”Don’t turn off the lights”: Modelling of human light interaction in indoor environments

Irtiza Hasan[1], Theodore Tsesmelis[2], Alessio Del Bue[2], Fabio Galasso[3], and Marco Cristani[1]

[1] University of Verona, Italy
[2] Istituto Italiano di Tecnologia
[3] Corporate Innovation OSRAM GmbH

Abstract. Human activity recognition and forecasting can be used as a primary cue for scene understanding. Acquiring details from the scene has vast applications in different fields such as computer vision, robotics and more recently smart lighting. This work brings together advanced research in computer vision and the most modern technology in lighting. The goal of this work is to eliminate the need for any switches for lighting, which means that each person in the office perceives the entire office as all lit, while lights, which are not visible by the person, are switched off by the system. This can be achieved by combining lighting with presence detection and smart light control.

Keywords: Scene understanding, Activity forecasting, Activity recognition, Photometry

1 Introduction

A modern lighting system should automatically calibrate itself (determine the type and position of lights), assess its own status (which lights are on and how dimmed), and allow for the creation or preservation of lighting patterns, e.g. after the sunset. The lighting patterns should be adjusted in a way, that is optimal for people actions and locality. As most of our activities hold within a given light pattern [7] as illustrated in Fig. 1. Moreover, light influences our perception of space [8], for example we expect to see a certain illumination pattern in a musical concert etc. The essence of such a system would be to deploy an *invisible light switch*, where the change in illumination is not perceived by the user.

Furthermore, idea of a smart lighting system, is to deploy a dynamic illumination pattern for a given activity. In brief, *SCENEUNDERLIGHT* H2020-MSCA-ITN-2015 project encompasses both fundamental research in computer vision and innovation transfer in smart lighting with a goal being at researching and developing novel autonomous tools using advanced computer vision and machine learning approaches that seamlessly integrates into smart lighting systems for indoor environments.

SCENEUNDERLIGHT proposes a plan to create such an achievement, in light management systems, by enabling the understanding of the environment via long-term observation, that span days, weeks and even months, with a sensing device (i.e. RGB cameras or RGBD if including a depth sensor) for smart illumination and energy saving via an artificial intelligence (AI) processor (e.g. an algorithm to understand the scene and make decisions on lighting). More specifically in this Research and Development plan, top-view time-lapse images of the scene allow computer vision algorithms to understand it 1)To estimate the human activities from RGB and RGBD images: in particular, recognize which and where activities occur in the environment, using technologies of detection and tracking. 2) To forecast human activities, in order to predict what people are going to do and where they are going to move. Knowledge of scene is then used for light management. For example, switch off lights in areas which are not visible from the people currently acting within the scene. Activating/deactivating lights in relation to the predicted activities. This paper for the first time implements an invisible light switch: users have the feeling of all-lit, while their scene is only minimally lit, therefore providing a notable energy saving in the invisible.

Human activities can be characterized in several ways such as groups vs individual. As a proof of concept, we address activities that occur in indoor environments such as walking, working at the desk and discussing. Given our setup (top view camera), head orientation plays an important role as it identifies Visual Frustum of Attention (VFOA). VFOA approximates the volume of a scene where fixation of a person may occur. As head orientation captures attention [11], for example if I am looking at the a monitor most likely my activity has something to do with monitor. In this paper we demonstrate robust an dynamic modelling of VFOA for head orientation and scene understanding.

