

Surface Light Field from Video Acquired in Uncontrolled Settings

Gianpaolo Palma, Nicola Desogus, Paolo Cignoni, Roberto Scopigno
Visual Computing Lab, CNR-ISTI, Pisa, Italy
Email: firstname.familyname@isti.cnr.it

Abstract—This paper presents an algorithm for the estimation of the Surface Light Field using video sequences acquired moving the camera around the object. Unlike other state of the art methods, it does not require a uniform sampling density of the view directions, but it is able to build an approximation of the Surface Light Field starting from a biased video acquisition: dense along the camera path and completely missing in the other directions. The main idea is to separate the estimation of two components: the diffuse color, computed using statistical operations that allow the estimation of a rough approximation of the direction of the main light sources in the acquisition environment; the other residual Surface Light Field effects, modeled as linear combination of spherical functions. From qualitative and numerical evaluations, the final rendering results show a high fidelity and similarity with the input video frames, without ringing and banding effects.

I. INTRODUCTION

The photo-realistic rendering of the appearance of real-world objects is a challenging problem in Computer Graphics. Especially in the Cultural Heritage field, the acquisition of the appearance is fundamental: the characteristic of the material, the reflectance behavior and the texture of the artwork offer major perceptual and cognitive hints with respect to the visualization of the plain 3D geometry. On the other hand, while for the acquisition of the 3D geometry there exist several solutions that allow to obtain accurate and high resolution 3D model in a simple and fast way, the appearance estimation is still an open problem due to the reflectance characteristics of the object and the constraints imposed by the acquisition environment. A Cultural Heritage artwork can be composed by different materials, which could present several types of patinas and degradation processes that alter its appearance, making difficult the modeling of the reflectance using mathematical reflection models. Usually artworks cannot be moved in an acquisition lab and the lighting environment cannot be substantially modified, imposing to perform on-the-field acquisition without the possibility to used controlled lighting conditions, like a dark room. Even if solutions have been proposed to estimate the Spatially Varying BRDF in these challenging conditions ([1]), in some applications it is sufficient to estimate a further simplification of the appearance in a more automatic way. An example is a virtual museum, where, in order to reproduce a real visit, the artwork can be viewed by different positions but with fixed lighting conditions. In this context, the *Light Field Rendering* [2] and its extensions are appropriate solutions. One of the major benefits of these techniques is the ability to create realistic renderings of a wide range of physical surfaces (including anisotropic ones) with complex reflectance behavior, without passing through the complex definition and

choice of intricate reflection models. Nevertheless, they have some drawbacks in term of data complexity if we need to increase the accuracy and the quality of the final rendering, due to the huge amount of data to acquire. For this reason hybrid solutions have been proposed. The main idea is to combine acquired photos with some geometric information to synthesize new views of the scene (*Lumigraph* [3] and *Surface Light Field* [4]). An important issue in the image-based and the hybrid methods for the Light Field (LF) estimation is the sampling density of the view directions. Many state of the art algorithms require a dense and uniform photographic acquisition of the object, obtained for example with special devices like camera arrays or robotic arms. However, in order to make the method usable in the Cultural Heritage context, we need an automatic system that is able to reconstruct the appearance from partial and irregular acquisitions, reducing the expertise of the operator needed to evaluate if the acquired data are sufficient. Acquired data are often partial because of the need to perform a fast sampling (not supported by a careful planning) or due the physical limitations and constraints of the digitization environment, like the presence of obstacles and occluders, or the challenging position and size of the artwork, which limit the free movement of the camera around the object and make hard to capture a complete sampling.

We propose a method for the estimation of the Surface Light Field (SLF) of a real-world object, starting from a medium quality 3D model and some video sequences with an irregular sampling of the view direction. The typical video acquisition is composed by very simple movements around the object, returning a view sampling biased by the camera path: very dense only along the camera path and completely missing in the other directions (see Figure 1). The algorithm uses the idea of the Dichromatic Reflection Model [5] to separate the estimation of the diffuse color from the estimation of the other surface appearance effects. For the diffuse color, we use a statistical method, based on the selection of the samples with a higher probability to have a Lambertian behavior, while the residual color, that is the difference from the diffuse color, is employed to fit a linear interpolation of spherical functions (Spherical and Hemispherical Harmonics).

