

Human-centric light sensing and estimation from RGBD images: The invisible light switch

Theodore Tsesmelis^{1,2,3}, Irtiza Hasan^{3,1}, Marco Cristani^{2,3}, Alessio Del Bue^{2,†}, Fabio Galasso^{1,†}

Corporate Innovation OSRAM GmbH¹, Istituto Italiano di Tecnologia (IIT)², University of Verona (UNIVR)³

t.tsesmelis@osram.com, irtiza.hasan@univr.it

Abstract

Lighting design in indoor environments is of primary importance for at least two reasons: 1) people should perceive an adequate light; 2) an effective lighting design means consistent energy saving. We present the Invisible Light Switch (ILS) to address both aspects. ILS dynamically adjusts the room illumination level to save energy while maintaining constant the light level perception of the users. So the energy saving is invisible to them. Our proposed ILS leverages a radiosity model to estimate the light level which is perceived by a person within an indoor environment, taking into account the person position and her/his viewing frustum (head pose). ILS may therefore dim those luminaires, which are not seen by the user, resulting in an effective energy saving, especially in large open offices (where light may otherwise be ON everywhere for a single person). To quantify the system performance, we have collected a new dataset where people wear luxmeter devices while working in office rooms. The luxmeters measure the amount of light (in Lux) reaching the people gaze, which we consider a proxy to their illumination level perception. Our initial results are promising: in a room with 8 LED luminaires, the energy consumption in a day may be reduced from 18585 to 6206 watts with ILS (currently needing 1560 watts for operations). While doing so, the drop in perceived lighting decreases by just 200 lux, a value considered negligible when the original illumination level is above 1200 lux, as is normally the case in offices.

1. Introduction

People generally do not consider the impact that indoor illumination has on the monthly costs of large environments such as offices or warehouses. At the same time, they are pretty sensible to illumination, especially during office activities such as drawing and studying or doing precision works. As a consequence, office illumination is often al-

ways ON, at the maximum available lighting level, which increases the energy consumption.

The works of Kralikova and Zhou *et al.* [22, 43], show that the lighting consumption of a building can take more than 15% of the overall electricity consumption. While at peak periods, this can reach up to approximately a fourth or even more. It is clear that savings in the lighting are usually most evident and most easily feasible especially in environments where the human occupancy is limited. However, in dynamic environments where the human presence is more evident the power saving strategies are becoming more complex and harder to be addressed. The base energy saving techniques and strategies usually focus on the following principles: **a)** maximise the use of daylight; **b)** make lighting control as local as possible and get staff involved in energy saving planning; **c)** use bright coloured walls and ceilings; **d)** utilize and adjust the light sources to the most energy efficient lamp/luminaire combinations (see Figure 1).

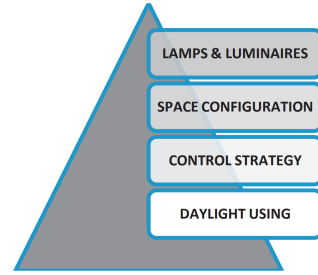


Figure 1. The pyramidal scheme shows energy saving strategies [35]. From the bottom to the top this implies the exploitation of natural light, the distribution and control of lighting locally, efficient space configuration with bright colors and open spaces for better light propagation and utilization of the light sources driven by energy saving customization. The order of the strategies in the pyramid shows the importance that each action should be applied in a green-oriented building maintenance.

However, lighting can be used for much more than to illuminate. It can enhance productivity, creating flexible spaces that adapt to the task at hand. Energy-efficient light-

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

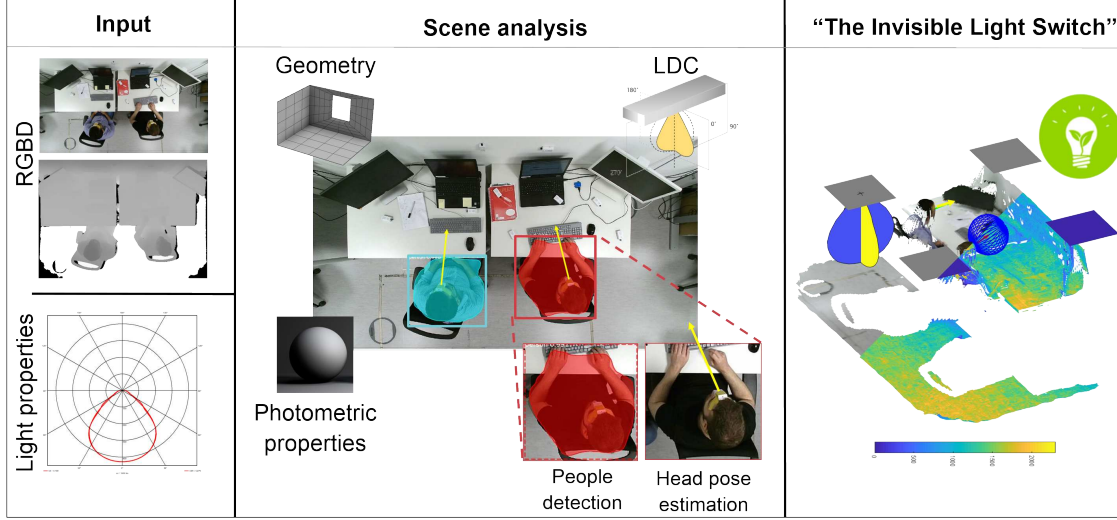


Figure 2. Overall pipeline of our system. We first acquire the RGBD input from the camera system (*left*) and together with the lighting system properties we use this information to create the Invisible Light Switch (ILS). That is, structuring the geometry of the scene, extracting the photometric properties of the material and applying a human centric analysis from where we detect the human presence in the scene and extract the possible head poses. Lastly we utilize the output of the scene analysis as the "Invisible Light Switch" application targeting a power saving framework.

ing solutions for industry can reduce environmental impact and save on costs, while at the same time increasing the life quality and productivity.

