

# LIT: a system and benchmark for light understanding

Theodore Tsesmelis<sup>1,2,3</sup>, Irtiza Hasan<sup>3,1</sup>, Marco Cristani<sup>2,3</sup>, Alessio Del Bue<sup>3,†</sup>, Fabio Galasso<sup>1,†</sup>  
Corporate Innovation OSRAM GmbH<sup>1</sup>, Istituto Italiano di Tecnologia (IIT)<sup>2</sup>, University of Verona (UNIVR)<sup>3</sup>

theodore.tsesmelis@iit.it, irtiza.hasan@univr.it

## Abstract

*A modern lighting system should automatically calibrate itself (light commissioning), assess its own status (which lights are on/off and how dimmed), and allow for the creation or preservation of lighting patterns (adjustability), e.g. after the sunset. Such a system does not exist today, nor (real) data, labels, or metrics are available to compare with and foster progress. In this paper we set the baselines to such a computational system, called LIT, and its applications. Using computational imaging we try to model and benchmark the light variations of indoor scenes with different illuminations (including natural light) and luminaire setups. We show that our lighting system can be easily trained with no manual intervention; after that, the benchmark allows to test automatic calibration (LIT-EST), status awareness (LIT-ID) and relighting (RE-LIT) as application.*

## 1. Introduction

A number of studies maintain that light is important for well-being and productivity of the occupants in buildings [17]. In particular, comfortable lighting conditions have a positive influence on vitality, depressive symptoms [15] and task performance [16]. In this scenario, the role of designing proper lighting solutions for indoor environments grew up in importance [20].

The design of lighting setups is usually done by rendering lighting patterns on CAD simulated scenes [12, 3]. Unfortunately, this does not consider future structural variations (different people and objects in the scene, the displacement of furniture may change) and that the natural light may drastically change the illumination (daytime, weather).

We introduce LIT, a novel computational framework and benchmark for complex indoor scenes to promote research on three necessary features of a smart lighting system. The first goal of the LIT system is to estimate the illumination layout, *i.e.* where the lights are positioned and what part of the scene they illuminate, which we call *Light Source Estimation* (LIT-EST). Next, we name *Light Source Identification* (LIT-ID) the understanding of the current lighting

pattern in the scene, *i.e.* which lights are on and to what intensity level. Finally, we propose as a third problem the *ReLighting* (RE-LIT), *i.e.* maintaining or setting a target illumination pattern.

The proposed solution to these problems builds upon factorization techniques. In particular, given the additive and non-negative property of light, we consider non-negative matrix factorization (NMF) [11] as a suitable decomposition technique for modelling the interaction of light sources in the scene. Concurrently we propose the LIT benchmark, the first applicable to a smart lighting system, comprising over  $\sim 50K$  images across the three defined tasks. We have acquired and manually labelled two time-lapse videos, in two scenes with natural light and different luminaire setups. Labels include the presence of a light source<sup>1</sup> and its dimming level.

In the rest of the paper, we define the overall LIT system (Sec. 3) and the benchmark (Sec. 4). Then we propose baseline LIT system solutions (Sec. 5) and evaluation (Sec. 6). Finally, we conclude (Sec. 7). Next we review related work.

## 2. Previous work

Light modeling and understanding are mostly studied in computer graphics for the creation of photo-realistic simulations. The light transport theory [21] physically simulates the path of transmitted light in a 3D environment and models the image as an integration process. This formalisation requires the 3D environment structure, the material surfaces and many other physical features [10].

The estimation in indoor environments is complicated by multiple light sources of different types (*e.g.* artificial, natural) and by the presence of objects of diverse sizes, shapes and spectral characteristics. Thus, most works assume known geometry or employ external sensors or tools, *e.g.* light probes, colorchecks. Reference spheres and a near light assumption are used in [19], achieving realistic results. However, without external sensor information, light estimation is a complex and an ill-posed problem. Lopez-Moreno

<sup>1</sup>We mean by *light sources* the natural light and the luminaires, providing light at a given intensity, but not a PC monitor for instance.

<sup>†</sup>These two authors contribute equally to the work.

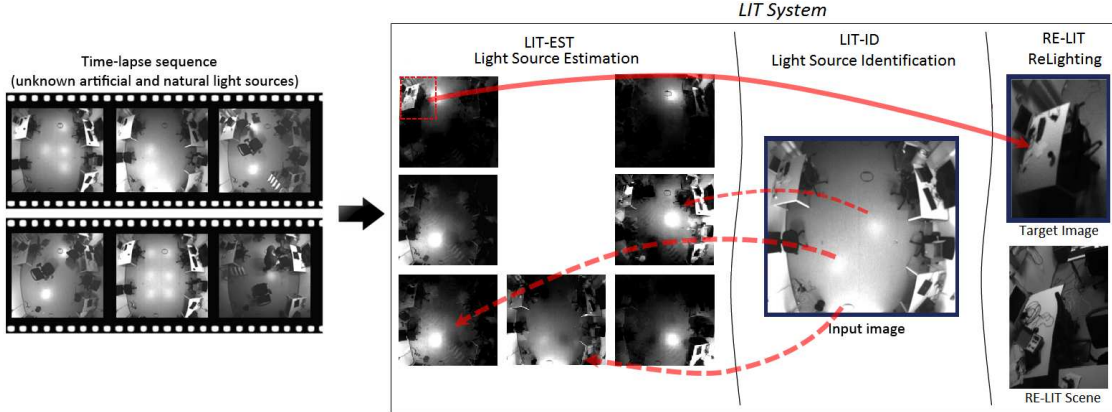


Figure 1: Overview of the LIT system. (Left) Time-lapse videos, depicting difference light sources active at different times. (Mid-left) LIT-EST estimates the single light sources of any type. (Mid-right) LIT-ID identifies which light sources are active and to what intensity. (Right) RE-LIT uses the sources to replicate the illumination pattern of a target image into a new scene.