Main contributions of our work are:

- a method to estimate the Surface Light Field without banding and ringing artifacts due to the spherical function fitting and the distribution of the input color samples in the visible hemisphere biased by the simple camera path (see Figure 2);
- a method to extract the Lambertian shading of the

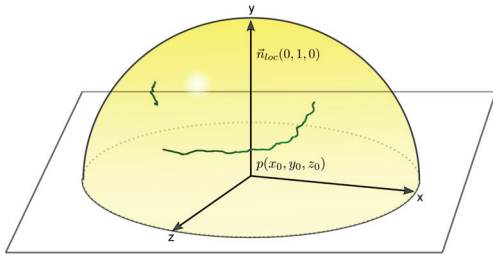


Fig. 1. Typical distribution of the acquired viewpoints (green line) in the visible hemisphere for a single surface point

object using statistical operations, which simplifies the modeling of the other Surface Light Field effects and takes advantage from a rough estimation of the acquisition environment map.

Even if the proposed method is not able to accurately predict the view dependent effects from not acquired viewpoint, due to the irregular acquisition, the main goal is to obtain realistic renderings without artifacts (Figure 2).



Fig. 2. Comparison of the finale SLF obtained from a viewpoint not acquired by the video sequences: (Left) rendering by trivial fitting of the input samples with the first 25 Spherical Harmonics functions; (Right) rendering of our algorithm. The trivial fitting produces artifacts, like ringing and banding effects, due to the irregular viewpoint sampling and the missing color samples in wide areas of the visible hemisphere.

II. RELATED WORK

In the Light Field and Surface Light Field literature, two different approaches have been used to solve the problem of a non uniform sampling of the view direction: hardware based and software solutions.

In the hardware based approach the proposed solutions use different type of special hardware devices that allow the acquisition of a uniform sampling of the view direction, like

computer controlled gantry with planar camera motion ([2]), camera arrays ([6], [7], [8], [9]), or camera systems with additional devices (microlens array [10] [11], attenuating mask [12]).

In the software approaches, the main idea is to use an algorithmic solution to fill the gaps and the holes coming out from a sparse and non-uniform directional acquisition. Gortler et al. [3] used a hand-held camera with a special designed stage for camera pose estimation. They rearranged the acquired samples in a two-plane parameterization and used a pull-push algorithm to fill the gaps on each plane due to an irregular distribution of the acquired views. Heigl et al. [13] started from an image sequence taken by a hand-held camera and relaxed the restrictions imposed by the regular LF structure. They rendered new views directly from the calibrated images with the use of a local depth map, mapping directly the original images onto one or more planes viewed by a virtual camera. Isaksen et al. [14] proposed a new parameterization, based on a general mathematical formulation for a planar data camera array, which allows to represent moderately sampled LF with wide variations in depth. The analytical formulation of the minimum sampling rate for LF rendering is derived in [15] using the sampling theorem and the Fourier spectral analysis of the LF signals. Buehler et al. [16] proposed a generalization of two image-based rendering algorithms (LF Rendering and View-dependent Texture Mapping[17]) guided by the distribution of the camera and the quality of the geometry. Chen et al. [18] proposed the approximation of the LF data by partitioning it over the mesh vertices and factorizing each part into a small set of lower-dimensional functions. They rearranged the input data into a matrix and then computed an approximated factorization of the matrix with two different methods: Principal Component Analysis and Non-negative Matrix Factorization. In order to obtain a more precise factorization, they applied a resampling of the data by Delaunay triangulation of the original views in the xy plane of the local reference frame of the vertex. This resampling step requires a good distribution of the original view in the local reference frame, a condition that our input videos do not guarantee. Davis et al. [19] presented a system for interactively acquiring and rendering a LF using a hand-held commodity camera. They computed and provided a coverage map to the user in real-time to show the views captured so far and to help the achieving of a dense coverage. A visual feedback mechanism for capture guidance is used also by Jachnik et al. [20]. They proposed a real-time system for the capture of the SLF of planar surface based on the splitting of the diffuse and specular components. The method has two limitations: it assumes an uniform specular behavior on the whole surface, modeled as a Phong lobe, which cannot be easily extended to objects with a more complex reflectance; it does not take advantage of the approximation of the environment map in the estimation of the SLF. The separation of the diffuse and specular components is also proposed by Wood et al. [21]. To generate a more compact SLF, they proposed two methods to represent each lumisphere as a weighted sum of a small number of prototype lumispheres. To improve the compression performance they applied a median removal algorithm that allowed to store the median texture map and to encode the residual separately. It should be noted that these approaches cannot be generalized for irregular sampling density. They

require some prototype lumispheres with a good and uniform sampling density that are difficult to obtain with a simple camera path.

We want to remark that, even if the amount of our input data (about 2000-3000 video frames) is bigger than the previous methods, these data can be very irregularly distributed in the visible hemisphere for each surface points (see Figure 1 for an example), due to few careful acquisition or to the physical limitations imposed by the digitization environment, a condition that the previous software solutions cannot manage and resolve without decreasing the quality of the final rendering.