The International Association of Lighting Designers [2] states that optimal lighting consists of achieving an optimal balance among *human needs*, *architectural considerations*, and *energy efficiency* (Fig. 3).

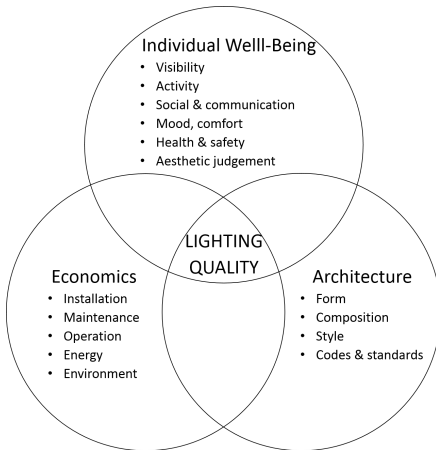


Figure 3. Framework of lighting quality according to the International Association of Lighting Designers (IALD) [2].

In this paper we present the Invisible Light Switch (ILS), a smart lighting framework for dynamically adjusting the illumination level in an indoor environment. ILS takes into account the geometry of the scene, the presence of people and their light perception with the goals of maximizing the

human comfort in terms of perceived light and, at the same time, with the lowest cost in terms of energy consumption. Our framework builds upon a light estimation system capable of estimating the light in a given 3D point of a multi luminaire indoor environment [38]. In this work [38], the radiosity model has been customized to take into account a realistic model of light propagation, outclassing even industrial software in the task.

This paper enriches the model by including the human aspect, and showing how the interplay between the light estimation system and the human activity may lead to a consistent energy saving framework. The invisible light switch summarises the idea: an individual has the feeling of an environment which is globally illuminated, while in reality an automated light switch dims the luminaires in a way which is invisible to the users. This is possible by estimating the position of a person in the sensed environment, its head orientation, and understanding the light which is perceived by him. In fact, the lighting sensed by a human can be assumed as the light contained in a conic volume departing from the mean point connecting the humans eyes in the direction of the nose. Given this, it is possible to determine which luminaries could be switched off/dimmed down while maintaining the level of perceived light unchanged. The head pose is provided by detecting the person first and then estimating the head orientation. The former is carried out by means the state-of-the-art detector Mask R-CNN [20] with ResNet [21] as a backbone architecture, while head pose is done using Hasan's *et al.* method [19].

To test the system, a novel dataset has been built where 2

people are present in two rooms with 2 portable luxmeters attached to their forehead, well suited to mimic the human perception. This provides us with the ground truth information which can be considered when evaluating the light estimations of the system, given a particular setup of the luminaires. Experiments show how reliable is the system in detecting the people position along with head orientations and the related perceived light. A margin of 100 Lux error is observed within a global illumination estimation of 1200 lux. As reported by [29], this delta is barely perceivable by the human eye, so the error can be considered within the accepted range. Thereafter, it is shown that the ILS allows to heavily modify the illumination setup by only affecting the human perception of a delta of 200 Lux, still within the range of non-perceivable changes. We finally show that our system is promising in terms of energy saving: since in the most aggressive scenario could indicate for up to 66% power efficiency.

The rest of the paper is organized as follows, next we review related work (Sec. 2) and define the overall proposed pipeline (Sec. 3). We present the results and evaluation in (Sec. 4). Finally, we conclude in (Sec. 5).

2. Related Work

2.1. Lights and Behaviour

Relationship between human activities and lights is a widely studied topic in perceptual sciences [5, 13, 15]. Recently, it was shown by [41] that light intensifies people's perception. It triggers emotional system leading to intensified effective reactions. Light changes our perception of space [14], we tend to associate different illumination patterns to different social gatherings (musical concert *vs.* candle light dinner). People seem to share more details in bright light than darkness [10], we as human beings also rely on facial expressions which are only visible in light. Moreover, light provides sense of security [15], people choose roads and streets in night due to the illumination [36]. Recently, studies targeting the office environments revealed a strong connection between people's productivity and the lights [22, 23, 33]. Eyeing the importance of lighting on humans, related communities such as Human Computer Interaction (HCI) [28] deployed interactive lighting in a city square, providing a sense of "belongingness" to the residents. Furthermore, ubiquitous computing [16] and architectural design [26] have also investigated this topic. However, there are also studies that question the relationship between the light perception and the actual measured spatial illumination [9, 29].

2.2. Modelling human activities

Despite receiving a wide scale attention, the literature in computer vision seems to have ignored the modelling of

light and behaviour. Only recently Hasan and Tsesselis *et al.* [18] presented the idea of jointly modeling the relationship of light and human behavior via long term time-lapse observation of the scene by recognizing and forecasting activities using the head pose estimation as a proxy for the gaze.

