# Point Light Source Position Estimation from RGB-D Images by Learning Surface Attributes

Sezer Karaoglu, Yang Liu, Theo Gevers, *Member, IEEE*, and Arnold W. M. Smeulders, *Member, IEEE*,

***Abstract*—Light source position estimation is a difficult yet an important problem in computer vision. A common approach for estimating the light source position (LSP) assumes Lambert's law. However, in real-world scenes, Lambert's law does not hold for all different types of surfaces. Instead of assuming all that surfaces follow Lambert's law, our approach classifies image surface segments based on their photometric and geometric surface attributes (i.e. glossy, matte, curved etc.) and assigns weights to image surface segments based on their suitability for LSP estimation. In addition, we propose the use of the estimated camera pose to globally constrain LSP for RGB-D video sequences.**

**Experiments on *Boom* and a newly collected RGB-D video datasets show that the state-of-the-art methods are outperformed by the proposed method. The results demonstrate that weighting image surface segments based on their attributes outperforms the state-of-the-art methods in which the image surface segments are considered to equally contribute. In particular, by using the proposed surface weighting, the angular error for light source position estimation is reduced from $12.6°$ to $8.2°$ and $24.6°$ to $4.8°$ for *Boom* and RGB-D video datasets respectively. Moreover, using the camera pose to globally constrain LSP provides higher accuracy $(4.8°)$ compared to using single frames $(8.5°)$.**

## I. Introduction

Images are the result of complex interactions between the light source, objects and recording devices. Being the creator of the image before anything else, the light source is often ignored as an important cue to its understanding. The light source may reveal low-level (e.g. surface structure [1], [2], [3], [4]) and high-level (e.g. material property [5]) information. Such information is used by humans in their daily activities. For instance, we can distinguish whether a surface is matte/glossy and what material the surface is made of (e.g. a matte plastic or a shiny metal) [6]. Or we can interpret the underlying geometry of objects (due to shading cues). The interaction between the light source and objects is also used in visual art where artists exploit the characteristics of light in different ways. Specular reflections in an human eye give the lively twinkle but are in fact a direct reflection of the light source. Shading indicates the curvature of the body and reveals collimation and direction of the light. Spotlight steers salience, and backlight in art photography renders the subject radiant. We consider that light source is an inevitable factor of forming and understanding images. Therefore, in this paper, we focus on detecting the light source position.

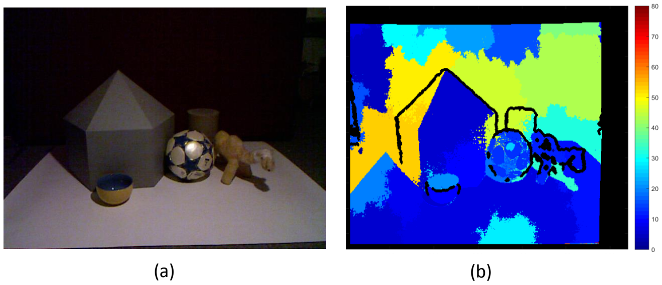


Fig. 1. (Best viewed in color) (a) An image sample taken from [7]. (b) Image surface segments and their individual angular errors. The color represents the angular error of each image surface segment. The corresponding angular error for each color is at the colorbar next to each plot. Dark blue regions represent smaller angular errors whereas dark red regions represent larger angular errors. The figure illustrates that curved (such as ball and bowl) and less textured (such as table and toy castle) image surface segments have lower LSP estimation errors. Moreover, regions with shadows or highlights have difficulties to estimate LSP.

Light Source Position (LSP) estimation has caught attention in references such as [8], [9], [10], [11], [12], [13], [14]. Most of these algorithms are based on Lambert's law, assuming that the pixel intensity is proportional to the angle between the light and surface (normal) direction. LSP estimation algorithms infer the position by assuming certain 3D-shapes of objects in the scene [8] (to obtain surface normals). Often these assumptions fail, and hence the applicability of these methods is limited.

A recent approach is to use low cost RGB-D cameras (e.g. Kinect and Asus Xtion) as they acquire color images with their depth in real-time. The use of RGB-D cameras alleviates the requirement of assuming certain object shapes, because the surface normals can be readily computed [7], [15], [16], [14], [17]. Assuming Lambert's law, the light source position can be estimated using the surface normals and pixel intensities. In particular, LSP is obtained by minimizing the residuals between the re-rendered (scene generated using a hypothesized light source position) and the original scene. This is straight-forward as long as Lambert's law holds, such as for matte surfaces. However, a glossy surface is prone to specular highlights and material-to-material inter-reflections which are not considered by Lambert's law. Moreover, due to imperfections of recording devices, the surface normals may be noisy on rough, crinkled and grained surfaces. Hence, these

S. Karaoglu is with the Computer Vision Group, University of Amsterdam, The Netherlands (e-mail: s.karaoglu@uva.nl).

Y. Liu is with the Computer Vision Group, University of Amsterdam, The Netherlands (e-mail: lawyoung529@gmail.com).

T. Gevers is with the Computer Vision Group, University of Amsterdam, The Netherlands, and also with the Computer Vision Center, Universitat Autònoma de Barcelona, 08193 Barcelona, Spain (e-mail: th.gevers@uva.nl).

A. W. M. Smeulders is with the Intelligent Sensory Information Systems Lab, Amsterdam, University of Amsterdam, 1098 XH Amsterdam, The Netherlands (e-mail: a.w.m.smeulders@uva.nl).

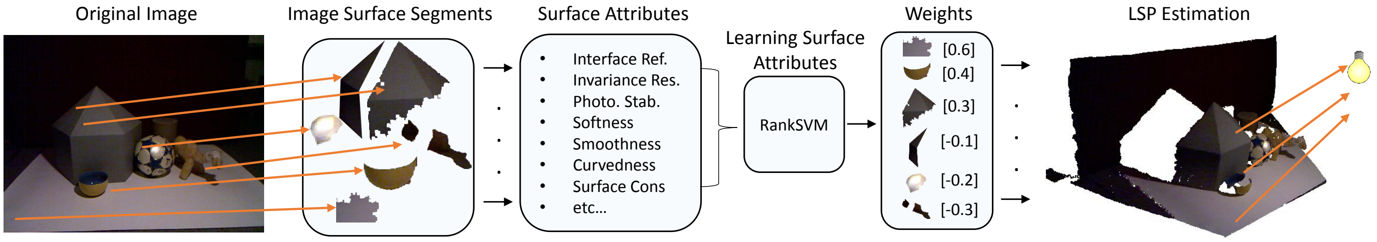


Fig. 2. The flow of the proposed approach. First, the image is divided into image surface segments. Surface attributes (e.g. glossy, matte, highlight, curved etc) are extracted from these regions aiming to rank them based on their suitability for proper LSP estimation. The proposed method assigns more importance to image surface segments which are more suitable for LSP estimation. This is in contrast to the state-of-the-art methods [7], [15] which assume equal importance to all image surface segments. Moreover, to improve the performance further, we introduce temporal constraints (for video sequences).

type of surfaces may negatively influence the LSP estimation. In Fig. 1, we estimate the LSP for each image surface segment (obtained by [18]). We measure the angular error between the estimated LSP and ground-truth LSP. The figure illustrates that different image surface segments have varying LSP estimation errors.