*et al.* [14] and Lombardi *et al.* [13] make use of isolated reference objects (user input) to estimate multiple light sources from a single image.

Karsch *et al.* [7] introduced the idea of light estimation and correction through a rendering-based optimization procedure. They optimize for the light position and intensity by a cost function, encoding the pixel difference between the real image and the simulated one. As for [14], [7] also requires shape, position and color of light sources.

In recent works, Kasper *et al.* [8] tries to improve the computationally expensive technique of path-tracing for light estimation with the use of finite nonlinear parametrization while in [4] they introduce a learning based approach. Finally in [18] similarly to our work, light separation through a factorization technique is part of the computational imaging study on the electrical grid for bulb type and light phase estimation.

In this work we do not require any scene nor light calibration. Additionally, to the best of our knowledge, we are the first to address non-synthetic scenes of entire offices.

### 3. The LIT system

The LIT system uses a set of time-lapse images showing varying light conditions and changes in the indoor area structure due to objects displacement and human activity. These images are used to first extract an indoor lighting model and then perform identification of the single sources for active relighting of the indoor scene. The system consists of three parts, as shown in Fig. 1, which we overview in this section (detailed description in Section 5).

#### 3.1. Light sources estimation (LIT-EST)

LIT-EST models indoor lighting variations by estimating each light source independently from the other (the first processing block in Fig. 1), in an unsupervised manner (with-

out need for annotated data). More in details, we process time-lapse video sequences (**training set**) and extract images of single lighting modules (luminaires or the natural light), although possibly no frame showed them in isolation.

More formally, we consider an image  $I$  to be the result of an image formation process, completely defined by the set  $\mathcal{L}$  of  $L$  images of the individual light sources  $\mathcal{L} = \{S_k\}_{k=1}^L$  that are active for the image  $I$ . We assume that an unknown function exists, such that:

$$I = f_I(S_1, \dots, S_L), \quad (1)$$

where  $f_I$  is the function of the image formation process given by the light sources  $S_k$ .

Note that the above formalism and assumption is general and common to several lighting models as used in photometric stereo where each image  $S_k$  is a linear representation, for instance, of surface normals (surface normals and albedo [6], higher order spherical harmonics [1], reflectance and illumination components [13]). Also note that a simple but effective model for  $f_I$  is a weighted linear summation *i.e.*  $I = w_1 S_1 + \dots + w_n S_n$ , as we show in Section 5.

We name  $I_{EST}$  the distinct subset of the LIT dataset which is used for the LIT-EST light source estimation. Note that the LIT-EST process is unsupervised, but we collect the true light source appearances (ground truth) for benchmarking its quality.

#### 3.2. Light sources identification (LIT-ID)

LIT-ID identifies which light sources are active and to what intensity, *i.e.* the dimming level of a luminaire. More formally, LIT-ID estimates a light activation vector  $\mathbf{a}$  for each image  $I$  of the **testing set**:

$$\mathbf{a} = [a_1, \dots, a_L]^T, \quad a_k \in [0, 1] \quad (2)$$

where  $a_k$  indicates whether the light source  $S_k$  is switched off ( $a_k = 0$ ), lit to its maximum intensity ( $a_k = 1$ ), or

partially powered,  $0 < a_k < 1$ .

Likewise to LIT-EST, the LIT-ID process is also unsupervised. We collect though a distinct set of images  $I_{ID}$ , which we annotated with ground truth activation vectors for the sake of benchmarking.

### 3.3. ReLighting (RE-LIT)

RE-LIT stands for the problem of setting the illumination pattern of a scene according to a target image. Here we consider lighting a desk in the same way across different scenes (novel luminaires and furniture) and mimicking the natural light by only using the artificial lighting system.

We benchmark the latter quantitatively, by first estimating the light sources of the target scene with LIT-EST (from  $I_{EST}$ ), and then reconstructing the images of a distinct set  $I_{RE}$  that maintains the desired lighting pattern. In our experiments,  $I_{RE}$  is a subset of  $I_{ID}$  containing natural light or not according to the application. RE-LIT is benchmarked by directly comparing the target and re-lit scene images.

## 4. The LIT benchmark

We introduce the first benchmark for light estimation (LIT-EST), identification (LIT-ID) and relighting (RE-LIT). This comprises a dataset, a few subsets for ablation studies and specific metrics for each task.

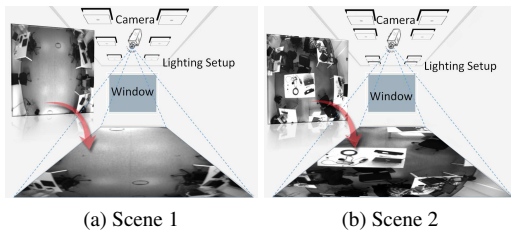


Figure 2: 3D models of the two indoor scenes used for the data acquisition. Images are captured in raw format, presenting the raw luminance that reaches to the sensor CMOS.