Estimation of head pose is inherently a challenging task due to subtle differences between human poses. However, in the past several techniques ranging from low level image features to appearance based learning architectures were used to address the problem of head pose estimation. Previously, [17, 40] used neural networks to estimate head pose. While authors in [8] adopted a randomized fern based approach to estimate head orientation. Limited accuracy was achieved though due to several reasons such as two images of the same person in different poses appeared more similar than two different people in same pose or due to difficulty to compute low level image features in low resolution images. Recently, decision trees have been reported to achieve state of the art results [24]. However, they rely on local features and are prone to make errors when tested in real world crowded scenarios. We address the issue of having a head pose estimator that can work in unconstrained real world scenarios by utilizing the power of deep neural network models which in recent past, it has been used for pose estimation [37].

Having a strong similarity with head pose, some studies focused on estimating visual frustum (VFOA) on low resolution images [6, 30, 34, 34] together with the general pose of the person. VFOA has been used as a reliable cue for identifying social interactions: in [7] the head direction is used to estimate a 3D visual frustum as approximation of the VFOA of a person. Given the VFOA and spatial layout, human-human interactions are estimated: the concept is that people who are in a close proximity and having their VFOA intersecting with each other are engaged into a human-human interaction. The same concept has been studied in [31]. On the other hand, in [32], the VFOA was defined as a line directed towards the focus of attention by taking into account subject's gaze in low resolution images: in that work the goal was to understand the gazing behavior of people in front of a shop window. The VFOA was projected on the floor and modeled as a Gaussian distribution containing "samples of attention" in front of a person [12]: the higher concentration, depicts stronger likelihood that in that area the eyes' fixation would be present. In a physiologically motivated study [39] the VFOA is represented by an angle θ (head orientation), an aperture $\alpha = 160$ degree and a length l . The θ corresponds to the variance of the Gaussian distribution modelled around the spatial proximity of a person. In the same study, the density of attention was used to measure the likelihood of a visual fixation: a more concentrated sampling was conducted at locations closer to the

person, smoothly decreasing the frequency of sampling as one goes away. The visual frustum is constructed by sampling from the above Gaussian kernel and only considering ones confining inside the cone of attention composed by the angle α . Finally in [42], the aperture of the cone was used to study frequent and less frequent regions of interest.

2.3. Modelling light in indoor environments

In previous studies, indoor light modeling is mostly a field of research in visual computing and computer graphics. However, in these fields more emphasis is given on the generation of lifelike and photorealistic renderings rather than the actual spatial lighting measurement. On the other hand in the lighting field the focus is given on commercial CAD-design modeling software products, *e.g.* Relux [4], DIALux [1] and AGi32 [3], which are broadly used for offline measurement and evaluation of lighting solutions in a simulated environment. To the best of our knowledge, the RGBD2Lux approach [38] is the first to bridge the gap across the two fields, bringing together visual computing and lighting design. By using only RGBD input, the method obtains a dense light intensity estimation of an indoor environment.

3. Ego-light-perception

Any light management system that has to autonomously adjust the illumination of the environment has to be aware of two main factors: the human occupancy and their activity in the environment (human centric analysis) and the existing ambient illumination over time considering how is this influenced from the scene structure, the object materials and the light sources (scene composition analysis).

These two aspects are tightly intertwined, since the structure of the scene allows and constrains human activities, but at the same time the human activities influence the scene structure. Consider for example a warehouse as a scene: its structure continuously changes due to the different arrangement of the goods, the latter being a direct consequence of the human activities carried out in the environment. In other words, the structure of the scene and the human have to be considered as parts of a whole, accounting in addition for their continued temporal evolution.

To this end, the major goal in this work is to provide a new computer vision system for estimating the illumination map along with the human occupancy and attention from a single view. We do this by bringing together individual works into a unique pipeline as we show in Figure 2.

3.1. People detection and head-pose estimation

We aim to detect people and estimate their head pose (their viewing angle). For the first task we adapt the Mask R-CNN [20] object detector, while for the second one the head pose estimator proposed in [19].

The R-CNN [20] detector has the ResNet-101 [21] as a backbone architecture, trained on 80k images and 35k subset of evaluation images (trainval35k) of MS COCO dataset [25]. We fine-tune the detector on our top-view dataset (see Sec. 4.1), adopting a specific training portion of the data. We randomly partition the data into training and testing set, keeping 70% of the data for training and 30% for testing. Since the top-view images are different from the frontal-view images of the COCO dataset [25], the fine-tuning has a crucial role. We adopt a similar procedure for training the head pose estimator as in [19]. It is worth noting that the input for the head pose is the whole body detection bounding box: this is because [19] has been specifically designed for managing small-sized head patches, exploiting the body as contextual cue for a better final head orientation classification. In particular, 4 and 8 classes related to angles have been taken into account.

During testing time, a cascaded approach is followed, first by applying the people detector and then feeding the detected body bounding box as input into the head orientation module.

3.2. Spatial light estimation

To obtain an estimate of a dense spatial illumination map, we adapt our work in [38]. In this work we make use of a radiosity model [11] for estimating the spatial illumination over time by just using the input from an RGBD camera. Furthermore, we extract the information regarding the photometric properties of the material of the scene based on a photometric stereo baseline approach that is applied on the time-varying RGB images. This approach extracts a scalar albedo at each pixel by using a set of images with different light sources that are switched on/off during the day. Having the light sources position and intensity, the scalar albedo under Lambertian assumptions, and the depth map from the sensor, our proposed method in [38] shows that it is possible to obtain a dense measurement of the light emitted by a 3D patch in the indoor environment. In order to provide more realistic estimates, we model real lighting systems that, differently from point-like sources, emit light given a specific light distribution curve (LDC). The LDC is custom for each lighting system and their properties are considered to be known when estimating the light intensity. The proposed method shows that, even by accounting the non-linearities of LDC, it is possible to solve for the radiosity equation with Least Squares and so obtain a more reliable measure of the light intensity. For the evaluation of our approach we used point-to-point sensory equipment *aka.* luxmeters installed across the scene.