In this paper, we propose a method which exploits the influence of various surface attributes for LSP estimation. First, surface attributes are computed from image surface segments. These attributes are used in a supervised learning scheme to rank the suitability of each surface for proper LSP estimation. Higher importance is assigned to image surface segments which have proper photometric (i.e. Lambertian reflectance) and geometric surface attributes. This is an advantage over the state-of-the-art methods [7], [16] which assume equal importance to all surfaces. To improve the performance further, we introduce temporal constraints, extending the method to video sequences. For static objects, it is assumed that the light source position does not change during the recording of the video. We derive a temporal constraint by estimating the camera pose. The camera pose is used to estimate a global LSP which minimizes the residuals between the re-rendered video frames and the original video frames. See Fig. 2 for the outline of the proposed method. The proposed method is tested on two different datasets (*Boom13* [7] and our newly collected video dataset). Experiments on these datasets show that surface weighting provides a significant improvement over the state-of-the-art methods.

The paper has the following contributions. First, surface attributes are differentiated according to their importance for LSP estimation. Second, in contrast to state-of-the-art methods, which declare all surface contributions as equally important, we derive weights for individual image surface segments based on their suitability. Third, a geometry-based initialization is proposed to make the light source LSP estimation specific to the underlying image and to ensure fast convergence. Fourth, for videos, we introduce a global consistency term to constrain the light source positioning by estimating the camera pose. Finally, we prove the viability of the method on a new (video) dataset, to be made publicly available.

## II. Related Work

In general, light source positioning algorithms assume certain $3D$ object shapes or known objects in the scene. For instance, [11] assumes that the position of objects in the scene is known and uses cast shadows to estimate the light source direction. [19] uses a fisheye camera to recover the position of the light source for indoor scenes. Hara et al. [20] proposed two different methods to estimate light source position. The first method has strong requirements such that (1) inter-reflection and cast shadows are avoided. Therefore, the proposed method is limited to convex objects. (2) saturated pixel values are avoided, (3) $3D$ geometric model of an object is given and (4) at least one specular peak is visible on the surface of the target object. The method tries to iteratively fit Lambertian reflection model and uses the difference between original and reconstructed diffuse component images to fit specular component. Finally, the iterative process is repeated until the light source position does not change. To eliminate the limitation caused by the lambertian approximation of diffuse reflection of the first method, the second method assumes that in addition to 3D geometric model, specular component image is given as input. Then the light source position is estimated by minimizing the linearity of log-transformed Torrance-Sparrow model using the given specular component image. The second method also has similar requirements like the first method (1) inter-reflection and cast shadows are avoided. Therefore, the proposed method is limited to convex objects, (2) $3D$ geometric model of an object is given, (3) specular component image is given and (3) the specular reflectance can be modelled by the Torrance-Sparrow model. These methods considers diffuse and specular reflections separately while minimizing the reconstruction error. However, to consider diffuse and specular image components separately, the algorithm either requires an information as input (e.g. requiring specular component) or constrains the imaging condition (e.g. at least one specular peak is visible, inter-reflection, cast shadows avoided and saturated pixel values are avoided). Others, e.g. [13], use GPS and compass information to determine the sun position in outdoor scenes. In [12], various cues are extracted from the sky, vertical surfaces and the ground to estimate the direction of the sun. Although these methods may be suited to estimate the light source position, they impose strong assumptions on the imaging conditions.

More recently, a number of methods have been proposed that use depth information. For instance, [15] decompose the image into specular, diffuse and albedo layers. The specular and diffuse layers are used to constrain the difference between the re-rendered and original image. However, detecting specular parts in a 2D still image is a difficult problem. Moreover, each surface segment is considered to equally contribute to LSP estimation. [21] also makes use of specular and diffuse image components. [21] combines separately estimated light source positions from diffuse and specular components into final light source positions. [14] tries to estimate the lighting environment using spherical harmonics. The authors assume the entire scene obeying a diffuse Lambertian reflectance model. In [22] the illumination is estimated based on inverse rendering. However, the algorithm requires an off-line process to recover the reflectance and manually select sample points to estimate the illuminant in the target scene. [7] assumes that image surface segments with the same color have the same albedo. The authors re-render the scene from a hypothesized LSP while minimizing the error between the synthesized and original image. Moreover, the authors gain substantial speed by enabling the power of GPU processing in [23]. There is a certain performance drop, however performing under 1seconds makes this approach attractive for real-time application.

In contrast to the previous methods, we propose to learn the suitability of surfaces to estimate the LSP based on their surface attributes. We assign weights to the image surface segments rather than treating them all equally.

Some recent works solve the problem of lighting estimation as intrinsic image decomposition and try to decompose $RGB-D$ scenes into reflectance and shading components [24], [25]. Barron et al. [24] formulates the problem as a non-convex function and iteratively decompose the $RGB-D$ scenes into reflectance and shading images and smooth the depth geometry while using the mixture of the shape and the mixture of the illuminant as priors. Chen et al. [25] smooths the depth image using an off-line smoothing algorithm. Then formulates the problem as a convex function by not considering smoothing the depth image during optimization. Thus the objective function can be reliably optimized and generate more robust results.

## III. LSP Estimation Using Surface Suitability

According to Lambert's law, an intensity pixel value $I$ can be modeled by:

$$I(u) = \rho(u) min(\boldsymbol{n}(u)(\frac{L-p(u)}{\|L-p(u)\|})^t \imath, 0) \quad , \tag{1}$$

where $t$ represents the transpose. The intensity value $I$ at pixel $u$, depends on the surface albedo $\rho$, the surface normal $\boldsymbol{n}$, the light source direction and the intensity $\imath$ of the light. The light source direction is defined as the direction between the light source position $L$ and the point $p$ in $3D$ coordinates. From the RGB-D sensor, both the intensity $I$ and the depth images are provided. Further, $\boldsymbol{n}$ is computed from the depth image using integral images and average 3D gradient method provided by PCL library [26]. For LSP estimation, the albedo $\rho$ and the light intensity $\imath$ are unknown. Hence, it is necessary to estimate $\rho$ and $\imath$. Image surface segments are generated by using [18]. It is assumed that image surface segments have uniform albedo [7], [16], [15]. Thus, $\rho$ does not change within an image surface segment. Moreover, it is assumed in [7] that $\imath$ remains constant because the light source distance does not vary significantly over neighboring pixels (image surface segment). We also do not use the inverse square law to adjust $\imath$. We set $\imath$ to be 1. Under these assumptions, the LSP is estimated by minimizing the error between the reconstructed and original image surface segments based on the hypothesized LSP.

### A. Surface Reconstruction Using Surface Suitability

To reconstruct the image surface segment $S$, $\rho$ values are required. Given an arbitrary $L$, $\rho$ values for surface $s_i$ are computed by:

$$\rho(u) = \frac{I(u)}{\boldsymbol{n}(u)(\frac{L-p(u)}{\|L-p(u)\|})^t \imath}, u \in s_i \quad . \tag{2}$$

The median of $\rho$ values are used to obtain a single $\rho$ value for image surface segment $s_i$. Then, the reconstructed image $I_r$ for $s_i$ is computed by:

$$I_r(u) = \rho_{s_i} min(\boldsymbol{n}(u)(\frac{L-p(u)}{\|L-p(u)\|})^t \imath, 0), u \in s_i \ , \tag{3}$$

we obtain the light source position $L$ by minimizing the error $E$ between the original intensity values $I$ and the reconstructed intensity values $I_r$.

$$E_i = \sum_{u \in s_i} f(s_i) \|I(u) - I_r(u)\| \quad , \tag{4}$$

$$E = \sum_{i \in S} E_i \ . \tag{5}$$

Unlike other methods [7], [15], where each image surface segment contributes equally to the total reconstruction error, we propose to assign weights $f$ to each image surface segment based on its suitability to compute the LSP. The aim is to assign more importance to image surface segments which are more suitable for LSP estimation under the assumption of Lambert's law. To this end, surface attributes, characterize the surface suitability for LSP estimation, are extracted. Then, the weights $f$ are learned using these attributes in a supervised learning. The surface attributes and learning procedure are detailed in sections III-B and III-C.