### 4.1. Time-lapse video dataset

We considered two scenes, as illustrated in Fig. 2, recorded time-lapse videos for an entire day using a fixed camera and labelled the active lights and their intensity for each frame. Both scene data is used for LIT-EST and LIT-ID. We only use Scene 2 data for RE-LIT because only luminaires in Scene 2 are dimmable, thus setttable to floating point activations, necessary for re-lighting.

Scene 1 data consists of 33,280 images (15,310 assigned to  $I_{EST}$  for light source estimation, 16,450 in  $I_{ID}$  for identification). Scene 2 data contains 20,700 images (9450 in  $I_{EST}$ , 11250 in  $I_{ID}$ ). A subset of 2600 images from  $I_{ID}$  are used for  $I_{RE}$ . The  $I_{EST}$  sets do not contain images of single sources (*i.e.* one luminaire active only), but we provide them separately for ablation studies (*c.f.* below).

Note that both  $I_{EST}$  and  $I_{ID}$  are ground-truth-annotated. Since we are targeting an unsupervised solution, we have used both entire sets (separately) for benchmarking the tasks. A train+test separation on each though is possible to allow learning-based techniques to be used.

### 4.2. Ablation data-subsets

We propose five subsets for ablation studies.

**LIT-Artificial.** This excludes all frames with natural light, which are supposedly more difficult for the processing. This contains 11,020  $I_{EST}$  images and 11,940  $I_{ID}$  frames for Scene 1. For Scene 2 the corresponding  $I_{EST}$  and  $I_{ID}$  frames are 16,500 and 18,000 respectively.

**LIT-Static.** This excludes all frames with motion of people, or objects. The Scene 1 *LIT-Static* comprises 7,240  $I_{EST}$  images and 7,490  $I_{ID}$  frames respectively.  $I_{EST}$  and  $I_{ID}$  sets for Scene 2 contain 8,900 and 9,700 respectively.

**LIT-Artificial-Static.** This is a subset of the *LIT-Artificial* where, having excluded frames featuring motion. This simplest scenario comprises 2,900  $I_{EST}$  and 3,400  $I_{ID}$  images for Scene 1, and 7,700  $I_{EST}$  and 8,500  $I_{ID}$  images for Scene 2.

**LIT-Sources.** This further adds to the full LIT the individual light sources (one light activated at a time), so as to test the decomposition techniques capability to disentangle lights. The *LIT-Sources* comprises 16,640  $I_{EST}$  and 16,450  $I_{ID}$  frames for Scene 1, 20,700  $I_{EST}$  images and 20,600  $I_{ID}$  images for Scene 2.

### 4.3. LIT benchmark metrics

We define metrics for each stage of the LIT system.

**LIT-EST measure.** This is given by the distance between the estimated light source images  $\tilde{S}_k$  for  $k = 1, \dots, L$  and their ground truth images  $S_k$ , which we define as the LIT-EST error measure  $\varepsilon_{EST}$ :

$$\varepsilon_{EST} = \frac{1}{L} \sum (1 - SSIM(S_k, \tilde{S}_k)), \quad (3)$$

where SSIM stands for the Structural Similarity Index [22] defined as:

$$SSIM(S_k, \tilde{S}_k) = \frac{(2\mu_{S_k}\mu_{\tilde{S}_k} + c_1)(2\sigma_{S_k}\sigma_{\tilde{S}_k} + c_2)}{(\mu_{S_k}^2 + \mu_{\tilde{S}_k}^2 + c_1)(\sigma_{S_k}^2 + \sigma_{\tilde{S}_k}^2 + c_2)} \quad (4)$$

with  $\mu, \sigma$  being the average and variance over the considered image, and  $c_{1,2}$  a small constant for the division stability. The SSIM index specifically tests how similar two images are in terms of luminance, contrast and structure (*c.f.* details in [22]).

Prior to compare true and estimated light source images in Eq. (3), we first match them according to their best SSIM score. In case of competing matches (*e.g.* two sources matching the same true source), we leave the worse unassigned and penalize the term with a 0 similarity.

**LIT-ID measure.** We compare the estimated  $\tilde{\mathbf{a}}$  and ground truth activation vector  $\mathbf{a}$  in terms of normalized  $l_1$ -norm distance:

$$\varepsilon_{\text{ID}} = \frac{1}{n} \sum_{i=1}^n \frac{\|\mathbf{a}_i - \tilde{\mathbf{a}}_i\|_1}{L} \quad (5)$$

where  $L$  is the number of light sources and  $n$  is the number of images in  $\mathcal{I}_{ID}$ .

**RE-LIT measure.** We measure the distance between a target image  $\mathbf{I}_{RE}$  (e.g. an image containing the natural light) and the re-lit image estimate  $\tilde{\mathbf{I}}_{RE}$  (e.g. generated with the sole artificial lights, estimated from  $\mathcal{I}_{EST}$ ), in terms of SSIM such that:

$$\varepsilon_{\text{RE}} = \frac{1}{L} \sum (1 - \text{SSIM}(\tilde{\mathbf{I}}_{RE}, \mathbf{I}_{RE})) \quad (6)$$

## 5. LIT baselines

In the following we propose baselines methods for LIT-EST, LIT-ID and RE-LIT.