3.3. Gaze-gathered light modelling

Light measurements are practically made using a luxmeter sensor. This sensor measures the perceived light that is

in function of the distance to the light, the orientation and other manufacturing characteristics. These properties are resumed by the Luxmeter Sensitivity Curve (LSC) as in Figure 4a. The LSC illustrates the perception characteristic of every luxmeter sensor which in this work we adopt in order to meet the measuring requirements of the collected ground truth data and to simulate the human light perception. We have chosen this solution because this is the standard de facto in the lighting industry and it provides satisfactory solutions when doing light commissioning [27].

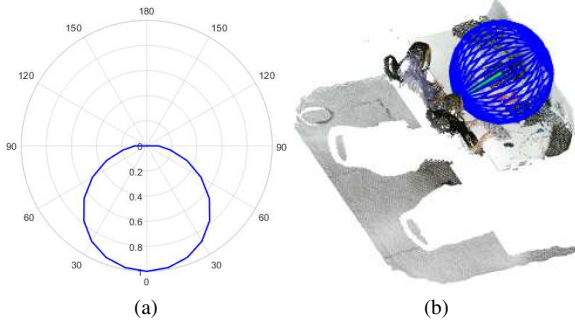


Figure 4. Modeling of the Luxmeter Sensitivity Curve (LSC) as a human light perception model.

The key idea in this procedure is that, once we have detected a person in the image and estimated his head positioning and orientation as described in Sec. 3.1, we extract his posture in the 3D space by mapping the 2D image coordinates of his detected head to the corresponding depth information. Thereafter, once we have the positioning of the head in the 3D space as well as its orientation (where the person looks at), we estimate the light that arrives to his/her face (or to the luxmeter as in our case) by applying a ray-casting procedure where we simulate the human field of view (FOV). Such view frustum is obtained by using emitted rays starting from the estimated head position towards the corresponding estimated head orientation. The total illumination arriving to the person is computed by adding the related spatial illumination (radiance) from the patches of the scene that are in the direct visibility of the person. The rays project in the space as a uniform generated sequence over the unit sphere and weighted accordingly, based on the modelled luxmeter’s LSC, towards the visible patches from the FOV of the sensor. The contribution of each patch to the total amount of lighting perceived by the occupant, is computed by estimating the percentage of rays intersecting that patch.

4. Results

Experiments are organized as follows, Sec. 4.1 presents the recorded dataset with all the different ablation studies both for light measurements as well as for top-view detection and head-pose estimation. Sec. 4.2 reports the results

regarding the person occupancy and head pose estimation study, while Sec. 4.3 describes in more details the evaluation for both spatial and the gaze-gathered light estimation. Finally, Sec. 4.4 evaluates the Invisible Light Switch as a power saving application.

4.1. Dataset overview

To the best of our knowledge, Tsesmelis’ *et al.* work [38] is the first to introduce a dataset for benchmarking light measurements with ground truth sensory data in real scenes. In this work we extended this dataset by introducing two more scenes with human activity, one based on a normal office environment and a second one representing a relaxing area (see Figure 5).

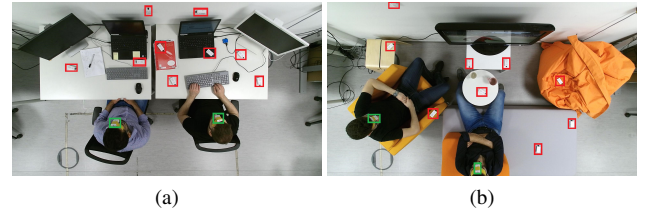


Figure 5. Illustration of the two indoor scenes used for evaluation: (a) illustrates a normal office environment and (b) shows a relaxing area. Red and green bounding boxes are showing the location of luxmeters within the space covering the spatial and gaze-gathered illumination ground truth measurements respectively.

Both scenes comprehend different human activities *e.g.* watching TV, working on a desk area, chatting, *etc.*, as well as different head orientations (VFOA) and multiple light combinations. In this work, VFOA is a cone with vertex in the middle of a person’s eyes, oriented as the gaze direction and an aperture angle of $\alpha = 30^\circ$.

In both rooms there is a controlled light management installation, where the position, type and properties (*e.g.* luminous intensity, light distribution curve, *etc.*) of the luminaires (eight in total) are considered known, see Figure 6.

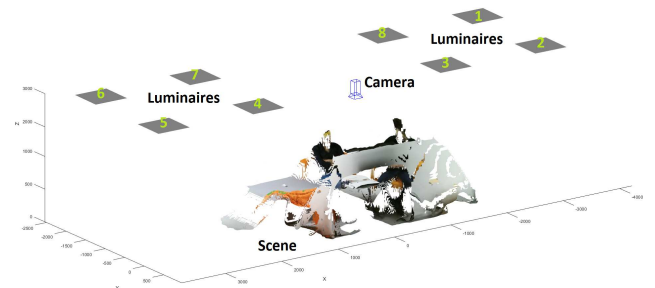


Figure 6. Illustration of the light management installation.

For obtaining the ground truth data we have installed and used a number of sensory equipment. A calibrated and aligned RGBD camera system (Kinect v2) is installed in the

Thereafter, we evaluate 24 and 30 different scenarios with different luminaire activations (luminaires switched on/off) for each room respectively (see Fig. 7). We target the use of RGB and depth input just for light measurement, the use of luxmeters as ground truth, and all other provided information for evaluation studies.