### B. Surface Attributes for LSP Estimation

Lambert's law assumes a surface which diffusely reflects the light. The surfaces which satisfy this condition are more suited to estimate LSP. For instance, matte surfaces are preferred over glossy surfaces to estimate the LSP. The specular reflections are not considered by Lambert's law. Consequently the surfaces with highlight will negatively influence the LSP estimation. A cast shadow is caused by the occlusion of the light source position. That means that there are no light rays reaching the surface directly coming from the light source. Hence, intensity values would be misleading for LSP estimation (assuming Lambert's law). Therefore, LSP estimation

| | Intensity | Chromatic | Normalized Chromatic | Hue |
|---|---|---|---|---|
| Representation | $O_3$ | $[O_1, O_2]$ | $[\frac{O_1}{O_3}, \frac{O_2}{O_3}]$ | $\frac{O_1}{O_2}$ |
| Invariant to | - | Highlights | Shadows | Highlights<br>Shadows |

TABLE I
OPPONENT COLOR SPACE IMAGE REPRESENTATIONS AND INVARIANT PROPERTIES [29].

from a shadow region would be prohibitive. Not only the photometric attributes, but also the geometric attributes of a surface is influential to estimate LSP. For instance, the surface normals on rough surfaces are prone to be more noisy than smooth surfaces. The intensity value is determined by the angle between the surface normal and the incident light direction. Therefore, noisy surface normals will negatively influence LSP estimation. Subsequently, smooth surfaces are preferred over rough surfaces.

It is clear that some surfaces have preferred attributes to estimate LSP. To this end, we define surface attributes. These attributes are further used in a learning scheme to measure the suitability of a surface to estimate LSP.

*1) Photometric Representations:* We aim to represent surfaces by their photometric attributes (e.g. glossiness). Moreover, it is important to identify surfaces under different photometric changes (e.g. highlights and shadows). The opponent color space is used to represent different photometric invariants [27], [28], [29]. The transformation of $RGB$ to $O_1O_2O_3$ is given by:

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} \frac{R-G}{\sqrt{2}} \\ \frac{R+G-2B}{\sqrt{6}} \\ \frac{R+G+B}{\sqrt{3}} \end{pmatrix}. \tag{6}$$

The different properties of the opponent color and their combinations are summarized in Table I. The intensity information is represented by $O_3$. It has no invariant properties. Therefore, changes due to shadows and highlights are represented [27], [29]. Color information is contained in $O_1$ and $O_2$. Due to the subtraction in $O_1$ and $O_2$, they are invariant to shifts in illumination such as highlights [27], [29]. In addition, we use $hue = \frac{O_1}{O_2}$ to account for both shadow and highlight invariances.

*2) Surface Attributes:* Extracting material characteristics from images has been studied in [29], [30], [31], [32], [33]. These works consider material recognition as a texture classification problem. [34], [35] perform a detection in the image for common materials such as stone, wood, metal, fabric etc. Material characterization of a surface reveals important information about the surface property. For instance, in general, metal is hard and glossy whereas plastic is soft and matte. Surface attributes would highly benefit from material characterization. Therefore, in this paper, material characterization is used to extract surface attributes to assign importance to the surfaces based on their suitability to estimate LSP. The opponent color and combinations (see section III-B1), $RGB$ and depth images are used to extract surface attributes (surface attributes related to [29] are extracted). We detail these attributes in this section.

**Invariant Response.** A cast shadow is caused by occlusion of the light source. Therefore, the light source position cannot be derived from a shadow region (assuming Lambert's law). Moreover, highlights are generated by specular reflections. The assumption of Lambert's law does not hold for highlight regions. To be able to distinguish image surface segments with shadows and highlights, we use the method proposed by [28] which measures the average gradient magnitude ratio defining the invariant response ($\frac{|\bigtriangledown O_3|}{\sqrt{|\bigtriangledown R|^2+|\bigtriangledown G|^2+|\bigtriangledown B|^2}}$ where $|\ .\ |$ stands for gradient magnitude). The image surface segments consist of uniform colors. Hence, the average gradient magnitude is expected to be low. However, shadows and highlights cause photometric edges. Therefore, the image surface segments with shadows and highlights are expected to have high invariant response.

**Photometric Stability.** Invariant representations contain instabilities. For instance, hue is unstable for colors with low saturation $\frac{O_1}{O_2}$ [29]. The surface reflection characteristics make the instabilities to vary for different surfaces. To account for the influence of instabilities, we use the method proposed by Everts et al. [29]. Mean intensity ($\mu(O_3)$) and saturation ($\mu(\sqrt{O_1^2+O_2^2})$) statistics are considered to measure the photometric stability of a surface.

**Interface Reflectance.** Lambert's law assumes a surface which diffusely reflects the light. Interface (specular) reflectance is not defined by Lambert's law. Therefore, it is difficult to estimate the LSP from glossy surfaces. Moreover, the depth sensor is sensitive to glossy surfaces (e.g. shiny metal, mirror). The depth estimation becomes unstable on glossy surfaces. Therefore, surface normals are mostly noisy. To this end, we propose to extract an attribute which aims to detect glossiness of surfaces. Motoyoshi et al. [33] propose that the skewness (third-moment) of the intensity histogram is highly correlated with interface reflectance (gloss) and inversely correlated with diffuse reflectance (matte). Others, such as Sharan et al. [32] use the standard deviation whereas Dror et al. [30] use the kurtosis to account for interface reflectance. We also use skewness, standard deviation and kurtosis to measure the amount of interface reflectance using the $O_3$ component.

**Colorfulness.** Hue is invariant to shadows and highlights. Therefore, these photometrical changes should not influence the hue distribution of a surface segment. The assumption is that the albedo does not change within a surface segment (see section III-A eq. 2). However, the variation in hue distribution most likely corresponds to the albedo change. Thus, the same albedo assumption may mislead the light source position estimation. To this end, we propose to use colorfulness by computing the hue entropy as in [29] ($-\sum(\mathbf{P}\log_2\mathbf{P})$). $\mathbf{P}$ represents the histogram of hue pixels.

**Softness.** Softness is useful to distinguish surfaces having diffuse (i.e. soft-plastic) or specular (i.e. hard-metal) reflection. Hu et al. [35] state that metal tends to have hard edges and sharp corners whereas plastic has soft edges and round corners. To this end, we measure the softness using the standard deviation of the gradient orientation ($\sigma(\bigtriangledown O_3)$) and magnitude ($\sigma(\mid \bigtriangledown O_3 \mid)$).

**Texturedness.** Most of the LSP algorithms use segmentation to group similar colored surfaces assuming that the pixels of the same surface segment have the same albedo. However, surfaces may also contain similarly colored textures such as crinkles in leather or grains in paper. These crinkles or grains will cause sharp intensity changes which may negatively effect the light source position estimation. To this end, we compute two Weibull parameters for the $O_3$ as proposed by Yanulevskaya and Geusebroek [31] to measure the amount of texturedness of a surface.

**Micro-texture.** The local non-uniformities on surfaces can be used to describe surface structure. Less micro-texture indicates polished glossy surfaces (e.g. metal) whereas more micro-texture indicates matte surfaces (e.g. fabric). Because of the diffuse reflection assumption of Lambert's law, these two types of surfaces are expected to influence the LSP estimation differently. The method proposed by Liu et al. [34] is used to measure the amount of micro-texture, In particular, we use the sum of residuals between a bilaterally smoothed $O_3$, $h(O_3)$, and the original $O_3$ ($\sum(h(O_3) - O_3)$).