---

### Algorithm 1: LIT-EST algorithm.

---

- 1 Training image sequence  $\mathcal{T} = \{\mathbf{I}_1, \dots, \mathbf{I}_n\}$ ;
  - 2 Create matrix  $\mathbf{V} \in \mathbb{R}^{m \times n}$  such that:  

$$\mathbf{V} = [ \text{vec}(\mathbf{I}_1) \quad \dots \quad \text{vec}(\mathbf{I}_n) ]$$
;
  - 3 Estimate  $[\mathbf{W}, \mathbf{H}]$  minimizing  $\min_{\mathbf{W}, \mathbf{H} \geq 0} \|\mathbf{V} - \mathbf{W}\mathbf{H}\|^2$  where  

$$\mathbf{W} \in \mathbb{R}^{m \times L}, \mathbf{H} \in \mathbb{R}^{L \times n} \text{ and } 1 \leq L \leq \min(m, n)$$
;
  - 4 **for**  $i = 1 : L$  **do**
  - 5     For each basis  $i$ , get the reconstructed images:  

$$\hat{\mathbf{V}}_i = \mathbf{w}_i \mathbf{h}_i^\top$$
;
  - 6     Compute the activation as  $\text{AoI}_{(i,j)} = \|\hat{\mathbf{V}}_{(i,j)}\|$ ;
  - 7     Find the maximum activation  

$$\text{idx}_{max} = \arg \max(\text{AoI}_{(i)})$$
;
  - 8     Assign  $\tilde{\mathbf{S}}_i = \hat{\mathbf{V}}_{(i, \text{idx}_{max})}$ ;
  - 9  $\tilde{\mathcal{L}} = \{\tilde{\mathbf{S}}_1, \dots, \tilde{\mathbf{S}}_L\}$ ;
- 

### 5.1. LIT-EST modeling

Given a set of  $n$   $\mathcal{I}_{EST}$  images extracted from a time-lapse video  $\mathcal{T} = \{\mathbf{I}_{EST1}, \dots, \mathbf{I}_{ESTn}\}$ , the goal of the LIT-EST application is to find, in an unsupervised way, the set  $\mathcal{L}$  of  $L$  source images that explain the illumination of the scene (cf. Eq. 1 of Sec. 3.1). This resembles the classical source separation problem, commonly approached with PCA, ICA and the like. For selecting the best technique of this family, we should note two main physical properties of light: **(a)** light from general illumination sources is incoherent and the light intensity from multiple sources combines linearly; **(b)** each light source contributes to the image formation with a non-negative weight term. This translates into an image formation function where non-negative light bases are combined with non-negative weights. Such conditions

elects Non-negative Matrix Factorization (NMF) [11] as the natural choice for light decomposition.

**NMF problem statement.** Given an  $m \times n$  non-negative matrix  $\mathbf{V}$ , NMF finds two compact non-negative matrices  $\mathbf{W}_{m \times r}$  (containing  $r$  horizontally stacked  $m$ -dimensional bases), and  $\mathbf{H}_{r \times n}$  (containing for each image the coefficients associated to the basis), such that:

$$\min_{\mathbf{W}, \mathbf{H}} \|\mathbf{V} - \mathbf{W}\mathbf{H}\|_F, \text{ subject to } \mathbf{W} \geq 0 \text{ and } \mathbf{H} \geq 0. \quad (7)$$

In our case the positive data matrix  $\mathbf{V}$  is given by the images in the video sequence by vectorizing them ( $m$  being the number of pixels) and concatenating all  $n$  training dataset:

$$\mathbf{V} = [ \text{vec}(\mathbf{I}_1) \quad \dots \quad \text{vec}(\mathbf{I}_n) ] = [ \mathbf{v}_1 \quad \dots \quad \mathbf{v}_n ],$$

where  $\text{vec}(\bullet)$  is the (column-wise) vectorizing operator. The positive matrix  $\mathbf{W}$  contains the  $r$  light bases as a set of images containing each light sources, while the matrix  $\mathbf{H}$  holds the linear positive weights that sum each basis in order to obtain  $\mathbf{V}$ .

**NMF techniques.** Motivated by their efficiency and robustness to noise, we focus on three recent NMF techniques:

- **Diagonalized Newton Algorithm for Non-negative Matrix Factorization (DNA-NMF) [5].** The DNA-NMF shows a faster convergence compared to NMF thanks to the diagonalized Newton optimization. The estimation with DNA-NMF remains tractable even for high-rank problems.
- **Block Principal Pivoting Non-negative Matrix Factorization (BPP-NMF) [9].** BPP-NMF improves further the computational burden by allowing the exchanges of multiple variables between working sets. This method also allows a constrained solution by inducing a sparsity prior with the use of  $L_1$ -norm.
- **Non-negative Matrix Under-approximation with Priors (PNMU) [2].** PNMU extends standard NMF by solving it sequentially and incorporating spatial and sparsity priors. This biases the algorithm towards bases that have active pixels (those illuminated by the light source) spatially close, and in general fewer active pixels per basis.

Further to the NMF approaches, we experiment with PCA and ICA decompositions. The comparative evaluation should highlight the importance of the non-negativity constraints in the LIT optimization.

**Light source image reconstruction.** The light decomposition bases  $\mathbf{W}$  computed from Eq.(7) via NMF techniques are important for the light source estimation quality. The bases should in fact disentangle the light sources, such that each basis would only contribute to a single light source.

Matrices  $W$  are however not yet light source images because their values need to be weighted by the weight matrix  $H$ . For example, small bases values may still generate strong light sources if their weight is large. This means that we should find a reference image for each basis to drive the light source estimation, *i.e.* the weighted basis.