III. BACKGROUND

The SLF was introduced in [4] to sample LF on a parametric surface directly. It is modeled as a four-dimensional function $L(u, v, s, t)$, where (u, v) defines a point on the surface and (s, t) represents the orientation angles identifying the view direction. Given a novel camera position, the rendering is done by computing the view direction values (s, t) for each surface point (u, v) and then indexing the SLF using (u, v, s, t) to extract the color value. The use of the geometry of the object permits to overcome some of the limitations of a generic LF representation: the camera is restricted to certain regions of space; the finite angular resolution leads to depth of field effects; objects become blurry in proportion to their distance from the image plane.

Our method is based on the separation of the diffuse component of the surface appearance from the other view dependent lighting effects. This separation process allows to avoid rendering artifact due to the fitting and interpolation process of the spherical functions. This idea is mathematically supported by the Dichromatic Reflection Model [5]. Intuitively, this model states that there are two independent reflection processes, the specular and the diffuse reflectance, each one characterized by a color whose magnitude varies with the light \vec{l} and view \vec{v} directions:

$$C(\vec{l}, \vec{v}) = m_d(\vec{l})C_d + m_s(\vec{l}, \vec{v})C_s \quad (1)$$

where C is the final RGB color, C_d and C_s the color of the two components, and m_d and m_s the magnitude of the two components. However, in the context of the SLF, since we are under fixed lighting conditions and we assume to have a Lambertian diffuse component, we can simplify Equation 1, that becomes:

$$C(\vec{v}) = C_d + m_s(\vec{v})C_s \quad (2)$$

IV. ALGORITHM

Our algorithm takes in input a medium-resolution triangular mesh of the object, with an associated texture parameterization, and some video sequences acquired moving the camera around the object. The target is to estimate the diffuse color and the other view dependent SLF effects in two different steps. In the first one, we estimate the diffuse component using statistical operations. In the second step, we reconstruct the other SLF effects, as linear combination of spherical functions. The final color of a point p with texture coordinates u, v is given by the following formula:

$$C(u, v, s, t) = D^{(u,v)} + \sum_{i=0}^n x_i^{(u,v)} h_i(s, t) \quad (3)$$

where (s, t) are the spherical coordinate (θ, ϕ) of the view vector in the local reference frame of p , $D^{(u,v)}$ is the diffuse texture color of the point, $x_i^{(u,v)}$ are the coefficients that are associated to the selected basis of spherical functions $h_i(s, t)$. The outputs of the algorithm are a texture map with the diffuse color $D^{(u,v)}$ and a binary file with the coefficients $x_i^{(u,v)}$, organized as an OpenGL texture array where the i -th layers contains the coefficient $x_i^{(u,v)}$. The estimation process is organized in four steps:

- 1) the video-to-geometry registration (Section IV-A);
- 2) the estimation of the direction of the main light sources in the acquisition environment (Section IV-B);
- 3) the approximation of the surface diffuse color (Section IV-C);
- 4) the fitting of the view dependent SLF (as residual from the diffuse color) in a set of basis spherical functions (Section IV-D).

A. Video-to-Geometry Registration

Using the method proposed in [22], the intrinsic and extrinsic camera parameters are recovered for each video frame. The result of the registration process allows to compute the color samples $C_{u,v} = \{I_{u,v}^{(j)} \in RGB\}$ and the relative quality value $q_{u,v}^{(j)}$ projected by each frame j for each texel (u, v) . The quality value is computed using the framework proposed in [23]. It is equal to the product of three measures: the distance in image space from the nearest depth discontinuity, to penalize possible wrong color samples due to small misalignment; the depth of the texel in camera space, to give a higher quality at the samples acquired by closer views; the view angle, to give a higher quality at the samples that are less orthogonal with respect to the view direction.

B. Lighting Estimation

In this step, the main goal is the approximation of the direction of the main light sources that were in the acquisition lighting environment, assuming distant lights. We start with the construction of a rough environment map using a simple accumulation approach along the specular mirror direction, followed by a clustering method to detect the main light sources.

In the specific, for each texel (u, v) independently, we select all the projected color samples that have a higher probability to show a not diffuse behavior. We use a trivial approach based on luminance thresholding, which selects all the color samples near or in the saturation area of the camera sensor. In the specific we select all the samples with a luminance greater than a fixed threshold $t = 0.98$ (with a luminance range $[0.0, 1.0]$). For all those samples, we compute the specular mirror direction \vec{r} of the view vector \vec{v} , and then we increment the value of the pixel in the environment map where the spherical coordinates of the vector \vec{r} are mapped. Then, the map is normalized with the distribution of all color samples in the environment along the specular mirror direction. This distribution is computed as the total number of samples that project in each specific pixel of the environment map along the direction \vec{r} . This normalization gives more robustness with

respect to the camera movement, especially when the temporal density of the acquisition is not uniform on the whole object (for example very fast in front of the object and more slow in the back). Finally, the environment map is normalized in the range $[0, 1]$ with respect to the maximum value. Even if the trivial luminance thresholding with a fixed threshold is prone to introduce noise in the estimation of the environment map, especially for white object, the high redundancy of the video data makes the estimation process more robust and stable.