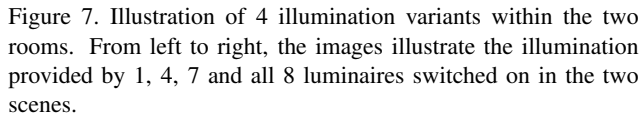


Figure 7. Illustration of 4 illumination variants within the two rooms. From left to right, the images illustrate the illumination provided by 1, 4, 7 and all 8 luminaires switched on in the two scenes.

4.2. Top-view detection and head-pose estimation

We fine tune both the person detector and the head pose estimator on our top-view dataset. We report an average precision (AP) of 98% in terms of people detection. As mentioned previously we test our approach on the testing set of our top-view dataset. For the head pose orientation fine tuning on the whole body has been crucial for the performance, since using the sole head region produced definitely worst scores. In particular, we adopt two different class numbers for head pose, namely 4 and 8. The corresponding confusion matrices are reported in Fig. 9, showing an accuracy of 43.2% (8 classes) and 70.7% (4 classes) respectively. The scarce performance in the 8-class case is due to the mix among adjacent viewing angles: actually, the average size of the head region in the dataset is approx. 40x50 pixels. For these reasons, we use the 4-class version in the light perception studies.

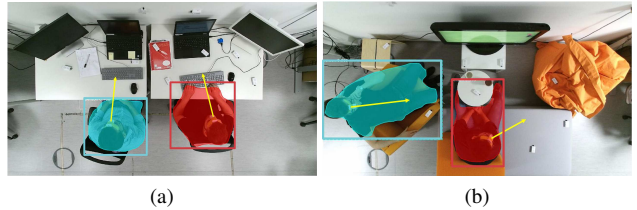


Figure 8. Illustration of people detection and head pose estimation. We detect people in the scene by using Mask R-CNN and then the detections are provided as input to the head pose estimator.

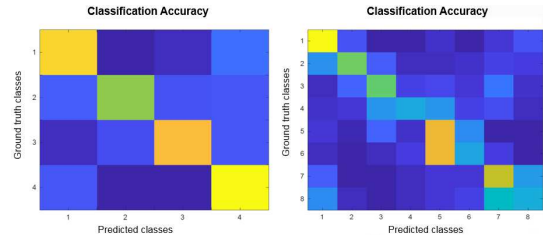


Figure 9. Confusion matrices of the head pose estimator. From left to right, the 4 and 8 classes confusion matrix respectively.

		Avg. error ε (in Lux)												Avg. (1-9)	Avg. (11-10)
		Luxmeters										11	12		
		1	2	3	4	5	6	7	8	9	10				
Scene 1	ε_{ext} (w.r.t. GT)	62.5	26.3	68.0	65.1	47.9	57.1	44.0	29.9	28.0	97.6	92.2	56.2	94.7	
	ε_{rand} (w.r.t. GT)	-	-	-	-	-	-	-	-	-	216.08	166.4	-	191.24	
Scene 2	ε_{ext} (w.r.t. GT)	35.3	33.8	44.0	20.1	31.5	39.6	23.6	27.9	27.3	41.7	69.2	35.8	55.4	
	ε_{rand} (w.r.t. GT)	-	-	-	-	-	-	-	-	-	55.42	151.93	-	103.68	

Table 1: The values represent the average estimated illumination error over the different lighting activation w.r.t. the ground truth measurements, for both scenes. Columns 1-9 corresponds to the spatial average values for the corresponding installed luxmeters in the environment. By contrast, values in columns 10-11 consider those luxmeters for evaluating the human light perception.

4.3. Person-perceived light estimation

Table 4.2 presents the quantitative results of our adopted light estimation approach. The table shows the average estimated error in lux values for both spatial (luxmeters 1-9) and gaze-gathered light estimation (luxmeters 10-11) cases. It can be easily noticed that the error, ϵ_{est} , for all luxmeters does not exceed the range of 100 lux, this yields an overall average light estimation error approx. 56 lux for Scene 1 and 36 lux for Scene 2. On the other hand, if we now consider only the luxmeters intended for evaluating the gaze-gathered light estimation, *i.e.* luxmeters 10 and 11, we notice that the error raises up to 94.7 lux and 55.4 lux for each scene respectively. This can be justified due to inaccuracies in the reconstruction of the 3D mesh areas corresponding to

the head position and orientation of the occupants, as well as to the fact that the inter-reflections from the wall towards the sensors are limited due to incomplete reconstruction as an outcome of the limited FOV of the depth sensor. In any case, the fact that the average light estimation error does not exceed 100 lux indicates that the estimated illumination map can be considered reliable for describing the global illumination of the scene.

Furthermore, to demonstrate the applicability of our model, we use as explained a real person detector and a head pose estimator (making the pipeline completely automatic). In Table 4.2 the $\varepsilon_{est,d}$ rows for column 10 and 11, illustrates the error based on the detectors output for both scene 1 and 2. It can be observed that while the average error w.r.t. the oracle is less than 100 lux, this error raises up to the range of 200 lux negative variation w.r.t. to the ground truth measurements. The last can be justified by erroneous head pose estimations, considering the large step size (90°) of the 4-class adapted classification problem. This further brings into discussion the fact that this error could further be substantially reduced by improving the head pose estimator.