Since the set of images  $I_{EST}$  may contain light sources partially lit, it is important to identify the one image where each weighted basis is strongest (across all pixels), therefore corresponding to one where the light source is fully lit.

In more details, let  $w_i$  be the basis vector in  $W$  that contributes to light source  $S_i$  and let  $h_j$  be the related coefficients. The reconstructed image is given by  $\hat{V}_{(i,j)} = w_i h_j^\top$ . The Activation of Image (AoI) pixel map is given by evaluating, for each image  $j$  in the training set, the following:  $AoI_{(i,j)} = \|\hat{V}_{(i,j)}\|$ . The maximum activation would then be given by:

$$\bar{j} = \arg \max_j AoI_{(i,j)}, \quad (8)$$

where  $\tilde{S}_i = \hat{V}_{(i,\bar{j})}$  is the desired light source image.

## 5.2. LIT-ID modeling

The  $L$  light source images estimated in the previous phase are then used to identify which light sources are active for each test image (LIT-ID). As defined in Sec. 3.2 this means estimating the activation vector  $\mathbf{a}_{ID}$  indicating the intensity of the  $L$  light sources.

We formalise this step as a constrained Least Squares problem since  $\mathbf{a}$  is bounded between 0 and 1. First, given all the light source images, we form the  $n \times L$  matrix  $\Lambda$  as:

$$\Lambda = [ \text{vec}(\tilde{S}_1) \quad \dots \quad \text{vec}(\tilde{S}_L) ].$$

Given a test image  $I_{ID}$ , the estimated activations  $\mathbf{a}_{ID}$  are computed as the solution of:

$$\min_{\mathbf{a}} \frac{1}{2} \|\Lambda \cdot \mathbf{a}_{ID} - \text{vec}(I_{ID})\|_2^2, \quad \text{subject to } 0 \leq \mathbf{a} \leq 1.$$

The constrained optimization ensures meaningful non-negative activation of light sources.

## 5.3. RE-LIT modeling

As for the LIT-ID task, we assume here to have previously computed the  $L$  light source images with LIT-EST. However, differently from the LIT-ID, the target image (the one providing the lighting pattern to replicate) may in general have been acquired within a different reference system, *e.g.* different luminaire and furniture layout, or within the same scene under different conditions, *e.g.* the target image may contain natural light but require its replication after the sunset with luminaires only.

Let us denote  $\Lambda$  the  $n \times L$  matrix, containing the vectorized  $L$  light sources. We further denote  $P$  the target image, providing the desired lighting pattern. The ReLighting activations  $\mathbf{a}_{RE}$  are estimated by solving the constrained optimization problem:

$$\min_{\mathbf{a}_{RE}} \frac{1}{2} \|\Lambda \cdot \mathbf{a}_{RE} - \text{vec}(P)\|_2^2, \quad \text{subject to } 0 \leq \mathbf{a}_{RE} \leq 1.$$

The  $\mathbf{a}_{RE}$  activations may be employed, as we do here, to synthesize the relit image:  $I_{RE} = \Lambda \mathbf{a}_{RE}$ . The same  $\mathbf{a}_{RE}$  vector may be input into the lighting system as luminaire dimming coefficients, for lighting the room as desired.

## 6. Benchmark Evaluation

Here we analyze the methods introduced in Sec. 5 on the tasks of LIT-EST, LIT-ID and RE-LIT.

### 6.1. Light source estimation results (LIT-EST)

Table 1 reports the light source estimation errors  $\varepsilon_{EST}$ , measuring the quality of the computed light source set  $\mathcal{L}$  with the source images of Scene 1 visualized in Figure 3.

In the first row of the table (Scene 1, LIT-All), one can immediately notice the inferiority of PCA and ICA, which confirms the importance of including non-negativity constraints in the modeling. Looking at the NMF techniques, BPP-NMF emerges as the best technique with an image reconstruction error of 31.2%. PNMU follows it tightly at the distance of  $\sim 3\%$ , while DNA-NMF under-performs by  $\sim 13\%$ . This is confirmed by the number of sources

$\varepsilon_{EST}$	Scene 1							Scene 2
	PCA	ICA	PNMU	PNMU sparse	DNA-NMF	BPP-NMF	BPP-NMF sparse	BPP-NMF sparse
<b>LIT-All</b>	0.6526 (4/7)	0.6063 (5/7)	0.3457 (6/7)	0.3887 (6/7)	0.4468 (5/7)	0.3119 (6/7)	<b>0.2877</b> (6/7)	0.5872 (6/9)
<b>LIT Artificial</b>	0.5856 (4/6)	0.5912 (5/6)	0.4022 (5/6)	0.3981 (5/6)	0.5058 (4/6)	0.3298 (5/6)	<b>0.3147</b> (5/6)	0.5584 (5/8)
<b>LIT Static</b>	0.7382 (4/7)	0.6406 (5/7)	0.3315 (6/7)	0.5750 (4/7)	0.3121 (6/7)	0.3055 (6/7)	<b>0.2409</b> (6/7)	0.2928 (7/9)
<b>LIT Artificial-Static</b>	0.5867 (4/6)	0.7607 (3/6)	0.3188 (5/6)	0.3514 (5/6)	0.0578 (6/6)	<b>0.0267</b> (6/6)	0.0289 (6/6)	0.2775 (7/8)
<b>LIT Sources</b>	0.6018 (5/7)	0.5070 (6/7)	0.3500 (6/7)	0.3964 (6/7)	0.2912 (6/7)	0.2963 (6/7)	<b>0.2653</b> (6/7)	0.5333 (7/9)

Table 1: Comparative evaluation of light source reconstruction techniques over the LIT-All set and subsets.  $\varepsilon_{EST} \in [0, 1]$  is the error measures (lower is better). Number in parentheses indicate the successfully matched light sources, see Section 6.1.

