

Guided Robust Matte-Model Fitting for Accelerating Multi-light Reflectance Processing Techniques

Ruggero Pintus¹
ruggero.pintus@crs4.it

Andrea Giachetti²
andrea.giachetti@univr.it

Giovanni Pintore¹
giovanni.pintore@crs4.it

Enrico Gobbetti¹
enrico.gobbetti@crs4.it

¹ CRS4, Visual Computing Group
Pula (CA), Italy

² University of Verona, CS Dept.
Verona, Italy

Abstract

The generation of a basic matte model is at the core of many multi-light reflectance processing approaches, such as Photometric Stereo or Reflectance Transformation Imaging. To recover information on objects' shape and appearance, the matte model is used directly or combined with specialized methods for modeling high-frequency behaviors. Multivariate robust regression offers a general solution to reliably extract the matte component when source data is heavily contaminated by shadows, inter-reflections, specularity, or noise. However, robust multivariate modeling is usually very slow. In this paper, we accelerate robust fitting by drastically reducing the number of tested candidate solutions using a guided approach. Our method propagates already known solutions to nearby pixels using a similarity-driven flood-fill strategy, and exploits this knowledge to order possible candidate solutions and to determine convergence conditions. The method has been tested on objects with a variety of reflectance behaviors, showing state-of-the-art accuracy with respect to current solutions, and a significant speed-up without accuracy reduction with respect to multivariate robust regression.

1 Introduction

Multi-light reflectance acquisition and processing techniques, such as Photometric Stereo (PS) [47], Reflectance Transformation Imaging (RTI) [28], and Polynomial Texture Maps (PTM) [26], aim at characterizing surfaces by observing them from a fixed point of view under different lighting conditions.

The generation of a matte model from image measurements is at the core of many successful approaches, which use it directly, or augment it with specialized methods for modeling high-frequency components (see Sec. 2). Such recovered low-frequency information, typically obtained by fitting low-order polynomials or Hemispherical Harmonics (HSH) to the captured data, is used for a variety of applications, including relighting [15, 26], enhanced

visualization[25], normal/albedo extraction[10], as well as material classification and segmentation [9, 44].

Reliably fitting the matte model to measurements in the presence of non-Lambertian phenomena (e.g., shadows, anisotropic behaviors, gloss, specularly) is easily prone to errors, and a variety of methods, from heuristic trimming of outliers to robust fitting procedures, have been proposed (see Sec. 2). The most general solutions are based on multivariate robust regression [10, 54], which is able to automatically separate inlier and outlier measurements according to the fitted models, to identify matte and non-matte behavior, and to provide a more reliable diffuse interpolation at non-sampled locations due to reduced risks of over-fitting. However, the highest accuracy of robust algorithms comes at the price of increasing computational costs, due to the need to evaluate a large number of hypotheses. This makes those solutions impractical in most scenarios, especially as the number of parameters increases [53].

Our contribution harnesses the capacity of robust statistics to be resistant to highly contaminated data, but exploits a computational framework that provides a considerable speed-up while keeping the fitting error comparable to current slow robust estimators, by using a guided sampling approach to drastically reduce the number of computed fitting trials. We first show how the efficiency of a Least-Median of Squares (LMS) robust regression can be improved by providing a fitting tolerance threshold and by associating to each measurement a weight proportional to the a-priori likelihood that the measurement is part of the solution. This additional information provides a means for likely generating good candidate solutions early-on and for recognizing an optimal candidate solution as soon as it is generated. We then show how we can generate the required information for such a guided algorithm by propagating the information from already computed pixels using a similarity-driven region growing strategy implemented using multi-core parallelism.