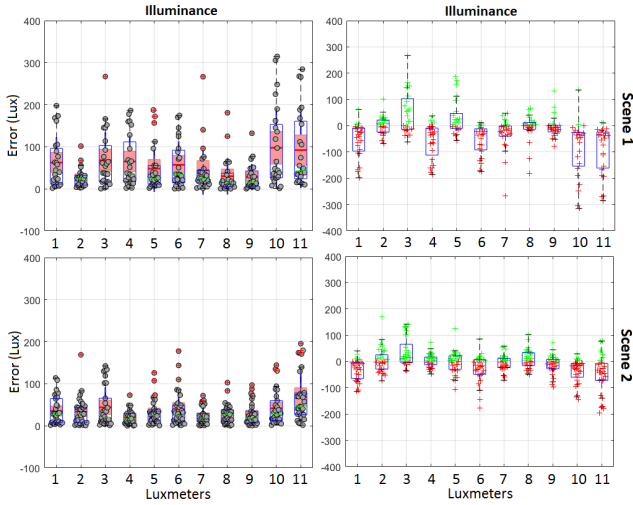


Figure 10. Scene 1 & 2 boxplot error evaluation (in Lux) using based on the presented framework. The boxplots in the first and second columns show the absolute and signed illumination estimated error for each lighting scenario in each scene respectively.

Figure 10 shows in a graph analysis the values presented in Table 4.2. The left graphs show the absolute light estimation error (y-axis), as estimated for each of the 11 (9 for spatial and 2 for the human light perception) used luxmeter sensors (x-axis). The gray dots, forming each of the box plot boxes, represent the estimated error of each of the lighting scenarios for each scene while the pink box represents the central 50% of the data. The upper and lower vertical lines indicate the extension of the remaining error points outside it and the central red line indicates the mean error

which comes in alignment with the values shown in Table 4.2. Similarly, the boxplots on the right present the signed illumination error accordingly. The green and red markers indicate whether the error is due to an over or under estimation of the illuminance at the sensor's location respectively. As it can be noticed in the most of the cases the error is a result of an under estimation of the illuminance which as explained earlier are a cause of the incomplete geometry of the scenes as we only consider the parts of the environment within the FOV of the camera sensors.

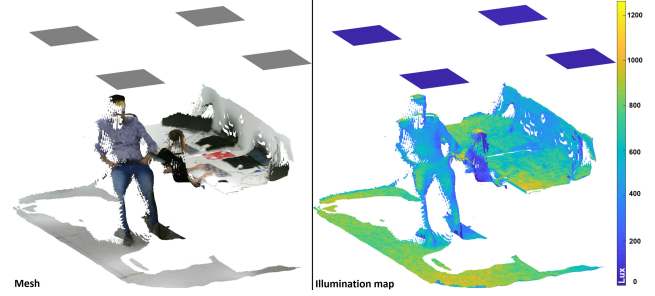


Figure 11. Illumination map of the full-lit scenario in scene 1 with a dense representation of the global illumination of the environment.

Finally, figures 11 and 12 visualise the illumination maps in the 3D space for one of the illumination scenarios in each of the scenes. As it can be seen the visualized illumination maps provide an accurate dense representation of the global illumination of the environment over time.

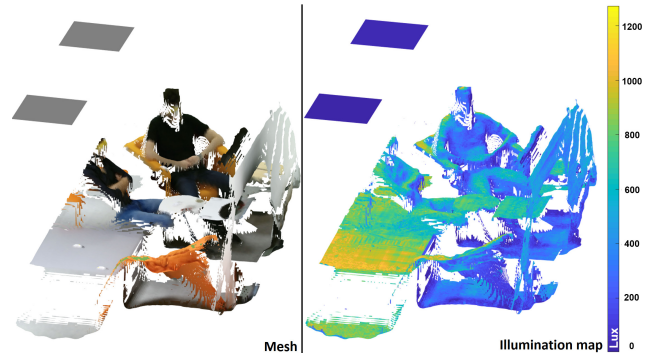


Figure 12. Illumination map of the full-lit scenario in scene 2. Notice the estimated illumination in the area in front of the occupants which is less bright in comparison to the one that are on their side. This is due to the body occlusion on the direct illumination coming from the luminaires from their back which is correctly estimated by the ILS.

4.4. The invisible light switch

The idea behind the Invisible Light Switch is straightforward: the proposed system controls and sets the illumination of the environment by taking into account the infor-

		Scene 1						Scene 2						
		VFOA 1			VFOA 2			VFOA 1			VFOA 2			
Luminaire activations		3 4 7 8	2 3 4 5	3 4	3 4 7 8	2 3 4 5	3 4	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
Luxmeter 10	Δ_{lux} (w.r.t. full-lit)	116.15	123.77	189.01	85.4	123.8	163.85	84.23	93.69	151.92	106.52	148.12	157.07	191.15
	ε_{est} (w.r.t. GT)	167.2	144.09	102.73	235.3	200.1	163.28	85.85	94.1	43.76	22.94	12.97	13.59	25.69
	Δ_{watt} (w.r.t. full-lit)	387.2	387.2	580.8	387.2	387.2	580.8	387.2	387.2	580.8	193.6	387.2	387.2	580.8
Luxmeter 11	Δ_{lux} (w.r.t. full-lit)	97.68	125.15	169.72	167.4	86.34	194.37	62.67	118.21	153.02	99.17	154.28	167.93	194.85
	ε_{est} (w.r.t. GT)	194.63	171.74	131.55	91.14	128.7	70.21	15.26	67.87	5.39	9.4	241.12	2.81	203.69
	Δ_{watt} (w.r.t. full-lit)	387.2	387.2	580.8	387.2	387.2	580.8	387.2	387.2	580.8	193.6	387.2	387.2	580.8

Table 2: Quantitative analysis of four different head orientation class studies (VFOA), two for each scene. Δ_{lux} shows the discrepancy of different lighting scenarios w.r.t. the full lit scenario (reference). ε_{est} shows the corresponding average error of the estimated light in regards to the ground truth lux measurements and Δ_{watt} shows the discrepancy of the power consumption in watts considering the active/non active luminaires for each corresponding scenario.

mation regarding the part of the scene that the user can see or cannot see, by switching off or dimming down the lights outside the user’s VFOA, and thus ensuring a consistent energy saving and productivity.