An example of the estimated environment map, with a comparison with the real one acquired with a reflective metal sphere, is shown in Figure 3. The map gives us some clues about the position of the main light sources in the scene, creating several clusters that correspond to the main real light sources. To detect the clusters we use a K-Mean clustering where the number of clusters can be selected by the user after the visualization of the estimated map. Finally we compute the centroid of each cluster to obtain an estimation of the light direction \vec{l}_k .



Fig. 3. Comparison of our approximation of the acquisition environment map (Right) with the real one obtained with a photo of a reflective sphere (Left).

C. Diffuse Color Approximation

In this phase we use the light directions computed in the Section IV-B to make more robust the estimation of the diffuse color for each texel. The computation is performed in three steps. The main idea is to take advantage of the high amount of color samples projected on each texel to discard all the ones that have a higher probability to introduce inaccuracy in the computation of the diffuse color, such as specular or wrong projected samples near color discontinuities.

In the first step, color samples are sorted by their luminance value to discard a percentage p_t of them having the lowest luminance. In our experiments we use $p_t = 15\%$. This step allows to increase the local coherence near the color boundaries, avoiding abrupt changes in luminance due to wrong projected color. In the second step, we discard all the color samples with a higher probability to have a view-dependent reflectance behavior, especially specular reflection. Knowing that a specularly occurs when the view vector is aligned with the specular mirror direction of the light vector, the idea is to determine which samples appear close to this direction and discard them. For this purpose, each light \vec{l}_k is reflected with respect to the surface normal \vec{n} to compute the mirror direction \vec{s}_k and then we discard the color samples such that $\arccos(\vec{v} \cdot \vec{s}_k) < \theta_t$, where \vec{v} is the sample view direction,

as depicted in Figure 4. We use a threshold $\theta_t = \pi/12$. The purpose of thresholds p_t and θ_t is not to discard all the samples with bad properties for the estimation of the diffuse color, but to decrease the number of them in order to reduce their influence in the following statistical computation of the diffuse color. They are chosen to work in the most general cases.

At this point, the set $\tilde{S}_{u,v}$ of the remaining samples is used to compute a threshold $t_{u,v}$, which is an upper bound of the median:

$$t_{u,v} = \mu_{u,v} + \sigma_{u,v}^2 \quad (4)$$

where $\mu_{u,v}$ and $\sigma_{u,v}$ are the mean and the average deviation of the luminance of the samples in $\tilde{S}_{u,v}$, weighted with the quality $q_{u,v}^{(j)}$. Finally, the diffuse color $D_{u,v}$ is computed as weighted average with the quality $q_{u,v}^{(j)}$ of the color samples that have a luminance value lower than the threshold $t_{u,v}$.

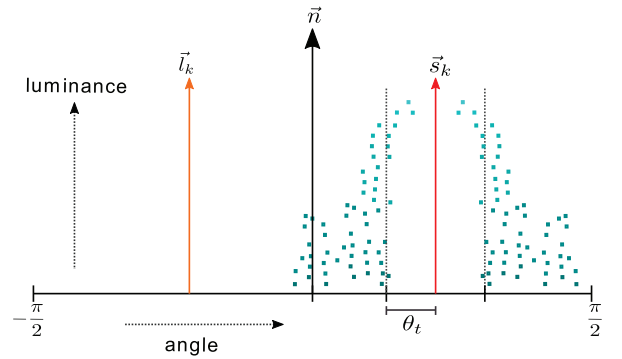


Fig. 4. Graph of the distribution of the samples with respect to the angle between the view direction and the surface normal. All the samples with an angle from the specular mirror direction \vec{s}_k of the light vector \vec{l}_k less than the threshold θ_t are discarded in the estimation of the diffuse color

D. Color Residual Fitting

The idea is to use the residual images, obtained as positive difference of the color samples from the diffuse color, to compute the coefficients x_i of a linear combination of a basis of spherical functions h_i , which just depend on the local spherical coordinates (s, t) of the view direction:

$$\sum_{i=0}^n x_i^{(u,v)} h_i(s, t) \quad (5)$$

We tested two different basis of spherical functions: Spherical Harmonics and Hemispherical Harmonics [24]. Using the Dichromatic Reflection Model, we assume a white light and we model the color residual using only the luminance difference between the diffuse color and the input color samples.