**Smoothness.** Due to the imperfections of recording devices, the surface normals may be noisy on rough, crinkled and grained surfaces. Lambert' law uses the angle between the surface normals and light source position to minimize the error between re-rendered and original image. Noisy surface normals will negatively influence the error minimization. Smoothness aims at differentiating between rough and smooth surfaces. Unlike micro-texture, we do not consider smoothness in micro-scale. The smoothness term is defined by larger scale geometry changes. Smooth surfaces do not consists of convex and concave shapes at the same time. We use the statistics of the surface normals (from the depth image) to measure surface smoothness ($\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{n_i} - \mu(\boldsymbol{n}))^2}$) ($N$ is the number of the points in a segment) and the mean of the gradient magnitude ($\mu(\mid \bigtriangledown \boldsymbol{n} \mid)$) of the surface normals. The first statistic is useful to observe overall deviation on an image surface segment whereas the second one is useful to observe local deviations. High values correspond to rougher surfaces. The surface normals are computed from the depth image using [26].

| Attribute | Definition | Information Channel |
|---|---|---|
| Invariant Response | $\frac{\mid\bigtriangledown O_3\mid}{\sqrt{\mid\bigtriangledown R\mid^2+\mid\bigtriangledown G\mid^2+\mid\bigtriangledown B\mid^2}}$ | $[O_3, RGB]$ |
| Photometric Stability | $\mu(O_3)$, $\mu(\sqrt{O_1^2 + O_2^2})$ | $[O_1, O_2, O_3]$ |
| Interface Reflectance | $Skew., \sigma, Kurt.$ | $O_3$ |
| Colorfulness | $-\sum(\mathbf{P}\log_2\mathbf{P})$ | $\frac{O_1}{O_2}$ |
| Softness | $\sigma(\bigtriangledown O_3)$, $\sigma(\mid \bigtriangledown O_3 \mid)$ | $O_3$ |
| Texturedness | $[\gamma, \beta] = weibull$ [31] | $O_3$ |
| Micro-texture | $\sum(h(O_3) - O_3)$ | $O_3$ |
| Smoothness | $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\boldsymbol{n_i} - \mu(\boldsymbol{n}))^2}$,<br>$\mu(\mid \bigtriangledown \boldsymbol{n} \mid)$ | $depth$ |
| Area | $\#pixels$ | $RGB$ |
| Surface Consistency | $L - \mu(L')$ | $[O_3, depth]$ |
| Curvedness | $\sigma(\mid \bigtriangledown \boldsymbol{n} \mid)$ | $depth$ |

TABLE II
DEFINITIONS OF EXTRACTED ATTRIBUTES.

**Area.** The surface normals and intensity values may be noisy due to imperfections of recording devices. The variation of intensity distributions is important to alleviate these errors. Therefore, larger image segments are expected to contribute more to proper LSP estimation than smaller segments.

**Surface Consistency.** Estimated light source positions $L' = \{L_i\}_{i=1}^{m}$ ($m$ is the number of the image surface segments in an image) should be consistent. Therefore, we express the surface consistency attributes by measuring the deviation from the average estimation for each image surface segment. For the $i^{th}$ surface segment, surface consistency is measured by $L_i - \mu(L')$.

**Curvedness.** Curvedness distinguishes surfaces which have non-flat surfaces. Considering the richer surface normal representations of curved regions, we expect LSP estimation to be more precise for highly curved surfaces. To this end, using surface normals, the curvedness attribute is expressed by the standard deviation of the gradient magnitude ($\sigma(\mid \bigtriangledown \boldsymbol{n} \mid)$).

### C. Image Surface Segment Suitability by Ranking

The suitability of an image surface segment is defined by the angular error between the estimated and the ground-truth light source position. The aim is to measure which image surface segment is preferred over others. Therefore, we consider the learning process as a ranking problem.

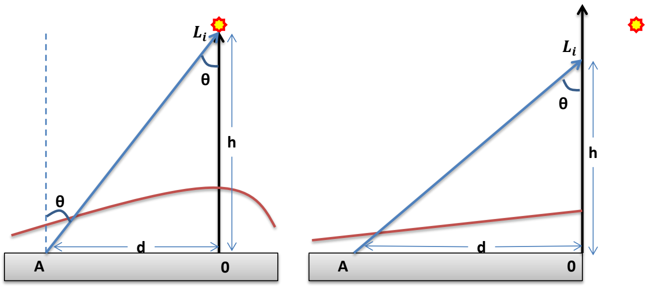


Fig. 3. Initial light position $L_i$ based on $3D$ geometry constraint. $O$ is the point with maximum intensity on the surface. $A$ is a random point on the surface of which the intensity is known. $L_i$ is estimated as an initial guess for LSP which is on the direction of the surface normal of point $O$ and satisfies Lambert's law for both points.

We use learning to rank (L2R) [36] to measure the suitability of different image surface segments for LSP estimation. The training set consists of image surface segments $S = \{s_i\}_{i=1}^{m}$ ($m$ is the number of image surface segments) and ground-truth label $y$ expressed in terms of angular error between ground-truth and estimated light source positions. Image surface segments are generated by using [18]. Attributes $\Phi$ are extracted from image surface segments as explained in section III-B. $\Phi$ and $y$ are used to learn $f$ which is described as follows:

$$f(s_i) = w^t \Phi(s_i). \tag{7}$$

The objective function to optimize the weight $w$ is described as follows:

$$\min_{w} \frac{1}{2} w^t w + C \sum_{i=1}^{l} \xi(w; \Phi(s_i), y_i), \tag{8}$$

where $C$ and $\xi$ represent regularization parameter and loss function respectively. The default value for the C parameter (=1) is used without tuning whereas various loss functions are used based on the choice of the learning to rank algorithm (i.e. pointwise, pairwise). Support vector classifier $SVC$ and support vector regressor $SVR$ are used as pointwise methods. $RankSVM$ [37] is used as a pairwise method. Their loss functions $\xi$ for $SVC$, $SVR$ and $RankSVM$ are $\max(0, 1 - y_i w^t \Phi(s_i))$, $(\max(0, |y_i - w^t \Phi(s_i)| - \epsilon))^2$ and $\max(0, 1 + w^t \Phi(s_i) - w^t \Phi(s_j))$ respectively [38]. The sensitiveness of the loss functions are determined by $\epsilon$ parameters. The scores of $f$ are used in eq. 4 to assign importance to the image surface segments. The image surface segments with high scores are more suitable for light source position estimation. Image surface segments which have a negative influence on LSP estimation (surfaces which have negative scores) are filtered out.

### D. LSP Initialization and Search

The downhill simplex method [39] is used to minimize $E$ in eq. 4. The state-of-the-art method [7] uses the camera viewpoint to initialize the light source position. However, to solve the minimization problem, a proper initial light source position is important to obtain fast convergence. Unlike the state-of-the-art [7], we propose to use constraints imposed by the $3D$ geometry to select the initial points. Assuming Lambert's law, the points on the object surface that receive the light close to a perpendicular angle have maximum intensity. Let $O$ be the perpendicular projection of $L$ on the planar surface. $\theta$ is the angle between the light source direction and the surface normal at surface point $A$. Then, $\cos(\theta) = I(O)/I(A)$, where $I(O)$ and $I(A)$ are intensity values at positions $O$ and $A$. The distance $d$ between points $O$ and $A$ is computed by the 3D coordinates, see Fig. 3. Finally, the height $h$ of the light source is given by:

$$h = d \frac{\cos(\theta)}{\sqrt{1 - \cos(\theta)^2}}. \tag{9}$$

The perpendicular projection point $O$ may not always be available on a surface. That is the reason to still run the full optimization. Otherwise, it would be possible to estimate light source position directly based on above equation 9. Nevertheless, the point with the maximum intensity on a surface will still be the closest to have minimum angle to the light source position (due to Lambert's Law). Therefore, the position provided by the light source position initialization will be closer to light source than random positions. This is visualized in Fig. 3 (sample on the right).