In Table 4.3 we examine the applicability of the invisible light switch from the human perspective aspect (luxmeters 10-11) for different head orientation cases (VFOA) in the two scenes. The value Δ_{lux} provides the information regarding what is the impact to the light perceived from the occupants (based on the ground truth sensor measurements) on different light source combination scenarios. As we can see this gives us a range of 0-200 lux negative variation even to the most aggressive scenario of having only two luminaires active (the ones to the direct view of the occupants each time). If we connect this with the amount of watts that we can save for this corresponding lighting scenario, *i.e.* $\Delta_{watt} = 580.8$ watt w.r.t. to the full lit case, this can give us a total power efficiency of 12379.2 KWatt through a whole day. The value ε_{est} reports the light estimation error based on our framework, which as we can see again it settles within a range of 0-200 lux overall negative variation. This error shows us how our system aligns with the ground truth measurements, *i.e.* a lower ε_{est} error the better, and whether the same pattern described above could be followed. A visual example of the VFOA 1 case for scene 1 (see Table 4.3) can be seen in Figure 13. As it can be easily noticed the estimated illumination over the desk areas have the less affect as we switch off the peripheral light sources and still providing an optimally lit scenario while it is minimally lit.

5. Conclusion

This paper highlights the importance of a human-centric aided lighting management system which targets productivity over a power saving framework. As a result, in this work

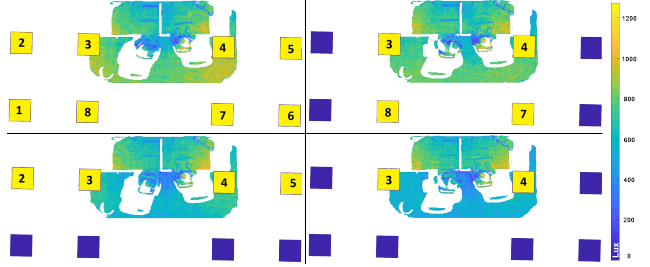


Figure 13. Qualitative illustration of the VFOA 1 ablation study for Scene 1 presented in table 4.3. The top left corner shows the illumination map of the full lit case, in comparison to three other light scenarios. As it can be seen the estimated illumination over the desk area where the two occupants have their attention is less affected in comparison to the areas behind them. This show in practice how the invisible light switch application could be established.

we proposed and evaluated a practical (application-wise) system which tries to encapsulate all these three aspects, *i.e.* ambient illumination, human activity and power efficiency. We also for the first time presented a complete system that estimates both the spatial and the individual gaze-gathered light intensity based on a camera-aided solution. We illustrated a possible 66% of power saving by deploying our framework as the “Invisible Light Switch” application which can be used to exploit an optimal illumination pattern for a given human activity.

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.