For each texel (u, v) we retrieve the set of color samples $S_{u,v}$ that have a positive residual from the diffuse color and we solve a system of linear equations:

$$Ax = \mathbf{b} \quad (6)$$

where A is an $m \times n$ matrix that for each row, one for each sample in $S_{u,v}$, contains the values of the spherical functions computed for the view direction of the sample, \mathbf{x} is the vector of the n coefficients x_i to estimate, and \mathbf{b} is the

vector with the luminance difference from the diffuse color. To solve the overdetermined system in Equation 6 we use a Weighted Singular Value Decomposition (SVD) in order to take advantage of the quality information $q_{u,v}^{(j)}$ related to each sample. In this way, we compute a weighted least square solution of the system:

$$\min_{\{x_1, \dots, x_n\}} \left(\sum_{j=1}^m q^{(j)} r_j^2 \right) \quad (7)$$

where

$$r_j = \left(\sum_{i=0}^n x_i h_i(s_j, t_j) \right) - b_j \quad (8)$$

The weighted least square solution allows to reduce the influence of a combination of misalignment in the video-to-geometry registration and artifacts in the video frames, which can alter the realism of the result. In general the samples in $S_{u,v}$ used to solve the equation 7, cover only a small part of the visible hemisphere. To avoid that the fitting procedure creates artifacts (banding and ringing effects) in the not sampled areas, we add some virtual samples, uniformly distributed in the uncovered regions, with a residual color equal to zero (at most 400 samples distributed with a Poisson-Disk strategies with respect to the existing samples). Finally, to reduce the GPU memory footprint need for the rendering of the SLF, we compress the floating-point coefficients x_i with a simple 8-bit quantization.

V. RESULTS

For our experiments, we used three different objects of different materials:

- the DWARF, a terracotta statue (30cm tall) that presents different types of specularity, in size and intensity: sharper and with a high-medium intensity on the dress; wider on the face; almost completely absent on the beard;
- the GNOME, a ceramic statue (15cm tall) that has very sharp and high specularity on the hat and a near diffuse behavior on the body;
- the SLEEPING BUDDHA, an acrylic resin Buddha (10cm tall) with different types of coatings: a gold paint on the body; a reddish specular paint on the dress; a diffuse black paint on the hair.

The videos were acquired with a full HD video camera at the highest quality to reduce the compression artifacts, while the 3D models were generated by 3D laser scanning and then simplified to obtain a medium resolution model for the computation of the texture parameterization. Today 3D scanning is a mature technology that guarantees the generation of precise and accurate models in short time (both for the acquisition and the processing of the data), in an easy and completely automatic way and with low cost hardware (e.g. consider the NextEngine 3D Scanner HD). For each object we used a 2048×2048 texture. All the data about the datasets are shown in Table I (the size in triangles of the 3D model, the time for the generation of the medium resolution 3D model, the length in frames of the videos used for the estimation of

the SLF, the time required for the alignment of the video on the mesh using the method in [22], the time required for the computation of our approximation of the SLF). We performed our test on a PC with an Intel Core i7 950 with 12GB of RAM and a NVIDIA GTX580 1536MB.

The figures 5, 6 and 7 show a comparison of the results of our method with an original frame of the video used by the algorithm. In the specific for each figure, we show the original frame (a), our estimation of the diffuse color (c), and our estimation of the Surface Light Field using two different basis of spherical functions: Hemispherical Harmonics (HSH) in the sub-figures (e)(f)(g); Spherical Harmonics (SH) in the sub-figures (h)(i)(l). In SLF estimation we tested an increasing number of coefficients (4, 9 and 16 coefficients for HSH and 9, 16 and 25 coefficients for SH) up to a number that allows the real-time rendering (above 25 fps) on the last common GPUs. From a qualitative evaluation of these results, we can say that: the method proposed in the Section IV-C estimates a good approximation of the Lambertian shading of the object without artifacts and discontinuities, removing all the other reflection effects, like the specularities (compare the (a) and (c) in Figure 5, 6 and 7); the final SLF has a good similarity with the original frame, a similarity that increases with the number of spherical functions that are used.

To have a more objective evaluation of our results (Figures 5, 6 and 7), we used two different metrics to measure the image fidelity between the original frames and the obtained renderings: the Mean Squared Error (MSE); the Structural SIMilarity (SSIM) index [25], a perceived fidelity measure (the value is a percentage and the value 100% is returned when we compare an image with itself). Table II contains the value of the two metrics for our test cases. We included as well the error between the original images and the renderings with only the diffuse component (sub-figure (c)), in order to highlight the improvement introduced by the term that models the residual color. From the data in the table we can see how the accuracy and fidelity of the SLF increase with the number of used spherical functions and, given a number of coefficients, the HSH representation outperforms the SH representation.