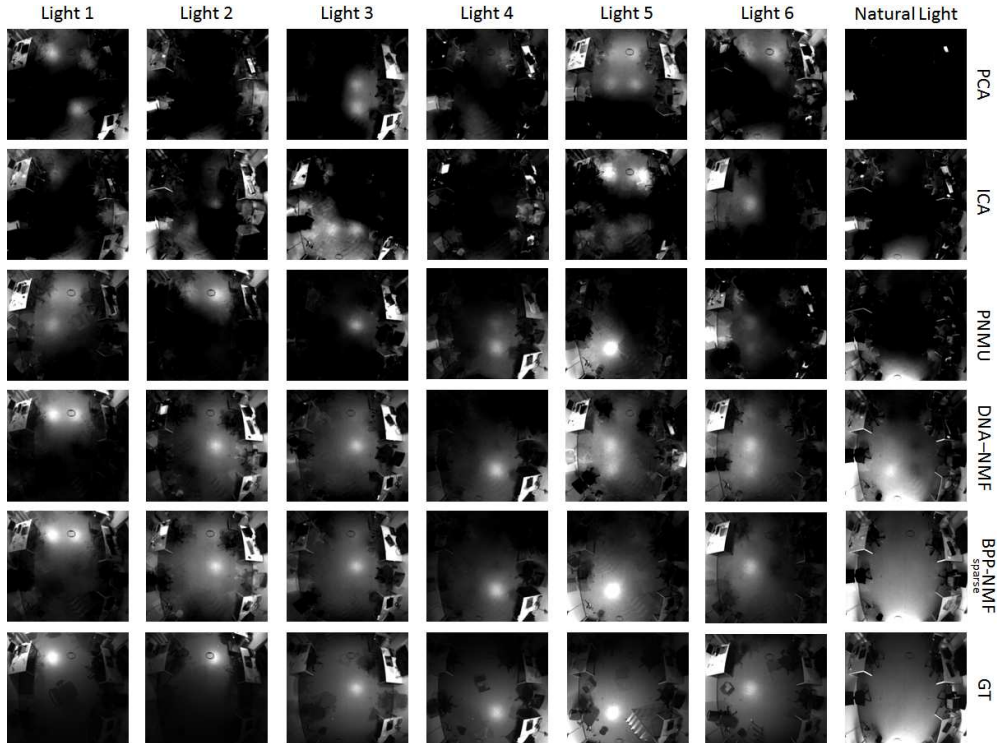


Figure 3: Light source estimates for each selected technique on the LIT-All set and subsets of Scene 1. PNMU and BPP-NMF techniques are shown only by the best performing alternatives, see Tables 1, 2. Column-wise, correct light source images should look alike to the ground truth, GT (last row).

which the technique manages to reconstruct and match to the ground truth, 6 out of 7 for all but the DNA-NMF (5 out of 7).

Notably, sparsity may improve results, but only if carefully designed. As an example, the BPP-NMF-sparse improves by  $\sim 3\%$  on BPP-NMF, which results in the best reconstructions. By contrast, PNMU-sparse degrades PNMU by  $\sim 4\%$ . It is plausible that this is influenced by the type of sparsity,  $l_1$ -norm for the BPP-NMF-sparse (seemingly preferable),  $l_2$ -norm for the PNMU-sparse.

**Ablation studies.** First, we consider the influence of the natural light by removing it from the LIT-All benchmark (LIT-Artificial). The rank of the techniques does not

change, confirming our earlier conclusions, but the decrease in performance of all techniques *wrt* LIT-All (e.g. BPP-NMF from 28.9% to 31.4% error) is somewhat surprising. Intuitively, the natural light is more difficult to model, as coming from a larger source (the sun, through a large window) and changing in the course the day. We question the puzzling performance decrease by visual inspection of the data and we realize that most scene motion (people, chairs, objects) takes place after the sunset, apparently a more active working time in the room. We conclude from the LIT-Artificial that modelling the scene changes is a harder task for matrix factorization techniques, compared to the natural light variability.

$\varepsilon_{ID}$	Scene 1							Scene 2		
	PCA	ICA	PNMU	PNMU sparse	DNA-NMF	BPP-NMF	BPP-NMF sparse	Oracle	BPP-NMF sparse	Oracle
LIT-All	0.3647	0.3470	0.2727	0.3149	0.3153	0.2537	<b>0.2405</b>	0.0756	0.2346	0.1056
LIT Artificial	0.3754	0.3722	0.3523	0.3402	0.3621	0.2773	<b>0.2662</b>	0.0627	0.2527	0.0829
LIT Static	0.3969	0.3653	0.2807	0.3893	0.2526	0.2469	<b>0.2165</b>	0.0652	0.1503	0.0413
LIT Artificial-Static	0.3527	0.4110	0.2933	0.2531	0.1164	0.0920	<b>0.0375</b>	0.0271	0.1574	0.0316
LIT Sources	0.3811	0.32224	0.2921	0.3306	0.2721	0.2500	<b>0.2311</b>	0.0756	0.2318	0.1056

Table 2: Comparative evaluation of selected techniques in terms of light activation error  $\varepsilon_{ID}$  over the LIT-All set and subsets.  $\varepsilon_{ID} \in [0, 1]$  (lower is better), see Section 6.2.