Initial light source positions are estimated for all the image surface segments in an image. These estimations are used to initialize the arbitrary light source position in eq. 3 by a weighted average. Moreover, to obtain surface consistency (see section III-B1), it is necessary to have a LSP estimation for each image surface segment. The light source position estimation accuracy should not change with or without good initialization because the optimization is allowed to run until convergence. However, $3D$ geometry-based initialization allows a speed-up of convergence for each individual estimation (5% faster convergence).

## IV. LSP Estimation from RGB-D Sequences

Temporal information can be used to improve the accuracy of the single frame-based LSP estimation. We assume that $L$, with respect to static objects in the scene, does not change during a single video recording. Hence, the only change is the relative position of $L$ with respect to the camera. Therefore, we propose to use the camera pose to provide temporal constraints in $RGB-D$ sequences. First, we estimate the camera pose to build correspondences between frames. Then, images are transformed to the same coordinate system to create consistency between estimations of different frames.

### A. Camera Pose Estimation

Considering static objects in the scene, we propose to estimate the camera pose as a rigid body movement. In our framework, the iterative closest point (ICP) algorithm is used to estimate the camera pose [40]. ICP estimates the camera pose by aligning the data. Data alignment problem is treated as a nonlinear optimization problem in which correspondences between recordings (depth images) are approximated using the closest pairs of points found between successive depth images [41], [40]. After the computation of corresponding points, ICP aims to find a single transformation matrix $\mathbf{T}$ with minimal point-to-plane error [40]:

$$\arg\min \sum_{u} \|(\mathbf{T}v_i(u) - v^g_{i-1}(u))^t \boldsymbol{n}^g_{i-1}(u)\|^2. \tag{10}$$

The error is measured by how good each point $v_i(u)$ in the current frame fits the tangent plane at its corresponding point $v^g_{i-1}(u)$ in the previous frame [41], [40]. $\boldsymbol{n}^g_{i-1}(u)$ is the surface normal of the corresponding point $v^g_{i-1}(u)$ in the previous frame. Using a global coordinate $g$, the camera pose $\mathbf{T}$ is used to transform point $v_i(u)$ from the image coordinate to the global coordinate. Then, a linear approximation is adopted to solve this system. In our approach, a GPU-based implementation of ICP is used which provides real-time camera pose estimation.

### B. Global LSP Refinement

After the camera pose is estimated by ICP, the proposed LSP estimation method is applied. To incorporate all video frames, $L$ is transformed from local image coordinates to a global one. As a result, given an image $I_i$, its corresponding $\mathbf{T}$ and $L$, $\rho$ values in $s_j$ of eq. 2 are modified as follows:

$$\rho(u, \mathbf{T}) = \frac{I_i(u)}{\boldsymbol{n}(u)(\frac{\mathbf{T}L - p(u)}{\|\mathbf{T}L - p(u)\|})^t \imath}, u \in s_j. \tag{11}$$

Then, $I_r$, for a given $s_j$, is computed by:

$$I_r(u, \mathbf{T}) = \rho(u, \mathbf{T}) \min(\boldsymbol{n}(u)(\frac{\mathbf{T}L - p(u)}{\|\mathbf{T}L - p(u)\|})^t \imath, 0), u \in s_j, \tag{12}$$

$$E_{i,j}(u, \mathbf{T}) = \sum_{u \in s_j} f(s_j)\|I_i(u) - I_r(u, \mathbf{T})\|, \tag{13}$$

and the residual error of $I_i$ is computed as follows:

$$E_i(u, \mathbf{T}) = \sum_{s_j \in I_i} E_{i,j}(u, \mathbf{T}). \tag{14}$$

Given an image sequences $I_c$, the energy function for the light source position is then defined by:

$$E = \sum_{I_i \in I_c} E_i(u, \mathbf{T}). \tag{15}$$

The estimated light source position is obtained by minimizing $E$ given by eq. 15. Finally, the light source position which minimizes the residuals between the reconstructed and original video sequence is selected as the final estimation.

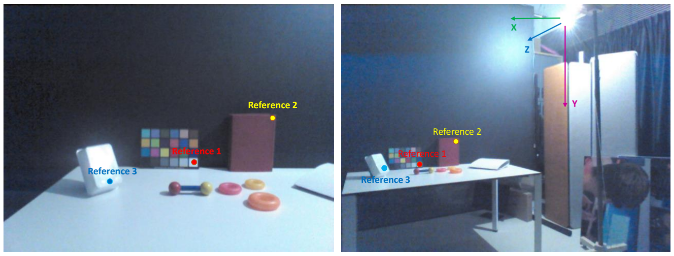


Fig. 4. A sample scene of the setup of the recorded $RGB-D$ Video dataset and three manually picked salient points on this sample image.

## V. Experiments

**Datasets and Evaluation Metric.** The proposed light position estimation algorithm is evaluated on the dataset proposed by [7] and our newly collected dataset. Thanks to the authors of [7], they have provided $Boom13$ dataset with ground-truth annotations. Currently, the dataset is publicly available[1].

Our RGB-D video dataset is collected using Kinect in a dark room. The room is isolated from light sources. First, we have placed the objects and a Philips daylight simulator bulb in the room (See Fig. 4). The room is lightened only using this bulb. The light source position and objects are fixed for each video sequence but vary between different video sequences. The relative light source position with respect to the objects varies within 3 meters. We have selected three salient control points (e.g. color checker white patch, book corner and bottom of white plastic). We have manually measured (using ruler) x, y and z distances between the light source and the control points (in real world). Then, we have performed video recordings. For each frame, we have obtained the coordinates of the control points with respect to camera. Accordingly, we have transformed the light source position to camera coordinates using the measured distances between light source position and control points. Three video sequences are recorded (approximately around $15fps$). The length of the videos vary between 20 to 60 seconds. While building the dataset, we randomly subsample (i.e. 1fps) the videos for the simplicity of the annotation process. At the end, the dataset consists of 71 frames.

The angular and euclidean distance error between the estimated and ground-truth light source positions are used to measure the accuracy.

**Implementation.** Two types of learning to rank (L2R) methods are used, namely, pointwise and pairwise [36]. The main difference between these methods are their objective functions [38]. Pointwise methods aim to minimize the error based on single instances, whereas pairwise methods minimize the disorder between pairs. Pointwise methods require numerical scores for training labels, whereas pairwise methods use preferences between pairs. For pointwise and pairwise methods the $Liblinear$ [42] and Joachims [37] implementations are used respectively. The default parameter settings are used as provided by the implementations. Leave-one-out is used for training and testing the performance of the proposed method.

[1] http://www.dtic.ua.es/jgpu12/lightEstimation/

| Method | Performance (Mean Angular Error) |
|---|---|
| Boom et al. [7] | 12.6°±6.4° |
| Proposed-SVC20 | 9.9°±6.1° |
| Proposed-SVC25 | 8.6°±5.6° |
| Proposed-SVC30 | 9.8°±6.3° |
| Proposed-SVR | 9.9°±5.7° |
| **Proposed-RankSVM** | **8.2°**±5.1° |

TABLE III
LSP estimation performance on *Boom13 dataset*. There is no threshold for angular error to consider an image surface segment to be good/bad. Varying thresholds are used to specify positive or negative labels. The numbers next to $SVC$ represent the angular error threshold used. The proposed attribute-based LSP algorithm outperforms [7] which assumes equal importance to all surfaces. Various learning algorithms are also tested, namely, support vector classifier $SVC$, support vector regressor $SVR$ and $RankSVM$. The results show that learning the surface attributes outperforms the method without learning [7] regardless the choice of the learning algorithm. $RankSVM$ performs best.