While guided sampling approaches have been successfully proposed to speed-up computation in many Computer Vision areas, see Sec. 2, this is to the best of our knowledge the first time that an optimization of robust fitting is proposed for the specific case of parameter estimation from MLICs. In particular, we introduce in this context a specific guided sampling approach based on appearance profile similarity, as well as parallelizable techniques to propagate results from already known pixels to unknown ones. As demonstrated in Sec. 4, the proposed approach, evaluated both on synthetic and real-world data provides a significant speed-up without accuracy reduction, and can be employed as a building block in RTI, PTM or PS pipelines. Moreover, we show how generic robust fitting approaches (baseline or accelerated), without domain specific modifications, are generally on-par with specifically tailored state-of-the-art PS techniques, and offer the additional possibility of being applicable, without modification, to other estimation techniques, such as higher order polynomial fitting for PTM computation. While the method has been implemented, tested and validated by using a LMS solver, it is possible to apply our strategy to many other iterative robust regressions, such as M-estimators or iteratively re-weighted Least Squares (LS). Moreover, the algorithm is fully data-driven, and does not need user-defined, tuned parameters.

2 Related work

Computing surface properties from multi-light data is a very old and still very active research area [16, 25, 26, 28, 47]. We discuss here only the approaches strongly related to ours, and, in particular, works that aim at fitting a low-frequency model to reflectance measurements.

For a wider coverage, we refer the reader to established surveys in surface reflectance capture [39], multi-light computational frameworks [0], digital modeling of material appearance [9], and geometric analysis in cultural heritage[50].

The simplest technique is the classical Lambertian PS approach [27], which assumes image intensity proportional to the cosine between the surface normal and the light direction, and exploits it to recover albedo and normals for multiple images. PTM [26] is presented as a fitting-based generalization of the simplest PS, and extends the basic Lambertian model with a 6-term polynomial; two of its coefficients are the two components of the normal vector (Lambertian term), while the others try to capture the remaining low-frequency signals. A linear LS regression is used to fit the model. Subsequent works presented a wide range of follow-ups to RTI-based surface characterization [11, 17, 28, 46], and proposed alternatives to define and represent surface properties, such as spherical and hemispherical harmonics [29], bivariate Bernstein polynomials [19], bi-polynomial functions [35], and discrete modal decomposition [31]. Unfortunately, all those methods cannot cope well with the presence of non-diffuse phenomena such as gloss, highlights, specularity, or shadows. The underlying computational framework is strongly affected by those leverage points, and exhibits high error in the resulting interpolation, albedo and normal maps.

Many techniques have been proposed to detect/remove outliers from the fitting computation. The first attempts have been made in the field of Photometric Stereo. Since three inliers are enough to compute the normal maps in the diffuse case, four- or five-light setups are used to exclude the lights that generates significantly distorted outcome [3, 8, 33, 37]. In the case of N lights (i.e., N -dimensional appearance profile), Wenger *et al.* [45] heuristically eliminate the half darkest measurements and the brightest 10% before applying LS. Conversely, Willems *et al.* [46] gradually prune unreliable values through iterative normal estimations, while Yuille and Snow [52] rely on the integrability constraint. Monotonicity is another reflectance property exploited to deal with non-Lambertian diffuse behaviors [18, 34]. Some works exploit the difference between the light and object chromaticity, by exploiting the dichromatic reflection model [23, 50]. Component based modeling [27] is another more general way to separate reflectance terms by decomposing input into a local and global component. Others handle outliers within a rank minimization [48] or matrix factorization frameworks [20]. Recent techniques employ statistical models as labeling problems coupled with shadow graphs [9, 50], maximum-likelihood (ML) [43], or Markov random fields (MRFs) [40, 49], as well as non-parametric data-driven approximation of object reflectance [2]. Robust regression has proven to be a general powerful approach for RTI computational problems [10, 43, 53]. Ikehata *et al.* [20] present a hierarchical Bayesian approximation to deal with non-Lambertian corruptions, and to estimate reliable surface normals. Zhang and Drew [54] propose to use a LMS approach to solve for a 6-term polynomial diffuse component. Their robust method automatically discards outliers (up to 50% of input measurements), obtains a reliable interpolation, and retrieves a good estimation of albedo and normals. Once they compute a reliable matte component, they can isolate the remaining high-frequency signal, and model it with Radial Basis Functions (RBFs). Although some of the previous approaches can provide a good appearance representation, some important issues remain unsolved. The methods that throw away a fixed or proportional number of values with conservative heuristics do not ensure both the removal of leverage points or an adequate level of Signal-to-Noise Ratio (SNR) [3, 8, 33, 37, 45]. Others have complex rationales based on too constrained assumptions [58], initial guesses [43], or on a-priori knowledge [50]. Most of them have also to address the trade-off between accuracy and a heavy computational load [22, 53]. Previously proposed robust methods [10, 54] are very