The renderings show some small differences from the reference frames. In the diffuse color, we lost some of the very small details due to small misalignment in the video-to-geometry registration. Some highlights appear different from the reference frame due mainly to imprecise normals (for example on the hat of the GNOME). Our method does not reproduce some specular highlights due to Fresnel effects (for example on the nose and on the top-right part of the face of the DWARF) because this type of effect appears near to the silhouette of the object, where our algorithm gives a lower quality at the samples. Finally the highlights appear less bright than the original ones in the original frames due to the limited number of functions (at most 25 functions for SH and 16 for HSH) used to model the residual color. To be more specific, the limited number of spherical functions used does not allow to reproduce the narrower specularity peaks, obtaining a band-limited reconstruction of the specularity. This aspect is confirmed by the higher rendering quality obtained by increasing the number of functions. The use of a higher number of spherical functions can surely remove this type of imperfections but at the cost of lower rendering performance.

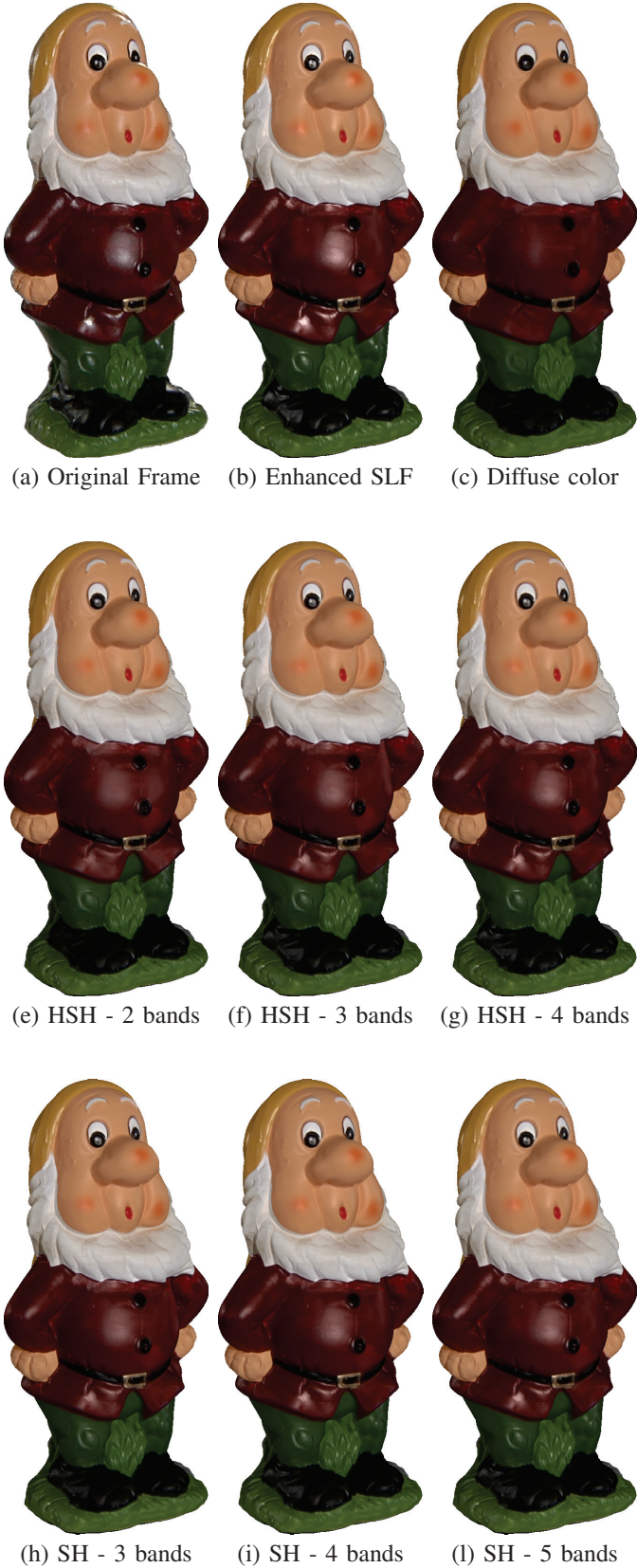


Fig. 5. DWARF results

In order to guarantee the real-time rendering of our SLF representation, we can obtain a further improvement of the visual results by introducing a small change in the rendering Equation 3, based on the enhancement of the residual component. We introduce a new term I_s , a user-defined parameter to change the intensity of the residual component, used in the following manner:

$$C(u, v, s, t) = D^{(u,v)} + I_s \sum_{i=0}^n x_i^{(u,v)} h_i(s, t) \quad (9)$$

The rendering results with this new parameter I_s are shown in the subfigure (b) of the figures 5, 6 and 7, using the SLF estimated using 5 bands of SH (subfigure (l)). The values assign to the parameter I_s are 1.6 for the GNOME and the DWARF and 1.2 for the SLEPPING BUDDHA. We can note how the visual differences between the rendering (b) and the original frame (a) are further reduced. This observation is confirmed from the estimation of the MSE and SSIM metric between the two images, reported in the last row of the Table II. The video supplemental material shows a real-time interaction with the our SLF approximation. The results are realistic and without artifacts.