This motivates the second ablation study (third row), whereby we remove any motion from the scene (LIT-Static). BPP-NMF-sparse remains the best with 24.1% error and improves *wrt* LIT-All by  $\sim 5\%$ . Indeed motion is a difficult element to model and there is much motion in our LIT, which challenges the selected techniques. Somewhat surprisingly, the worse performers (PCA, ICA, PNMU-sparse) get worse, confirming the importance of non-negativity and of a correct sparsity prior.

Finally, in LIT-Artificial-Static we remove both motion and the natural light from LIT, which results in most techniques “solving” the decomposition correctly (*e.g.* a residual 3% error for BPP-NMF). This resonates with our earlier observations.

While in LIT-All the techniques should reconstruct the light sources without ever observing them in isolation (each frame features at least two lights switched on), LIT-Sources test whether this would make a difference. The results match the intuition, though with marginal gain (*e.g.* BPP-NMF improving by 2%).

**Scene 2.** The same experiment, repeated on Scene 2, allows for the same conclusions as the errors of BPP-NMF-sparse, also here perform best (see Table 1 last column). Note the general higher error on LIT-All (58.7%), now larger than LIT-Artificial (by visual inspection, Scene 2 contained more motion in the day time). In fact, Scene 2 is more difficult “by design”, with 2 additional luminaires, further away from the camera view. This is witnessed by the non-trivial error on LIT-Artificial-Static (27.7%)<sup>2</sup>.

## 6.2. Light identification results (LIT-ID)

The light source activation errors  $\epsilon_{ID}$  for the selected techniques, applied on Scene 1, are reported in Table 2. The scores question the capability to identify the lighting setups in all LIT-ID frames, given the reconstructed light sources.

Table 2 brings to similar conclusions as from Table 1, with BPP-NMF-sparse being the best performer (24% error, LIT-All row), both due to the superiority of the BPP-NMF technique (*cf.* the higher errors of PNMU, DNA-NMF, PCA, ICA) and the type of adopted  $l_1$ -norm sparsity (*cf.* the slightly worse performance of BPP-NMF). Again the LIT-Artificial appears to be more challenging for all techniques due to the after-sunset motion (*cf.* second row); LIT-Static is in fact an easier task (BPP-NMF reducing its error to 21.6%); LIT-Static-Artificial just leaves most techniques with a residual error (3.7% for BPP-NMF).

Since the activation errors depend on the estimated light sources, we further compare the techniques to the Oracle (final column), which “knows” the true light sources. We notice though that the Oracle error on LIT-All is 7.6%, meaning that, even after a perfect light calibration, the light scene understanding still needs further research.

<sup>2</sup>Complete results for Scene 2 are reported in *supp. material*.

Second, the natural light emerges as the largest challenge (the Oracle error on LIT-Artificial drops to 6.2%). While the motion troubles the light source estimation the most (*cf.* 6.1), the understanding of the scene illumination is most complex due to the natural light because it comes from a from the larger window area, it traverses the entire scene and it crosses all other artificial lights.

**Scene 2.** Testing on Scene 2 (Table 2, last two columns) brings to similar conclusions. As discussed in Sec. 6.1, Scene 2 is more difficult, which is witnessed by the larger error in the LIT-Artificial-Static (15.7%).

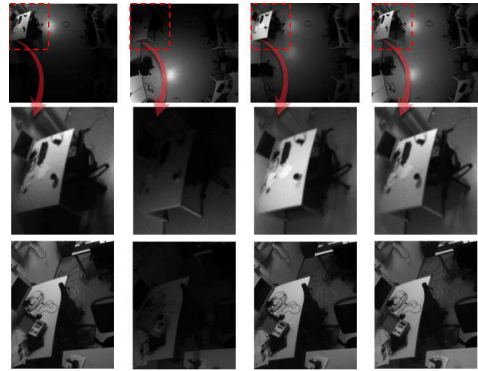


Figure 4: Relighting sample. Scene 1 lighting scenario (top row, zoom in middle row) is re-created in Scene 2 (bottom row), with different luminaires.

## 6.3. Relighting results (RE-LIT)

We propose two sample applications for the introduced system: a) maintaining a lighting pattern when the illumination setup changes and b) compensating the lack of natural light, by synthesizing its effect with the installed luminaires.

Towards the first application, in figure 4 we show qualitatively how the illumination level on a desk in Scene 1 (first two rows) can be replicated onto a corresponding desk in Scene 2 (third row). The gradient and overall illumination level should be as similar as possible along the columns (for each light scenario).

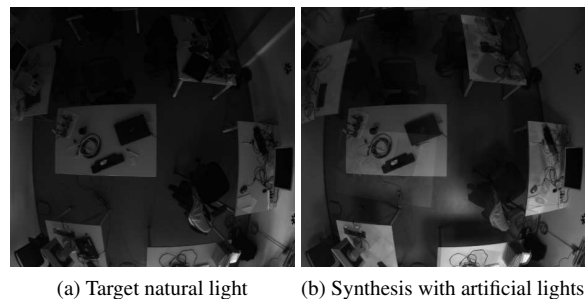


Figure 5: Reconstruction of natural light sample by the artificial sources, estimated with BPP-NMF for Scene 2.