general, since they do not impose any prior knowledge or heuristics about outlier distribution, but, since they achieve robustness and generality by blindly applying a classic robust fitting method [62] independently for each pixel, they require very high computational times even in parallel settings.

For this reason, we investigated possible strategies to reduce the computational complexity of robust fitting methods. Our solution follows the general idea of estimation by guided sampling, which has proven effective in many areas of Computer Vision, in particular for increasing the chance to pick inlier samples at each iteration of RANSAC-style estimators [6, 13, 27, 42]. The common idea underpinning guided sampling is to exploit prior knowledge to increase the density of samples in the regions where the correctness probability of the hypothesis is higher, rather than sampling the search space uniformly [44]. With an accelerated hypothesis generation [5], it is thus possible to reach a low residual error and an acceptable convergence within a reduced amount of iterations.

Our guided sampling solution introduces a speed-up strategy that harnesses spatial consistency, appearance profile similarity, and data-driven estimation of outlier distribution among neighboring pixels. To the best of our knowledge, this is the first algorithm that effectively exploits information from past pixels to drive robust per-pixel computation in the field of MLIC processing. Although tested with LMS, the proposed rationale is general enough to be easily adapted to increase computational efficiency of other, general iterative robust estimators applied to RTI processing pipeline. Moreover, while we validate our method by fitting polynomial bases, the theoretical underpinning of our solution can be generally used with other types of basis functions, e.g., HSH [14, 54].

3 Method

Multi-light reflectance data is acquired by taking from the same pixel-registered viewpoint a set of photographic images of an object, with each image illuminated from a different direction. As a result, a calibrated processing pipeline receives as input a so-called per-pixel appearance profile, i.e., a list of pairs that couple a measured value with a light direction and intensity. Surface shape and diffuse appearance can then be characterized for each pixel by converting this information to shape and appearance descriptors. In this work, we focus on matte model extraction, in which the set of multi-light measurements is fitted to an analytical low-frequency model (e.g., a polynomial). Our technique is based on two main pillars. The first (Sec. 3.2) is the modification of the standard robust regression method (Sec. 3.1) in order to provide additional parameters that allow for an early termination of the iterative process, and a resulting speed up of the atomic, per-pixel fitting coefficient computation. The second contribution (sec. 3.3) is a global strategy to automatically compute a data-driven sparse initialization of those parameters, and a region-growing procedure to propagate that information to the entire image domain. As we will see, the structure of the proposed algorithm is highly parallelizable, and can be implemented as efficiently as current pixel-independent solutions on multi-core machines.

3.1 Standard per-pixel robust estimator

The Least Median of Squares (LMS) estimator [32], which ensures up to 50% breakdown point (half input measurements can be outliers) [62], has already proved to be capable to provide excellent results for matte-model extraction in general settings [10, 48, 53]. It is an

Algorithm 1: Standard vs. Guided LMS algorithm.**Standard Least Median of Squares****Input:**

- AP : appearance profile
- ε : fraction of outliers
- P : probability of picking at least an inlier subset