References

- [1] DIAL GmbH: DIALux. <https://www.dial.de/en/dialux/>. Accessed: 2017-11-16.
- [2] International Association of Lighting Designers, IALD. <https://www.iald.org/Advocacy/Advocacy/Quality-of-Light>. Accessed: 2018-09-11.
- [3] Lighting Analysts, INC. <http://www.agi32.com>. Accessed: 2017-11-16.
- [4] Relux Informatik AG: ReluxDesktop. <https://relux.com/en/>. Accessed: 2017-11-16.
- [5] L. Adams and D. Zuckerman. The effect of lighting conditions on personal space requirements. *The journal of general psychology*, 118(4):335–340, 1991.
- [6] S. O. Ba and J.-M. Odobez. A probabilistic framework for joint head tracking and pose estimation. In *IEEE International Conference on Pattern Recognition (ICPR)*, 2004.
- [7] L. Bazzani, M. Cristani, D. Tosato, M. Farenzena, G. Paggetti, G. Menegaz, and V. Murino. Social interactions by visual focus of attention in a three-dimensional environment. *Expert Systems*, 30(2):115–127, 2013.
- [8] B. Benfold and I. Reid. Guiding visual surveillance by tracking human attention. In *British Machine Vision Conference (BMVC)*, pages 1–11, 2009.
- [9] C. Bernecker. Square Law Dimming: Presumed Perception or Reality? Leducation, New York, 2017.
- [10] S. J. Carr and J. M. Dabbs Jr. The effects of lighting, distance and intimacy of topic on verbal and visual behavior. *Sociometry*, pages 592–600, 1974.
- [11] M. F. Cohen, J. Wallace, and P. Hanrahan. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Inc., San Diego, CA, USA, 1993.
- [12] M. Cristani, L. Bazzani, G. Paggetti, A. Fossati, D. Tosato, A. Del Bue, G. Menegaz, and V. Murino. Social interaction discovery by statistical analysis of f-formations. In *British Machine Vision Conference (BMVC)*, pages 23.1–23.12, 2011.
- [13] J. E. Flynn, C. Hendrick, T. Spencer, and O. Martyniuk. A guide to methodology procedures for measuring subjective impressions in lighting. *Journal of the Illuminating Engineering Society*, 8(2):95–110, 1979.
- [14] A. D. Galasiu and J. A. Veitch. Occupant preferences and satisfaction with the luminous environment and control systems in daylit offices: a literature review. *Energy and Buildings*, 38(7):728–742, 2006.
- [15] R. Gifford. Light, decor, arousal, comfort and communication. *Journal of Environmental Psychology*, 8(3):177–189, 1988.
- [16] F. Gil-Castineira, E. Costa-Montenegro, F. Gonzalez-Castano, C. López-Bravo, T. Ojala, and R. Bose. Experiences inside the ubiquitous oulu smart city. *Computer*, 44(6):48–55, 2011.
- [17] N. Gourier, J. Maisonnasse, D. Hall, and J. L. Crowley. Head pose estimation on low resolution images. In *International Evaluation Workshop on Classification of Events, Activities and Relationships*, pages 270–280. Springer, 2006.
- [18] I. Hasan, T. Tsesmelis, A. Del Bue, F. Galasso, and M. Cristani. “don’t turn off the lights”: Modelling of human light interaction in indoor environments. In S. Battiato, G. M. Farinella, M. Leo, and G. Gallo, editors, *New Trends in Image Analysis and Processing – ICIAP 2017*, pages 143–151. Cham, 2017. Springer International Publishing.
- [19] I. Hasan, T. Tsesmelis, F. Galasso, A. Del Bue, and M. Cristani. Tiny head pose classification by bodily cues. In *Image Processing (ICIP), 2017 IEEE International Conference on*, pages 2662–2666. IEEE, 2017.
- [20] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask r-cnn. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2980–2988. IEEE, 2017.
- [21] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [22] R. Kralikova, M. Andrejiova, and E. Wessely. Energy saving techniques and strategies for illumination in industry. *Procedia Engineering*, 100:187–195, 2015.
- [23] R. Kralikova, M. Pinosova, and B. Hricova. Lighting quality and its effects on productivity and human health. *Int. J. Interdiscip. Theory Pract*, 10, 2016.
- [24] D. Lee, M.-H. Yang, and S. Oh. Fast and accurate head pose estimation via random projection forests. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1958–1966, 2015.
- [25] 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 *European conference on computer vision*, pages 740–755. Springer, 2014.
- [26] R. Magielse, B. J. Hengeveld, and J. W. Frens. Designing a light controller for a multi-user lighting environment. 2013.
- [27] I. S. of North America and I. T. P. C. P. of Light Sources Subcommittee. *The commissioning process applied to lighting and control systems*. New York, N.Y. Illuminating Engineering Society of North America, 2011. ”IES DG-29-11.”.
- [28] E. S. Poulsen, A. Morrison, H. J. Andersen, and O. B. Jensen. Responsive lighting: the city becomes alive. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 217–226. ACM, 2013.
- [29] M. S. Rea. The IESNA Lighting Handbook: Reference & application. In *The IESNA Lighting Handbook: Reference & application*. Illuminating Engineering Society of North America. New York, NY: Illuminating Engineering Society of North America., 2000.
- [30] N. Robertson and I. Reid. Estimating gaze direction from low-resolution faces in video. In *European Conference on Computer Vision (ECCV)*, 2006.
- [31] N. M. Robertson and I. D. Reid. Automatic reasoning about causal events in surveillance video. *EURASIP Journal on Image and Video Processing*, 2011.
- [32] K. Smith, S. O. Ba, J.-M. Odobez, and D. Gatica-Perez. Tracking the visual focus of attention for a varying number of wandering people. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(7):1212–1229, 2008.

- [33] K. C. Smolders, Y. A. de Kort, A. D. Tenner, and F. G. Kaiser. Need for recovery in offices: Behavior-based assessment. *Journal of Environmental Psychology*, 32(2):126–134, 2012.
- [34] R. Stiefelhagen, M. Finke, J. Yang, and A. Waibel. From gaze to focus of attention. In *Visual Information and Information Systems*, 1999.
- [35] P. Tak. Lighting design and energy savings. *Light, FCC Public s. r. o., Prague*, Vol.10, No.2., 2009.
- [36] L. H. Taylor and E. W. Socov. The movement of people toward lights. *Journal of the Illuminating Engineering Society*, 3(3):237–241, 1974.
- [37] A. Toshev and C. Szegedy. Deeppose: Human pose estimation via deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1653–1660, 2014.
- [38] T. Tsesmelis, I. Hasan, M. Cristani, F. Galasso, and A. Del Bue. RGBD2lux: dense light intensity estimation with an rgbd sensor. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 2019.
- [39] S. Vascon, E. Z. Mequanint, M. Cristani, H. Hung, M. Pelillo, and V. Murino. Detecting conversational groups in images and sequences: A robust game-theoretic approach. *Computer Vision and Image Understanding*, 143:11–24, 2016.
- [40] M. Voit, K. Nickel, and R. Stiefelhagen. A bayesian approach for multi-view head pose estimation. In *2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pages 31–34. IEEE, 2006.
- [41] A. J. Xu and A. Labroo. Incandescent affect: Turning on the hot emotional system with bright light. *ACR North American Advances*, 2013.
- [42] L. Zhang and H. Hung. Beyond f-formations: Determining social involvement in free standing conversing groups from static images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [43] 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*, 86:275–287, 2015.