.
- [24] J. F. Hughes and J. D. Foley, *Computer graphics: principles and practice*. Pearson Education, 2014.
- [25] S. Marschner and P. Shirley, *Fundamentals of computer graphics*. CRC Press, 2015.
- [26] J. Krivanek, A. Keller, I. Georgiev, A. S. Kaplanyan, M. Fajardo, M. Meyer, J.-D. Nahmias, O. Karlik, and J. Canada, "Recent advances in light transport simulation: Some theory and a lot of practice," ser. *SIGGRAPH*, 2014.
- [27] E. Veach and L. J. Guibas, "Optimally combining sampling techniques for monte carlo rendering," in *SIGGRAPH*. ACM, 1995.
- [28] M. F. Cohen and D. P. Greenberg, "The hemi-cube: A radiosity solution for complex environments," *SIGGRAPH Comput. Graph.*, Jul. 1985.
- [29] A. Keller, "Instant radiosity," ser. ACM SIGGRAPH, 1997.
- [30] H. W. Jensen, "Global illumination using photon maps," in *EGWR*, 1996.
- [31] T. Hachisuka, S. Ogaki, and H. W. Jensen, "Progressive photon mapping," *ACM Trans. Graph.*, Dec. 2008.
- [32] (2010) Relux informatik ag: Reluxdesktop. <https://relux.com/en/>.
- [33] (1994) Dial gmbh: Dialux. <https://www.dial.de/en/dialux/>.
- [34] (1990) Lighting analysts, inc. <http://www.agi32.com>.
- [35] (2006) Ward g.: Radiance renderer. <http://radsite.lbl.gov/radiance>.
- [36] (2016) Vrviz research center: Hilite. <http://www.vrviz.at/projects/hilite>.
- [37] K. Krösl, C. Luksch, M. Schwärzler, and M. Wimmer, "LiteMaker: Interactive Luminaire Development using Progressive Photon Tracing and Multi-Resolution Upsampling," in *VMV*, 2017.
- [38] C. Luksch, R. F. Tobler, T. Mühlbacher, M. Schwärzler, and M. Wimmer, "Real-time rendering of glossy materials with regular sampling," *Vis. Comput.*, 2014.
- [39] M. A. ul Haq, M. Y. Hassan, H. Abdullah, H. A. Rahman, M. P. Abdullah, F. Hussin, and D. M. Said, "A review on lighting control technologies in commercial buildings, their performance and affecting factors," *Renewable and Sustainable Energy Reviews*, 2014.
- [40] A. Williams, B. Atkinson, K. Garbesi, E. Page, and F. Rubinstein, "Lighting controls in commercial buildings," *Leukos*, 2012.
- [41] A. Tsangrassoulis, A. Kontadakis, and L. Doulos, "Assessing lighting energy saving potential from daylight harvesting in office buildings based on code compliance & simulation techniques: A comparison," *Procedia Environmental Sciences*, 2017.
- [42] A. Pellegrino, V. R. M. Lo Verso, L. Blaso, A. Acquaviva, E. Patti, and A. Osello, "Lighting control and monitoring for energy efficiency: a case study focused on the interoperability of building management systems," *Transactions on Industry Applications*, 2015.
- [43] J. Kim and R. De Dear, "Workspace satisfaction: The privacy-communication trade-off in open-plan offices," *J. of Environmental Psychology*, 2013.
- [44] H. Wang, A. Pandharipande, D. Caicedo, and P. P. J. v. d. Bosch, "Distributed lighting control of locally intelligent luminaire systems," in *Conference on Systems, Man, and Cybernetics (SMC)*, Oct 2012.
- [45] A. Pandharipande and D. Caicedo, "Adaptive illumination rendering in led lighting systems," *Transactions on Systems, Man, and Cybernetics: Systems*, Sep. 2013.
- [46] Q. Zhang, L. Xu, and J. Jia, "100+ times faster weighted median filter (wmf)," in *CVPR 2014*.
- [47] S. Paris, P. Kornprobst, and J. Tumblin, *Bilateral Filtering*. Hanover, MA, USA: Now Publishers Inc., 2009.
- [48] Q.-Y. Zhou, J. Park, and V. Koltun, "Open3D: A modern library for 3D data processing," *arXiv:1801.09847*, 2018.
- [49] R. Basri, D. Jacobs, and I. Kemelmacher, "Photometric stereo with general, unknown lighting," *IJCV*, 2007.
- [50] T. Tsesmelis, I. Hasan, M. Cristani, A. Del Bue, and F. Galasso, "LIT: A system and benchmark for light understanding," in *ICCVW*, 2017.
- [51] I. Hasan, T. Tsesmelis, F. Galasso, A. Del Bue, and M. Cristani, "Tiny head pose classification by bodily cues," in *ICIP*, 2017.
- [52] L. Masset, O. Brüls, and G. Kerschen, "Partition of the circle in cells of equal area and shape," Structural Dynamics Research Group, Aerospace and Mechanical Engineering Department, University of Liege, Institut de Mécanique et Genie Civil (B52/3), Tech. Rep., 2011.
- [53] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.
- [54] Wago dali module. <https://www.wago.com/us/controllers-bus-couplers-i-o/dali-multi-master/p/753-647>.
- [55] Wago dali configurator. <https://www.wago.com/de/dali>.
- [56] M. Demirkus, L. Wang, M. Eschey, H. Kaestle, and F. Galasso, "People detection in fish-eye top-views," in *VISAPP*, 2017.
- [57] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," in *ECCV*, 2014.