VI. CONCLUSIONS

We have presented a new method for the estimation of the Surface Light Field starting from a 3D model and some video sequences, acquired moving the camera around the object. The video sequences do not guarantee a uniform sampling density of the view direction. In order to avoid interpolation artifacts due to the very irregular video acquisition, characterized by a dense coverage sampling only along the camera path, we separate the SLF in two components: the diffuse color and the residual from the diffuse color. The first component is estimated using statistical operations that take advantage of the data redundancy of the video sequences. The main idea is to discard all the color samples that have a higher probability to exhibit a not diffuse behavior. We roughly estimate the direction of the main light sources by accumulation of the saturated samples along the specular mirror directions and then we delete all the samples with an angle between the view direction and the mirror direction of the light vector above a fixed threshold. The second component models the color residual, which is the difference from the diffuse color, as linear combination of a basis of spherical functions. The results on our test cases do not present artifact due to the interpolation and fitting of the spherical functions and the similarity measures proof a high fidelity degree between the renderings with our estimation of SLF and the original video frames used by the algorithm. Finally, we have introduced a small change in the rendering formula to allow the enhancement of the residual component in order to overcome the limitation of a band-limited fitting and, at the same time, preserve the real-time visualization of the model.

For the future work, we will investigate three possible directions: the study of the robustness of the algorithm with respect to the temporal sampling, evaluating the rendering quality obtained with a SLF estimated with a temporal down-sampled version of the input video sequences; a method to estimate the best parameter I_s in the Equation 9 automatically, using similarity metrics computed between the SLF rendering

	Geometry (triangles)	Mesh Processing (minutes)	Frames	Registration time (minutes)	SLF estimation (minutes)
DWARF	200k	55	3382	113	72
GNOME	135k	40	2092	73	61
SLEEPING BUDDHA	205k	70	2414	83	65

TABLE I. MODELS AND INPUT DATASETS CHARACTERIZATION

	DWARF		GNOME		SLEEPING BUDDHA	
	MSE	SSIM	MSE	SSIM	MSE	SSIM
Diffuse	0.003075	85.187%	0.003089	76.533%	0.005666	79.546%
HSH - 2 band	0.002027	87.169%	0.003059	76.879%	0.002728	83.788%
HSH - 3 band	0.001259	88.365%	0.002661	78.211%	0.001972	85.948%
HSH - 4 band	0.001020	89.040%	0.002543	79.024%	0.001538	87.322%
SH - 3 band	0.001317	87.905%	0.002695	78.099%	0.002237	84.402%
SH - 4 band	0.001066	88.478%	0.002581	78.673%	0.001788	85.650%
SH - 5 band	0.000932	88.818%	0.002515	79.232%	0.001635	86.178%
SH Enhanced- 5 band	0.000869	91.414%	0.002473	82.998%	0.001081	89.481%

TABLE II. ERROR MEASURES

and some peculiar frames of the input videos; the improvement of the rendering results of objects with a significant meso-structure by optimizing an additional bump map.

VII. ACKNOWLEDGMENTS

The research leading to these results was co-funded by EU FP7 project ICT FET “Harvest4D” (<http://www.harvest4d.org/>), G.A. no. 323567) and the UE FP7 project ICT Network of Excellence “V-MUST.net” (<http://www.v-must.net/>), G.A. no. 270404).