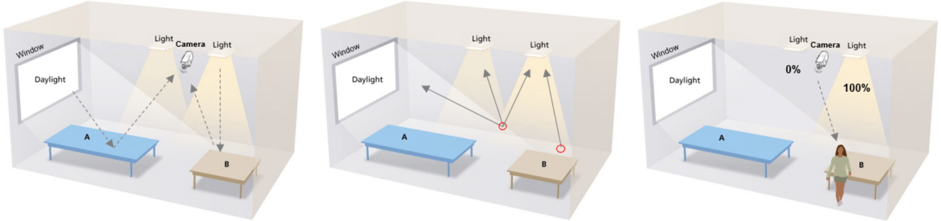

Fig. 1. Pipeline of an advanced lighting system. Left image, where the light sources are in camera view(calibration). Center image provides a visibility map(what they illuminate). Finally right image illustrates the adjustment of light sources based on people actions

2 State of the Art Review

Acquiring semantics from the scene is a fundamental requirement in several fields, ranging from computer vision to smart lighting. The project *SCENEUNDERLIGHT* is aimed at modelling the relationship between light and human behavior. We provide a review of some state of the art methods focused on the modelling of light and and behavior.

2.1 Lights and Behaviour

Relationship between human activities and lights is a widely studied topic in perceptual sciences [7, 9, 1]. Recently, it was illustrated by [25] that light intensifies people's perception. It triggers emotional system leading to intensified affective reactions. Light changes our perception of space [8], we 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 [5], as we human beings also rely on facial expressions which are only visible in light. Light provides sense of security [9], people adopted roads and streets in night due to the illumination [21]. Recently, studies targeting the office environments revealed a strong connection between people's productivity and the lights [19]. Eyeing the importance Eyeing the importance of lighting on humans, corresponding communities such as HCI [15] where interactive lighting deployed in city square provided a sense of "belongingness" to the residents. Furthermore ,ubiquitous computing [10] and architectural design [14] also have investigated this topic to an extent.

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. *SCENEUNDERLIGHT* for the first time models the relationship of light and human behavior via long term time-lapse observation of the scene by recognizing and forecasting activities in the scene. In this work, we propose the use of visual frustum of attention (VFOA) for scene understanding, activity recognition and activity forecasting. VFOA identifies the volume of a scene where fixation of a person may occur; It can be inferred from head pose estimation, and it is crucial in scenarios such as top-view office cameras and surveillance scenarios where precise gazing information cannot be retrieved.

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, [12, 24] used neural networks to estimate head pose. [4] adopted a randomized fern based approach to estimate head orientation. Only limited accuracy was achieved 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.Secondly, it was hard to compute low level image features in low resolution images. Recently, decision trees have been reported to achieve state of the art results [13]. 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 networks. in recent past, it has been used for pose estimation [22].

In this work, we plan to estimate VFOA with the help of an head pose estimator. We provide a review of the approaches that used VFOA in past in unconstrained scenarios.The earlier works that focus on estimating VFOA on low resolution images were [20] [20, 16] and [2], jointly with the pose of the person. VFOA has been used primarily for spotting social interactions: in [3] the head direction serves to infer a 3D visual frustum as approximation of the VFOA of a person. Given the VFOA and proximity information, interactions are estimated: the idea is that close-by people whose view frustum is intersecting are in some way interacting. The same idea has been explored, independently, in [17]. In [18], the

VFOA was defined as a vector pointing to the focus of attention, thanks to an approximate estimation of the gazing direction at a low resolution; in that work the goal was to analyze the gazing behavior of people in front of a shop window. The projection of the VFOA on the floor was modeled as a Gaussian distribution of “samples of attention” ahead of a pedestrian in [6]: the higher the density, the stronger the probability that in that area the eyes’ fixation would be present. More physiologically grounded was the modeling of [23]: in that work, the VFOA is characterized by a direction θ (which is the persons head orientation), an aperture $\alpha = 160$ *degree* and a length l. The latter parameter corresponds to the variance of the Gaussian distribution centered around the location of a person. Even in this case, samples of attention were used to measure the probability of a fixation: a denser sampling was carried at locations closer to the person, decreasing in density in zones further away. The frustum is generated by drawing samples from the above Gaussian kernel and keeping only those that fall within the cone determined by the angle α. In [26], the aperture of the cone can be modulated in order to mimic more or less focused attention areas.