### A. LSP Estimation from a Single RGB-D Frame

**Experiment I : Influence of Learning Surface Attributes** We evaluate our attribute-based LSP algorithm on the Boom13 dataset [7] and compare it with [7]. We follow the same steps for both algorithms. The main difference between the obtained results is that [7] gives equal weights to all image surface segments whereas our method assigns a weight to the each image surface segment based on its suitability to contribute to a correct LSP estimation. The results are summarized in Table III. The results show that the proposed algorithm outperforms [7]. The significant improvement over [7] indicates the importance of surface attributes to estimate LSP. Hence, image surface segments influence LSP estimation differently based on their appropriateness.

**Experiment II: Influence of Learning Algorithms** In this experiment, we evaluate three different learning algorithms to rank the image surface segments. Support vector classifier $SVC$ and support vector regressor $SVR$ are used as pointwise methods. $RankSVM$ [37] is used as a pairwise method. These algorithms differ mainly by their loss functions $\xi(w;\Phi(s_i),y_i)$ of eq. 8. $\xi$ for $SVC$, $SVR$ and $RankSVM$ are $\max(0,1-y_i w^t\Phi(s_i))$, $(\max(0,|y_i-w^t\Phi(s_i)|-\epsilon))^2$ and $\max(0,1+w^t\Phi(s_i)-w^t\Phi(s_j))$ respectively [38]. The sensitiveness of the loss functions are determined by $\epsilon$ parameters. $\Phi(s)$, $w$ and $y$ stand for the surface attribute, the weights and the labels respectively.

The angular error between the estimated and ground-truth light source positions are used as training labels. Since there is no threshold for the angular error to determine an image surface segment to be good/bad, we use varying thresholds for the angular error to specify image surface segments to be positive or negative labels for $SVC$. The angular errors are directly used as training labels for $SVR$. $RankSVM$ requires pairwise preferences between image surface segments. These preferences are created based on their angular errors.

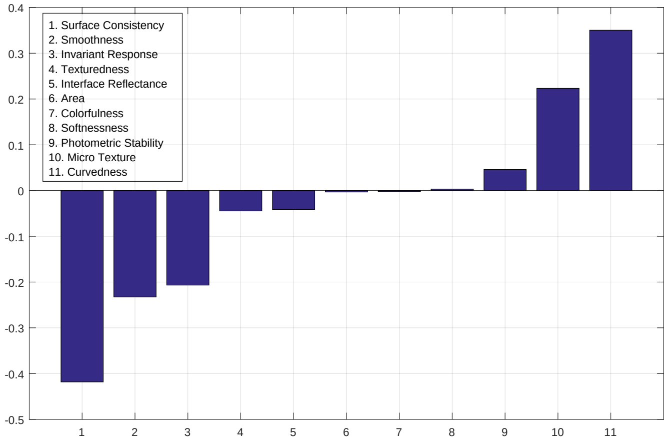


Fig. 5. Attribute weights: The weights are obtained by averaging the summed classifier weights of different dimensions of the same attributes. It illustrates that surface attributes influence the LSP estimation differently.

The results show that learning the surface attributes outperforms the method without learning [7] regardless the choice of the learning algorithm (See Table III). This indicates the importance of learning surface relevance for LSP estimation. $SVC$ using 25° error threshold performs comparable results to $RankSVM$. However, the necessity of choosing a labeling threshold makes $SVC$ less practical. $RankSVM$ performs the best without introducing any hand-crafted rules for labeling. Therefore, $RankSVM$ is used for the rest of the paper.

In addition, we also evaluated mean Euclidean distance between estimated and ground-truth light source position as error measure. The proposed method reaches $1.2m \pm 0.6m$ mean Euclidean distance error whereas [7] reaches $1.7m \pm 0.8m$.

**Experiment III: Influence of Surface Attributes** In this experiment, we study the influence of each individual surface attributes. The weights are obtained by averaging the summed classifier weights of different dimensions of the same attributes. The attribute importance is summarized in Fig. 5.

Surface consistency is defined by the deviation of an image surface segment LSP estimation from the average estimation of the other image surface segments in the image. The results show the importance of a global consistency condition. Another conclusion is that proper image surface segments vote for similar light source positions. Deviating from the average estimation of image surface segments negatively influence the importance of an image surface segment.

Deviation from smoothness of a surface is observed to be negatively related to the correctness of the estimation. This is due to noisy surface normals extracted from rough surfaces mislead the optimization algorithm. Moreover, rough surfaces are more prone to cast shadows due to the occlusion of the light source.

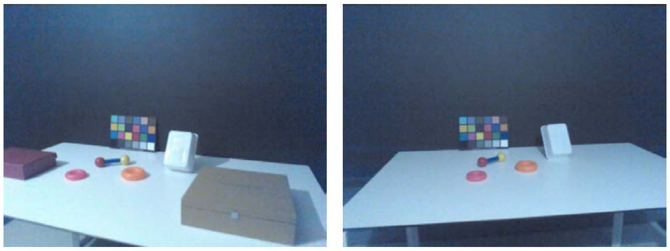


Fig. 6. Sample images from our $RGB-D$ video dataset. Images are from different video sequences.

The surfaces are segmented based on their color. Shadows and highlights cause gradient changes within the same colored segments. Invariant response takes into account this. Higher average gradient ratio corresponds to shadows and highlights. Considering Lambert's law, it is expected that light source position estimation is negatively affected by invariant response.

The amount of texturedness on LSP estimation is important. This is due to the surface homogeneity assumption of LSP algorithms. They assume that the intensity changes are caused by shading. However, changes caused by the surface texture negatively influence the optimization algorithm to reach convergence. Therefore, as expected, less textured regions are more useful for LSP estimation.

Interface reflectance is highly correlated with surface property of being matte or glossy. Surfaces become more glossy with an increasing amount of interface reflectance. Since Lamberts law assumes a surface which diffusely reflects the light, the light position estimation is negatively affected by interface reflectance.

The increase in colorfulness most likely indicates albedo change within a surface. This conflicts with the assumption, albedo does not change within a surface segment. Thus, light source position estimation is negatively affected by colorfulness.

The surfaces being photometrically stable has a positive influence on the light source position estimation.

Less micro-texture mostly indicates polished, glossy surfaces (e.g. metal) whereas more micro-texture indicates matte surfaces (e.g. fabric). Because glossy surfaces are more prone to be affected by interface reflectance, the amount of micro-texture positively influences the LSP estimation.

The amount of curvedness has a positive affect on the LSP estimation accuracy. Curved surfaces create more variations of surface normals. This provides intensity variations even for small regions. Whereas a flat surface usually changes monotonically and does not create such intensity variation.