Output: Solution S **begin**

 Compute number of trials $nTrials(\varepsilon, P)$

do

sAP_p = Random sampling of an AP 's subset of p cardinality

 Compute fitting coefficients from sAP_p

 Compute fitting coefficients from elements with the best half residuals (Refinement)

 Evaluate the median residual $M_J = med r_i^2$

 Update solution S if M_J is less than current minimal residual

while $J < nTrials$

 Compute inliers for S with $r_i^2 \leq (2.5\sigma)^2$

 Compute final fit S using all the inliers

return S **Guided Least Median of Squares****Input:**

- AP : appearance profile
- wAP : appearance profile weights
- th : residual threshold
- ε : fraction of outliers
- P : probability of picking at least an inlier subset

Output: Solution S **begin**

 Compute number of trials $nTrials(\varepsilon, P)$

do

sAP_p = Weighted random sampling of an AP 's subset of p cardinality

 Compute fitting coefficients from sAP_p

 Compute fitting coefficients from elements with the best half residuals (Refinement)

 Evaluate the median residual $M_J = med r_i^2$

 Update solution S if M_J is less than current minimal residual

while $(J < nTrials)$ or $(M_J < th)$

 Compute inliers for S with $r_i^2 \leq (2.5\sigma)^2$

 Compute final fit S using all the inliers

return S

iterative process that finds the fitting solution for a set of N elements i by minimizing the median of squared residuals r_i^2 . For a given N -dimensional appearance profile, the algorithm picks m random subsets of profile elements with size p (see Algorithm 1 left); p is the number of unknown coefficients in our fitting model. For each subset J , a generic solver (e.g., LS) is used to find the corresponding solution p_J . Then, the median value M_J is obtained as $M_J = med_{i=1, \dots, N} r_i^2(p_J)$. The final solution will be the p_J that corresponds to the minimal among all m M_J 's.

It should be noted that, since the solver used in each iteration takes the minimal number p of unknowns to compute each candidate solution, it might be prone to errors due to the presence of noise. To cope with this problem, at each iteration, it is common to compute the residual for each profile element, take the best 50% of them (assumed to be inliers), and launch a refinement step with a more overdetermined system. The presence of Gaussian noise is also an issue for the efficiency of LMS. After the robust iterations terminates a further refinement step is typically added to improve the solution. As proposed by Rousseeuw [62], we estimate a robust standard deviation as $\sigma = 1.4826 [1 + 5/(N - p)] \sqrt{M_J}$, and we use this to extract inlier elements ($r_i^2 \leq (2.5\sigma)^2$). We use this information to compute a least-squares procedure for all the inliers.

The number of trials m that ensures convergence of the algorithm is computed as $m = \log(1 - P) / \log[1 - (1 - \varepsilon)^p]$, where P is the probability that at least one of the m subsets contains only inliers, and ε is the percentage of outliers expected in the input data. This is the main reason of the heavy computational times. Since we want to ensure high accuracy, we need to set $P \approx 1$, and if we consider the highest allowed fraction of outliers, we have to set $\varepsilon = 0.5$ (50% breakdown point). We always use those values for our experiments. With those numbers, as p slightly increases (e.g., 6-term PTM polynomial), m tends to rapidly become very high; for instance, over 300 trials per pixel are required to have a probability of 99% of generating a non-contaminated base when up to 50% outliers might be present.

3.2 Guided per-pixel robust estimator

We modify that standard pipeline in order to increase efficiency but keeping a similar accuracy (see Algorithm 1 right, modifications highlighted in red). Typical LMS uses a Monte Carlo technique to draw the m subsets in no particular order, and considers a uniform distribution of the likelihood that an appearance profile element was an inlier or an outlier. It is therefore in general impossible to reduce the number of iterations for early convergence.