In all these approaches, VFOA has been employed to capture group formations. To the best of our knowledge, we propose here, for the first time, VFOA for use in a predictive model. In order to estimate VFOA, a robust Head Pose Estimator is required which can work well in un-constrained real life scenarios. To this end we propose a robust real time head pose estimator using convolutional neural networks.The preliminary results are encouraging, as we have not done any preprocessing of the input image.

Finally, the project *SCENEUNDERLIGHT* proposes to model the relationship between behaviour and light, by providing an invisible light switch. Where the main essence is to provide user’s the feeling of ”all-lit” while the scene is minimally lit. A step towards new generation type of lighting system.

3 Proposed Framework

Towards the understanding of the scene, we distinguish the scene structure material properties and the human-centric scene. The first regards the scene composition: its 3D structure, the objects materials, the light position and characterization (natural versus artificial) and their lighting patterns. The second regards the human activities and interactions, particularly the human-scene (walking, working at desk or reading, presenting at a board) and human-human interaction (where people meet, discuss, relax). 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 the static 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. For this reason, it appears convenient to deal with the two topics within the same research framework, for the first time in the literature as illustrated in Fig. 2.

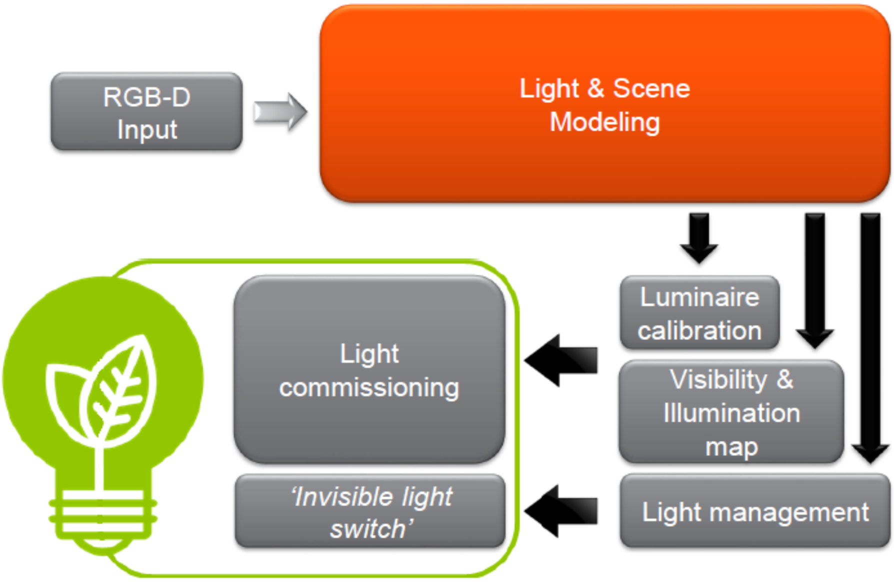

Fig. 2. Flowchart of the proposed system. An RGBD input image is used to model relationship between light and scene (human centric). The proposed system is capable of self calibration and finally implementing an invisible light switch, where the change in illumination is not perceived by the user.

3.1 Scene Composition Analysis

The structure of a scene consists of a number of material properties and their arrangement in the 3D space. This aspect is fundamental in order to understand the lighting propagation effects and the localization of the natural artificial sources. With the 3D scene structure, light propagation can be defined as an inverse problem called inverse lighting.

In this work, for the first time, inverse lighting is tackled in a real environment, typically indoor, that presents complex geometries and several types of lights (artificial and natural). Given the depth of the scene from a top-view RGBD sensor, the problem will result in the estimation of the photometric properties of the scene and objects material together with a coarse localization of the lights. In particular, we will rely on the fact that, given a large collection of images, inverse lighting becomes tractable. By leveraging a larger number of images depicting the same scene in time, it is possible to reduce the ambiguity of the problem by studying the evolving lighting conditions.