### B. LSP Estimation from a RGB-D Video Sequence

**Experiment I: Influence of Temporal Constraints** In this experiment, we conduct two experiments. First, we compare the performance of the proposed attribute-based LSP estimation algorithm with [7]. Second, we compare the performance of LSP estimation based on a single frame with video sequence. The results are summarized in Table IV: The results for the first experiment show that our attribute-based method (mean angular error $8.5°$) outperforms [7] (mean angular error $24.6°$). For the second experiment, we estimate a single global light source position for the whole video sequence using the proposed temporal constraints. To obtain the light source position for a single image, the estimated global light source position is transformed into local image coordinates (using the estimated camera pose). Then, the errors are measured. For each sequence, our temporally constrained LSP algorithm reduces the LSP estimation error from $8.5°$ to $4.8°$.

| Method | Performance<br>(Mean Angular Error) |
|---|---|
| Boom et al. [7] | 24.6°±6.0° |
| Proposed Attribute | 8.5°±2.4° |
| Boom et al. [7] + Proposed Temporal | 6.3°±4.1° |
| **Proposed Attribute+Temporal** | **4.8°±2.9°** |

TABLE IV
LSP ESTIMATION PERFORMANCE ON *Our dataset*. THE PROPOSED ATTRIBUTE-BASED METHOD OUTPERFORMS [7]. THE PROPOSED TEMPORAL CONSTRAINTS IMPROVES THE ACCURACY OF THE PROPOSED ATTRIBUTE-BASED METHOD AND AN OFF-THE-SHELF LSP ESTIMATION METHOD [7].

In addition, we evaluated mean Euclidean distance between estimated and ground-truth light source position as error measure. The proposed method reaches $0.5m \pm 0.2m$ mean Euclidean distance error whereas [7] reaches $1.9m \pm 5.8m$.

**Experiment II: Improving off-the-shelf LSP Estimation Method** In this experiment, we use a state-of-the-art LSP algorithm [7] and apply the proposed global refinement step. The objective function is replaced by the proposed temporal constraint. The mean error is reduced from $24.6°$ to $6.3°$ with respect to the original LSP algorithm [7] (See Table IV).

### C. Influence of the Segmentation Algorithm on LSP Estimation

The light source position estimation error minimization function heavily relies on the homogeneous segments due to the assumption that the pixels in the same segments have the same albedo. Therefore, the errors caused by the segmentation algorithm are expected to harm the light source position estimation. In this experiment, we compare the results of the proposed method and [7] using graph-based [18] and quick shift [43] segmentation algorithms. The results are reported in Table V.

It can be noted that the result of the proposed method is influenced by the choice of the segmentation algorithm whereas depending on the segmentation algorithm, the result of [7] significantly varies in $RGB-D$ Video dataset. The proposed surface attributes (i.e. colorfulness and texturedness) helps the proposed algorithm to give less importance to those surfaces with segmentation failures. Therefore, the proposed algorithm is less sensitive to segmentation algorithm and segmentation errors.

| Segmentation Method | Boom13 Dataset | | Proposed RGB-D Video Dataset | |
|---|---|---|---|---|
| | Proposed | Boom et al. [7] | Proposed | Boom et al. [7] |
| Felzenszwalb [18] | 8.2°±5.1° | 12.6°±6.4° | 4.8°±2.9° | 24.6°±6.0° |
| quick-shift [43] | 8.8°±5.1° | 12.9°±10.6° | 6.0°±2.5° | 12.9°±6.0° |

TABLE V
THE LSP RESULTS ON $Boom13$ AND $RGB-D$ VIDEO DATASETS USING DIFFERENT SEGMENTATION ALGORITHMS. THE RESULTS SHOW THAT THE PROPOSED METHOD IS LESS SENSITIVE TO SEGMENTATION ALGORITHM AND SEGMENTATION ERRORS.

### D. Discussion

**Complexity Analysis:** In this paper, we focus only on the theoretical contributions rather than practical contributions (real-time applicability). Therefore, the proposed method would highly benefit from a careful engineering. Given that, the proposed method has been tested on a desktop machine with an Intel(R) Core(TM) $i7-4810MQCPU$ @$2.803Ghz$. The operations and their time consumptions are reported in Table VI.

At this moment, the most time consuming step is the feature extraction. By exploiting the computing power of GPU, feature extraction can be computed in a parallel architecture, which will help to reduce the computing time significantly. A significant speed gain could be obtained in the error minimization step by minimizing the reconstruction error of each segment using different GPU threads.

The total execution time of [7] is reported as 25 seconds, however [23] shows that using the power of GPU and a compromise in performance can lead this algorithm to perform in less than one second. Therefore, for a real-time augmented reality application the total execution time (4 seconds) can be drastically drop by carefully revising the steps.

We note that the importance score introduced in the optimization function does not only improve light source position estimation but also it reduces the time consumed for error minimization. The error minimization for the proposed method takes $344.7ms$ whereas error minimization takes $8351.3ms$ for [7] (using multi-thread).

**Number of Training Samples:** The proposed $RGB-D$ video dataset is the largest light source position estimation dataset. However, the number of frames in the dataset can still be considered as small. The proposed method would learn ranking surfaces better by increasing the number of samples and varying the surface attributes in training. Moreover, larger number of video frames would also help to minimize the error of individual frame errors while using temporally constrained LSP estimation. Therefore, our method would additionally benefit from datasets with larger number of images.

On the other hand, not only number of frames but the variety of surface attributes would help our algorithm. Our surface attribute weighting method would benefit from more complex scenes (e.g. surfaces with photometric changes which do not obey lambertian reflectance rule and surfaces with more geometric variations). Moreover, camera pose estimation would also be more accurate in a more complex scene (i.e. ICP algorithm to estimate camera pose would benefit from surface variations). Therefore, a complex scene would also help to improve our results.

| Operation | Time Consumed) |
|---|---|
| Point Cloud Generation | 3ms |
| Normal Estimation | 42ms |
| Surface Segmentation | 195.8ms |
| Light Source Position Initialization | 1.07ms |
| Feature Extraction | 3016ms |
| Importance Score Prediction | 0.127ms |
| Error minimization | 344.7ms |
| Total | 3602.697ms |

TABLE VI
DIFFERENT STEPS OF THE PROPOSED METHOD AND THEIR TIME CONSUMPTIONS.

## VI. CONCLUSION

In this paper, we have exploited the influence of surface attributes on the accuracy of LSP estimation. Given a single $RGB-D$ image, we first analyzed the effects of photometric and geometric surface attributes. Then, surfaces are ranked using a supervised learning scheme. The ranking results are used to decide the contribution of an image surface segment for LSP estimation. Higher importance is assigned to those image surface segments which have suitable photometric (i.e. Lambertian reflectance) and geometric surface attributes. To speed up the LSP estimation, a geometry constrain has been introduced to initialize point selection. Moreover, the image surface segments which have a negative influence on LSP estimation are filtered out. Additionally, we introduce a temporal constraint to estimate LSP from a $RGB-D$ video sequence. LSP is optimized using the camera poses between successive frames. The results show that our method based on weighting image surface segments using their attributes outperforms the state-of-the-art methods. By using the proposed surface weighting, the angular error is reduced from 12.6° to 8.2° and 24.6° to 8.5° for *Boom* and our newly collected datasets respectively. Moreover, using the camera pose to temporally constrain LSP provides higher accuracy (4.8°) compared to using single frames (8.5°).

## ACKNOWLEDGMENT

This publication was supported by the Dutch national program COMMIT and European Union's Horizon 2020 project TrimBot2020 (grant No.688007).

## REFERENCES

[1] B. K. P. Horn and M. J. Brooks, "Shape from shading," *Cambridge Massachusetts: MIT Press*, 1989.
[2] D. Simakov, D. Frolova, and R. Basri, "Dense shape reconstruction of a moving object under arbitrary, unknown lighting," *ICCV*, 2003.
[3] N. Joshi and D. J. Kriegman, "Shape from varying illumination and viewpoint," *ICCV*, 2007.
[4] R. Zhang, P. sing Tsai, J. E. Cryer, and M. Shah, "Shape from shading: A survey," *TPAMI*, vol. 21, pp. 690–706, 1999.

[5] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *International journal of computer vision*, vol. 43, no. 1, pp. 29–44, 2001.