REFERENCES

- [1] G. Palma, M. Callieri, M. Dellepiane, and R. Scopigno, “A Statistical Method for SVBRDF Approximation from Video Sequences in General Lighting Conditions,” *Computer Graphics Forum*, vol. 31, no. 4, pp. 1491–1500, 2012.
- [2] M. Levoy and P. Hanrahan, “Light field rendering,” in *SIGGRAPH 96 Conference Proceedings*, Aug. 1996, pp. 31–42.
- [3] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, “The lumigraph,” in *SIGGRAPH 96 Conference Proceedings*, Aug. 1996, pp. 43–54.
- [4] G. Miller, S. Rubin, and D. Ponceleon, “Lazy decompression of surface light fields for precomputed global illumination,” in *Rendering Techniques '98*, 1998, pp. 281–292.
- [5] S. A. Shafer, “Using color to separate reflection components,” *Color Research and Application*, vol. 10, no. 4, pp. 210–218, 1985.
- [6] H. Schirmacher, L. Ming, and H.-P. Seidel, “On-the-Fly processing of generalized lumigraphs,” in *EUROGRAPHICS 2001 Conference Proceedings*, ser. Computer Graphics Forum, vol. 20, 3, Sep. 4–7 2001, pp. 165–173.
- [7] J. C. Yang, M. Everett, C. Buehler, and L. McMillan, “A real-time distributed light field camera,” in *13th Eurographics Workshop on Rendering*, 2002, pp. 77–86.
- [8] C. Zhang and T. Chen, “A self-reconfigurable camera array,” in *Eurographics Symposium on Rendering*, A. Keller and H. W. Jensen, Eds., 2004, pp. 243–254.
- [9] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, “High performance imaging using large camera arrays,” *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 765–776, Jul. 2005.
- [10] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, “Light field photography with a Hand-Held plenoptic camera,” Stanford University, Tech. Rep., Apr. 2005.
- [11] T. Georgiev and A. Lumsdaine, “Focused plenoptic camera and rendering,” *Journal of Electronic Imaging*, vol. 19, no. 2, 2010.
- [12] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, “Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing,” *ACM Transactions on Graphics*, vol. 26, no. 3, pp. 69:1–69:11, Jul. 2007.
- [13] B. Heigl, R. Koch, M. Pollefeys, J. Denzler, and L. J. V. Gool, “Plenoptic modeling and rendering from image sequences taken by hand-held camera,” in *DAGM-Symposium*, 1999, pp. 94–101.
- [14] A. Isaksen, L. McMillan, and S. J. Gortler, “Dynamically reparameterized light fields,” in *SIGGRAPH 2000 Conference Proceedings*, Jul. 23–28 2000, pp. 297–306.
- [15] J. Chai, X. Tong, S. Chan, and H. Shum, “Plenoptic sampling,” in *SIGGRAPH 2000 Conference Proceedings*, Jul. 23–28 2000, pp. 307–318.
- [16] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, “Unstructured lumigraph rendering,” in *SIGGRAPH 2001 Conference Proceedings*, 2001, pp. 425–432.
- [17] P. E. Debevec, Y. Yu, and G. D. Borshukov, “Efficient view-dependent image-based rendering with projective texture-mapping,” in *Rendering Techniques '98*, 1998, pp. 105–116.
- [18] W.-C. Chen, J.-Y. Bouguet, M. H. Chu, and R. Grzeszczuk, “Light field mapping: efficient representation and hardware rendering of surface light fields,” *ACM Trans. Graph.*, vol. 21, no. 3, pp. 447–456, Jul. 2002.
- [19] A. Davis, M. Levoy, and F. Durand, “Unstructured Light Fields,” *Computer Graphics Forum*, vol. 31, no. 2, pp. 305–314, 2012.
- [20] J. Jachnik, R. A. Newcombe, and A. J. Davison, “Real-time surface light-field capture for augmentation of planar specular surfaces,” in *ISMAR*. IEEE Computer Society, Nov. 2012, pp. 91–97.
- [21] D. N. Wood, D. I. Azuma, K. Aldinger, B. Curless, T. Duchamp, W. Stuetzle, and D. H. Salesin, “Surface light fields for 3D photography,” in *SIGGRAPH 2000 Conference Proceedings*, Jul. 23–28 2000, pp. 287–296.
- [22] G. Palma, M. Callieri, M. Dellepiane, M. Corsini, and R. Scopigno, “Geometry-aware video registration,” in *VMV*, 2010, pp. 107–114.
- [23] M. Callieri, P. Cignoni, M. Corsini, and R. Scopigno, “Masked photo blending: mapping dense photographic dataset on high-resolution 3d models,” *Computer & Graphics*, vol. 32, no. 4, pp. 464–473, Aug 2008.
- [24] P. Gautron, J. Krivánek, S. N. Pattanaik, and K. Bouatouch, “A novel hemispherical basis for accurate and efficient rendering,” in *Rendering Techniques 2004, Eurographics Symposium on Rendering*, June 2004, pp. 321–330.
- [25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transaction on Image Processing*, vol. 13, no. 4, pp. 600–612, Apr. 2004.



(a) Original Frame (b) Enhanced SLF (c) Diffuse color



(e) HSH - 2 bands (f) HSH - 3 bands (g) HSH - 4 bands



(h) SH - 3 bands (i) SH - 4 bands (l) SH - 5 bands

Fig. 6. GNOME results



(a) Original Frame (b) Enhanced SLF (c) Diffuse color



(e) HSH - 2 bands (f) HSH - 3 bands (g) HSH - 4 bands



(h) SH - 3 bands (i) SH - 4 bands (l) SH - 5 bands

Fig. 7. SLEEPING BUDDHA results