This information computed from the images will characterize the indoor scene by providing the estimation of the room light response, as a function of the different light sources and the different times during the day. This contributes to the final demonstrator with local scene estimates of how the current illumination differs from the one initially set, and how changing the lighting pattern can restore that.

3.2 Human Centric Scene Understanding

Further to the material properties, a scene is characterized by how humans interact with it. The research field of Ambient Intelligence (AmI) explicitly considers this, with the ultimate goal of designing transparent infrastructures, that actively adapts to the presence of people without prying into their lives . Similarly, lighting has to adapt to the specific activity of the human beings in the area in order to provide a light management system that follows the needs of the users. To this end, the project will design models and algorithms for estimating and forecasting the presence and the activity of humans, by exploiting the 3D+visual data (available from the RGBD images) and the inferred scene geometry.

As stated in previous sections, activities can be characterized into several categories such as individual vs group, etc. In this work we focus on activities that may occur in an office environment such as walking, discussing and working at the desk. Given the top view camera setup we propose the us a robust presence detector and dynamic modelling of human activities through VFOA. As VFOA identifies interest of people towards the scene, it can be used as a proxy for attention/gaze. Additionally this work for the first time proposes the use of head orientation in tracking. Given the fact that people usually walk in the direction they look at. Exploiting the information we can forecast future trajectories of humans in the scene. Activity analysis in a more robust fashion can be carried out using robust modelling of VFOA.

The goal is to discriminate among different human activities, intended as different trajectories and different elementary actions performed by the users (walking, writing at the PC, etc.). In this fashion, forecasting will be available, which will serve for implementing appropriate energy saving routines in the building (see next section). Once again, achieving such goals with RGBD data is a new challenge for the community: here, exploiting depth information could serve to ease the classification issue, that will be carried out using Social Affinity Maps.

3.3 The Invisible Light Switch

The idea behind the Invisible Light Switch is straightforward: the user controls and sets the illumination of the environment that he can see, while the proposed system acts on the part of the environment that the user cannot see, turning off the lights, thus ensuring a consistent energy saving. The study of the scene as discussed above serves this goal: knowing the 3D geometry of the scene and the map of inter-reflectance will allow to understand how the different light sources impact each point of the space; knowing where a user is located and what is his posture serves to infer what he can see and what he cannot, individuating potential areas where the light can be turned off. Being able to forecast his future activities will help understand (in advance) which lights should be turned on, avoiding the user to continuously act on the illumination system, and showing the user the illumination scenario that he wants to have.

4 Conclusion

The main aim of this research is to highlight the importance of smart lighting by implementing an *invisible light switch*. The key idea revolves around the fact knowledge of the static scene and light arrangement will allow the user to set a desired illumination pattern for the environment, which the system will maintain across daylight changes, e.g. augmenting the illumination level (given available light sources) when the sun sets. Secondly, detection, tracking and recognition of current and forecast human activities will allow an advanced occupancy detection, i.e. a control switch which turns lights on when the people are in the environment or about to enter it. Finally, this work joins research in smart lighting and computer vision towards the invisible light switch, which will bring both technologies together. The result light management system will be aware of the 3D geometry, light calibration, current and forecast activity maps. The user will be allowed to up an illumination pattern and move around in the environment (e.g. through office rooms or warehouse aisles). The system will maintain the lighting (given available light sources) for the user across the scene parts and across the daylight changes. Importantly, the system will turn lights off in areas not visible by the user, therefore providing energy saving in the invisible.