For the second application task, the natural light synthesis, we consider Scene 2, because it offers dimmable luminaires (where their light intensity can be adjusted). As described in Section 3.3, the artificial light sources (estimated from the LIT-Artificial set) are combined to compensate for the missing natural light. Figure 5 illustrates one such example, whereby the sole natural source is synthesized.

We measure the average relighting error  $\varepsilon_{RE}$  over all  $I_{ID}$  images (those containing natural lights). BPP-NMF-sparse achieves a **14%** error, while the oracle (*i.e.* the true artificial light sources) gives an error of **13%**. The sole 1% gain of the oracle indicates that the light source estimation of BPP-NMF-sparse is not the performance bottleneck. Rather most error is due to the complex reflection patterns which the natural light casts in the scene.

## 7. Conclusion

We have proposed LIT to stem computer vision research in the field of smart lighting. LIT addresses three key features of an advanced lighting system: 1) the automatic estimation of the light sources and their influence in the scene; 2) the approximation of the active light intensity; 3) and the capability to apply or preserve a desired lighting pattern.

Thereafter, we have introduced the LIT benchmark: data, labels and metrics to measure and foster progress in this new field. Our experiments highlight non-negative matrix factorization techniques as a promising modelling direction, in line with the light's physical nature.

In the future, we foresee more computer vision research in the field, including people detection and action recognition, towards an alternative way we are living lighting.

**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] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *International Journal of Computer Vision*, 72(3):239–257, 2007. **2**
- [2] G. Casalino and N. Gillis. Sequential dimensionality reduction for extracting localized features. *CoRR*, 2015. **4**
- [3] C. Dachsbacher, J. Krivánek, M. Hašan, A. Arbre, B. Walter, and J. Novák. Scalable realistic rendering with many-light methods. In *Computer Graphics Forum*, volume 33, pages 88–104. Wiley Online Library, 2014. **1**
- [4] M. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gamberetto, C. Gagné, and J. Lalonde. Learning to predict indoor illumination from a single image. 2017. **2**
- [5] H. V. hamme. The diagonalized newton algorithm for non-negative matrix factorization. *CoRR*, 2013. **4**
- [6] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *JOSA A*, 11(11):3079–3089, 1994. **2**
- [7] K. Karsch, V. Hedau, D. Forsyth, and D. Hoiem. Rendering synthetic objects into legacy photographs. *ACM Trans. Graph.*, 30(6):157:1–157:12, Dec. 2011. **2**
- [8] M. Kasper, N. Keivan, G. Sibley, and C. Heckman. Light source estimation with analytical path-tracing. *CoRR*, abs/1701.04101, 2017. **2**
- [9] J. Kim, Y. He, and H. Park. Algorithms for nonnegative matrix and tensor factorizations: a unified view based on block coordinate descent framework. *Journal of Global Optimization*, 58(2):285–319, 2014. **4**
- [10] J. Krivánek, A. Keller, I. Georgiev, A. S. Kaplanyan, M. Fajardo, M. Meyer, J.-D. Nahmias, O. Karlık, and J. Cañada. Recent advances in light transport simulation: Some theory and a lot of practice. In *ACM SIGGRAPH 2014 Courses*, SIGGRAPH '14, pages 17:1–17:6, NY, USA, 2014. ACM. **1**
- [11] D. D. Lee and H. S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999. **1, 4**
- [12] W.-C. Lin, T.-S. Huang, T.-C. Ho, Y.-T. Chen, and J.-H. Chuang. Interactive lighting design with hierarchical light representation. In *Computer Graphics Forum*, volume 32, pages 133–142. Wiley Online Library, 2013. **1**
- [13] S. Lombardi and K. Nishino. Reflectance and illumination recovery in the wild. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):129–141, 2016. **2**
- [14] J. Lopez-Moreno, S. Hadap, E. Reinhard, and D. Gutierrez. Compositing images through light source detection. *Computers & Graphics*, 34(6):698 – 707, 2010. Graphics for Serious Games Computer Graphics in Spain: a Selection of Papers from {CEIG} 2009 Selected Papers from the {SIGGRAPH} Asia Education Program. **2**
- [15] T. Partonen and J. Lönnqvist. Bright light improves vitality and alleviates distress in healthy people. *Journal of Affective disorders*, 57(1):55–61, 2000. **1**
- [16] 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*, 26(6):695–700, 2003. **1**
- [17] D. Serghides, C. Chatziniakola, and M. Katafygiotou. Comparative studies of the occupants behaviour in a university building during winter and summer time. *International Journal of Sustainable Energy*, 34(8):528–551, 2015. **1**
- [18] M. Sheinin, Y. Schechner, and K. N. Kutulakos. Computational Imaging on the Electric Grid. *IEEE CVPR*, 2017. **2**
- [19] T. Takai, A. Maki, K. Niinuma, and T. Matsuyama. Difference sphere: An approach to near light source estimation. *Comput. Vis. Image Underst.*, Sept. 2009. **1**
- [20] P. Tregenza and M. Wilson. *Daylighting: architecture and lighting design*. Routledge, 2013. **1**
- [21] E. Veach and L. J. Guibas. Optimally combining sampling techniques for monte carlo rendering. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 419–428. ACM, 1995. **1**
- [22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. **3**