In our case, we provide an additional list of weights that rank the elements by a given inlier probability (we will see in Sec. 3.3 how we obtain these probabilities). Now, each time we need to obtain a subset from the input data, we employ a weighted random picking, that ensures that the probability of picking an item is proportional to its weight. Such a weighted selection can be efficiently implemented using techniques employed in Reservoir sampling [12]. The basic idea is, at each time a subset has to be selected, to assign to each potential candidate i a key r^{1/w_i} where r is a random number. By partially sorting the items, and select the top p , we can generate a sample at each trial. Since such a guided weighted selection favors good candidate solutions (according to a-priori probabilities encoded in weights), it is likely that the inlier set is identified in the first few iterations. Thus, a strategy in which the process is stopped as soon as the median residual M_J is less than a given residual threshold likely decreases the computational cost.

Algorithm 2: Global fitting

Input:

- A Array: $n \times m$ 2D-array of appearance profiles
- M : number of sparse seed pixels

Output: $n \times m$ 2D-array of fitting coefficients

begin

 Compute similarity map of I

 Select a sparse set of M seed pixels S

for pixel $p \in S$ **do in parallel**

 Compute fitting coefficients with $th_p = 0$ and uniform weights

 Compute the residual threshold $th = \text{avg}_p (r_p + 2.5\sigma_p)^2$

do

 Select candidate pixel set C

for pixel $c \in C$ **do in parallel**

 Compute fitting coefficients with $th_c = th$ and weights from the most similar, already processed, neighbor of c

while C is not empty;

return $n \times m$ 2D-array of fitting coefficients

3.3 Parallel guided estimation

The per-pixel robust estimator presented in Sec. 3.2 is employed on a global basis to convey a fitting solution for the entire image domain (see Alg. 2). In particular, we employ a region-growing approach to propagate already computed solutions to neighbor pixels, and to exploit this data-driven information to determine convergence conditions and anticipate process termination.

At the beginning of the procedure, we do not have any information about both expected residual and profile weights, neither globally nor on a per-pixel basis. For this reason, we select a small amount of evenly distributed random pixels across the image domain, and for those corresponding appearance profiles we compute a standard robust fitting, recording the solution, together with the fitting residual and the residual error for each light. This is slow,

but feasible when applied only on very few seed pixels. For the rest of the pixels, computed using our accelerated guided solution, we set a convergence threshold equal to the median of seed residuals. We then compute the solution for all pixels using a region-growing approach, which computes, in multiple passes, new pixel solutions only in the neighborhood of already computed ones. The nature of this procedure is highly suitable for parallel execution by selecting and computing all the pixels at a given iteration in parallel.

Each iteration of the region-growing pipeline starts by selecting as candidate pixels for processing the spatial neighbors of already computed pixels that exhibit a large similarity with them. To favor a similarity driven propagation, only the top 50% of most similar pixels are retained at each propagation step. The basic idea of this similarity-driven flood-fill approach is that neighboring pixels having a similar profile have a high probability of exhibiting a similar illumination behavior. Thus, the weights required for weighted random sampling can be derived from the known errors associated to each light in the most similar neighbor. In practice, we set the weight to be proportional to the residuals r_i^2 for each appearance profile element i , remapped so that the minimum residual has a weight of 1, and the maximum residual has a weight of $1/255$. These weights, together with the estimated error threshold, are used to control the guided pixel estimator.

In order to quickly compute the most similar neighboring pixels, a similarity map, depending only on input data, is computed in a pre-processing step. There is a vast amount of methods to measure similarity between appearance profiles [24]. Here we adopt a straightforward correlation-based measurement, that proved to behave well in our framework. Anyhow, the proposed pipeline is independent from this choice, and more fancy metrics can be employed.

4 Results

The proposed computational framework has been implemented on Linux using C++. We test the performance of our algorithm by using the data and the evaluation results of the *DiLiGenT* benchmark [36] (Fig. 1), which is a photometric stereo image dataset with calibrated directional lights. The dataset contains ten objects of varying reflectance properties. For each object, images with resolution of 612x512 depict the model illuminated from 96 different calibrated lighting directions. Ground-truth normal maps aligned at sub-pixel precisions are provided. All the benchmarks of our method have been executed on a PC with a Intel Core i7-4790K CPU (4 cores, 8 threads, 4.00GHz, 8MB cache) and 8GB RAM.