References

1. Adams, L., Zuckerman, D.: The effect of lighting conditions on personal space requirements. The journal of general psychology 118(4), 335–340 (1991)
2. Ba, S.O., Odobez, J.M.: A probabilistic framework for joint head tracking and pose estimation. In: IEEE International Conference on Pattern Recognition (ICPR) (2004)
3. Bazzani, L., Cristani, M., Tosato, D., Farenzena, M., Paggetti, G., Menegaz, G., Murino, V.: Social interactions by visual focus of attention in a three-dimensional environment. Expert Systems 30(2), 115–127 (2013)
4. Benfold, B., Reid, I.: Guiding visual surveillance by tracking human attention. In: British Machine Vision Conference (BMVC). pp. 1–11 (2009)
5. Carr, S.J., Dabbs Jr, J.M.: The effects of lighting, distance and intimacy of topic on verbal and visual behavior. Sociometry pp. 592–600 (1974)
6. Cristani, M., Bazzani, L., Paggetti, G., Fossati, A., Tosato, D., Del Bue, A., Menegaz, G., Murino, V.: Social interaction discovery by statistical analysis of f-formations. In: British Machine Vision Conference (BMVC). pp. 23.1–23.12 (2011)
7. Flynn, J.E., Hendrick, C., Spencer, T., Martyniuk, O.: A guide to methodology procedures for measuring subjective impressions in lighting. Journal of the Illuminating Engineering Society 8(2), 95–110 (1979)
8. Galasiu, A.D., Veitch, J.A.: 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)
9. Gifford, R.: Light, decor, arousal, comfort and communication. Journal of Environmental Psychology 8(3), 177–189 (1988)
10. Gil-Castineira, F., Costa-Montenegro, E., Gonzalez-Castano, F., López-Bravo, C., Ojala, T., Bose, R.: Experiences inside the ubiquitous oulu smart city. Computer 44(6), 48–55 (2011)
11. Goffman, E.: Behaviour in public places: notes on the social order of gatherings (1963)
12. Gourier, N., Maisonnasse, J., Hall, D., Crowley, J.L.: Head pose estimation on low resolution images. In: International Evaluation Workshop on Classification of Events, Activities and Relationships. pp. 270–280. Springer (2006)

13. Lee, D., Yang, M.H., Oh, S.: Fast and accurate head pose estimation via random projection forests. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1958–1966 (2015)
14. Magielse, R., Hengeveld, B.J., Frens, J.W.: Designing a light controller for a multiuser lighting environment (2013)
15. Poulsen, E.S., Morrison, A., Andersen, H.J., Jensen, O.B.: Responsive lighting: the city becomes alive. In: Proceedings of the 15th international conference on Humancomputer interaction with mobile devices and services. pp. 217–226. ACM (2013)
16. Robertson, N., Reid, I.: Estimating gaze direction from low-resolution faces in video. In: European Conference on Computer Vision (ECCV) (2006)
17. Robertson, N.M., Reid, I.D.: Automatic reasoning about causal events in surveillance video. EURASIP Journal on Image and Video Processing (2011)
18. Smith, K., Ba, S.O., Odobez, J.M., Gatica-Perez, D.: 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)
19. Smolders, K.C., de Kort, Y.A., Tenner, A.D., Kaiser, F.G.: Need for recovery in offices: Behavior-based assessment. Journal of Environmental Psychology 32(2), 126–134 (2012)
20. Stiefelhagen, R., Finke, M., Yang, J., Waibel, A.: From gaze to focus of attention. In: Visual Information and Information Systems (1999)
21. Taylor, L.H., Socov, E.W.: The movement of people toward lights. Journal of the Illuminating Engineering Society 3(3), 237–241 (1974)
22. Toshev, A., Szegedy, C.: Deeppose: Human pose estimation via deep neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1653–1660 (2014)
23. Vascon, S., Mequanint, E.Z., Cristani, M., Hung, H., Pelillo, M., Murino, V.: Detecting conversational groups in images and sequences: A robust game-theoretic approach. Computer Vision and Image Understanding 143, 11–24 (2016)
24. Voit, M., Nickel, K., Stiefelhagen, R.: A bayesian approach for multi-view head pose estimation. In: 2006 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems. pp. 31–34. IEEE (2006)
25. Xu, A.J., Labroo, A.: Incandescent affect: Turning on the hot emotional system with bright light. ACR North American Advances (2013)
26. Zhang, L., Hung, H.: 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)