[6] A. Faisman and M. S. Langer, "How does lighting direction affect shape perception of glossy and matte surfaces?" in *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2013.

[7] B. Boom, S. Orts-Escolano, X. Ning, S. McDonagh, P. Sandilands, and R. Fisher, "Point light source estimation based on scenes recorded by a rgb-d camera," *BMVC*, 2013.

[8] Y. Wang and D. Samaras, "Estimation of multiple illuminants from a single image of arbitrary known geometry," in *ECCV*, 2002.

[9] A. O. Bălan, M. J. Black, H. Haussecker, and L. Sigal, "Shining a light on human pose: On shadows, shading and the estimation of pose and shape," in *ICCV*, 2007.

[10] Q. Zheng and R. Chellappa, "Estimation of illuminant direction, albedo, and shape from shading," in *CVPR*, 1991.

[11] I. Sato, Y. Sato, and K. Ikeuchi, "Illumination from shadows," *TPAMI*, 2003.

[12] J.-F. Lalonde, A. A. Efros, and S. G. Narasimhan, "Estimating the natural illumination conditions from a single outdoor image," *International Journal of Computer Vision*, vol. 98, no. 2, pp. 123–145, 2011.

[13] C. B. Madsen and B. B. La, "Probeless illuminantion estimation for outdoor augmented reality," *INTECH*, 2010.

[14] L. Gruber, T. Richter-Trummer, and D. Schmalstieg, "Real-time photometric registration from arbitrary geometry," in *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 119–128.

[15] N. Neverova, D. Muselet, and A. Trémeau, "Lighting estimation in indoor environments from low-quality images," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*, 2012, pp. 380–389.

[16] Z. S. Jiang, S. Rezvankhah, and K. Siddiqi, "Project report: Light source estimation using kinect," 2013.

[17] L.-F. Yu, S.-K. Yeung, Y.-W. Tai, and S. Lin, "Shading-based shape refinement of rgb-d images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1415–1422.

[18] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004.

[19] J. michael Frahm, K. Koeser, D. Grest, and R. Koch, "Markerless augmented reality with light source estimation for direct illumination," in *In Conference on Visual Media Production CVMP*, 2005.

[20] K. Hara, K. Nishino *et al.*, "Light source position and reflectance estimation from a single view without the distant illumination assumption," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 4, pp. 493–505, 2005.

[21] P.-E. Buteau and H. Saito, "[poster] retrieving lights positions using plane segmentation with diffuse illumination reinforced with specular component," in *Mixed and Augmented Reality (ISMAR), 2015 IEEE International Symposium on*. IEEE, 2015, pp. 202–203.

[22] Y. Ogura, T. Ikeda, F. De Sorbier, and H. Saito, "Illumination estimation and relighting using an rgb-d camera." in *VISAPP (2)*, 2015, pp. 305–312.

[23] B. J. Boom, S. Orts-Escolano, X. X. Ning, S. McDonagh, P. Sandilands, and R. B. Fisher, "Interactive light source position estimation for augmented reality with an rgb-d camera," *Computer Animation and Virtual Worlds*, 2015.

[24] J. T. Barron and J. Malik, "Intrinsic scene properties from a single rgb-d image," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 17–24.

[25] Q. Chen and V. Koltun, "A simple model for intrinsic image decomposition with depth cues," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 241–248.

[26] R. B. Rusu and S. Cousins, "3D is here: Point Cloud Library (PCL)," in *ICRA*, 2011.

[27] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE TPAMI*, 2010.

[28] A. Gijsenij, T. Gevers, and J. Van De Weijer, "Improving color constancy by photometric edge weighting," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 5, pp. 918–929, 2012.

[29] I. Everts, J. van Gemert, and T. Gevers, "Per-patch descriptor selection using surface and scene properties," in *Computer Vision ECCV*, 2012.

[30] R. O. Dror, T. K. Leung, E. H. Adelson, and A. S. Willsky, "Statistics of real-world illumination," in *CVPR*, 2001.

[31] V. Yanulevskaya and J. M. Geusebroek, "Significance of the weibull distribution and its sub-models in natural image statistics," in *International Conference on Computer Vision Theory and Applications*, 2009.

[32] L. Sharan, Y. Li, I. Motoyoshi, S. Nishida, and E. H. Adelson, "Image statistics for surface reflectance perception," *Journal of the Optical Society of America*, 2007.

[33] I. Motoyoshi, S. Nishida, L. Sharan, and E. H. Adelson, "Image statistics and the perception of surface qualities," *Nature*, 2007.

[34] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a bayesian framework for material recognition." in *CVPR*, 2010.

[35] D. Hu, L. Bo, and X. Ren, "Toward robust material recognition for everyday objects," in *BMVC*, 2011.

[36] T.-Y. Liu, *Learning to rank for information retrieval*. Springer Science & Business Media, 2011.

[37] T. Joachims, "Optimizing search engines using clickthrough data," in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2002.

[38] S. Karaoglu, Y. Liu, and T. Gevers, "Detect2rank: Combining object detectors using learning to rank," *Image Processing, IEEE Transactions on*, vol. 25, no. 1, pp. 233–248, 2016.

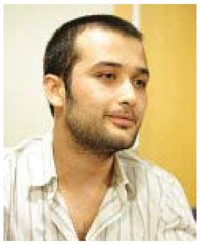
[39] J. A. Nelder and R. Mead, "A simplex method for function minimization," *The computer journal*, vol. 7, no. 4, pp. 308–313, 1965.

[40] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, and A. Fitzgibbon, "Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth." ACM Symposium on User Interface Software and Technology, 2011.

[41] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Mixed and augmented reality (ISMAR)*, 2011.

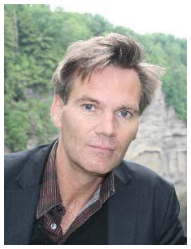
[42] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "Liblinear: A library for large linear classification," *JMLR*, 2008.

[43] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.



**Sezer Karaoglu** received his Ph.D. degree at Computer Vision Group, Informatics Institute, University of Amsterdam. He was selected as tuition fee scholar for a European Master degree from Color in Informatics and Media Technology (CIMET) program. He holds double master degree. His research interests are 2D-3D Computer Vision and Machine Learning. He is the CTO and Co-founder of 3DUniversum, spin-offs of the Informatics Institute of the UvA.

**Yang Liu** received the B.S. degree and Ph.D degree in Dept. of Computer Science of Shandong University in 2008 and 2013, respectively. His research interests are in the areas of Scene Understanding, Color Constancy and 3D Scene Reconstruction.



**Theo Gevers** Theo Gevers is a Full Professor of Computer Vision with the University of Amsterdam (UvA), Amsterdam, The Netherlands. His main research interests are in the fundamentals of image understanding, 3-D object recognition, and color in computer vision. He is the Founder of Sightcorp and 3DUniversum, spin-offs of the Informatics Institute of the UvA.

**Arnold W.M. Smeulders** is in charge of COMMIT/, a nation-wide, very large public-private research program distributed over the Netherlands on large-scale data, content, sensing and interaction. And he is professor at the University of Amsterdam UvA for research in the theory and practice of computer vision. The groups search engines have received a top-three performance for all 14 years in the international TREC-vid competition for image categorisation. He was recipient of a Fulbright fellowship at Yale University, and visiting professor in Hong Kong, Tuskuba, Modena, Cagliari and Florida. He was co-founder of Euvision Technologies BV, a company spin-off from the UvA. He is currently director of the Qualcomm - UvA and the Bosch - UvA labs. He is associate editor of the IJCV. He is fellow of the International Association of Pattern Recognition and elected member of the Academia Europaea (AE).