Figure 1: Sample images from the *DiLiGenT* benchmark [36]

4.1 Lambertian model fitting

The first test is the robust fitting of a Lambertian model aimed at the computation of the surface normals. This makes it possible to test the accuracy of the proposed approach with respect to the available ground-truth data in the *DiLiGenT* database, as well as to measure the performance in a worst-case situation for our general matte-model fitting method, given that only three parameters have to be derived from the measures. Although our aim is to propose a general strategy for guided robust fitting in the context of MLIC, we present here results and comparisons in the PS case, since it is a well researched subproblem with available benchmark data. Nevertheless, it should be noted that, while PS methods can exploit field based knowledge (monotonicity, symmetries, differential properties), we do not make any adaptation to our fitter.

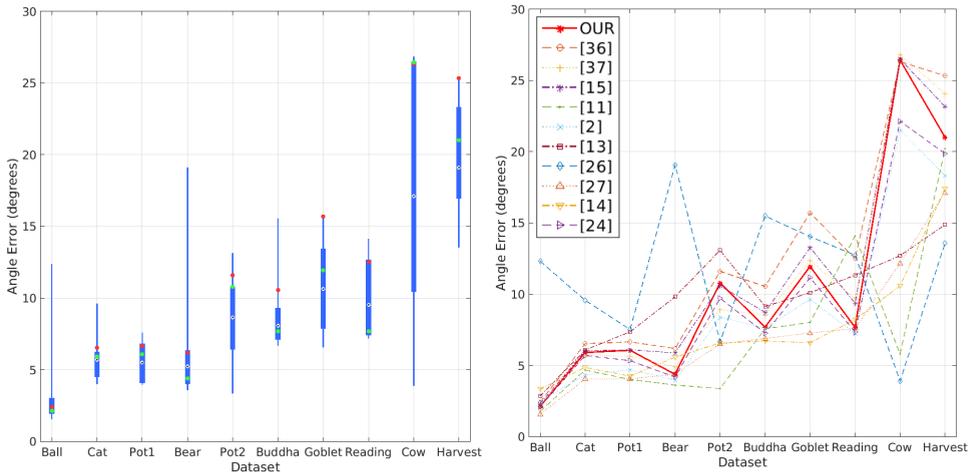


Figure 2: **Error statistics.** Comparison of our results with the state-of-the-art techniques surveyed in Shi *et al.* [36]. (Left) Distribution of the median angular error of all tested methods. The thin line depicts the minimum and maximum error interval, the box represents the 1st to the 3rd quartile range, while the central mark (white and black dot) is the median of all angular errors. The red dot is the error of the classic PS, while the error of our method is shown as a green dot. (Right) Median angular error for all approaches (our method highlighted in red).

We compare our error statistics with those presented in the *DiLiGenT* benchmark, which are the classic photometric stereo [47] as well as some non-Lambertian methods with state-of-the-art performances [0, 16, 18, 19, 20, 54, 35, 48], and with the classic robust fitting [52]. In Figure 2 we show how our method, that uses a robust statistics technique to automatically separate Lambertian from non-Lambertian behavior, provides state-of-the-art results in normals estimation for all the datasets. Both the standard full per-pixel LMS estimation and our accelerated results are within the most accurate normal fitting methods. As noted by Shi *et al.* [56], particularly complicated BRDFs (*Cow*), spatially-varying materials (*Goblet*), and concave shapes (*Reading*, *Harvest*) remain challenging types of objects. Methods tailored for general isotropic reflectance (e.g., Shi *et al.* [54]) work well for some complicated data (e.g., *Cow*), but have bad error behavior for easier objects (e.g., *Bear* or *Buddha*).

In order to evaluate the performance, for each dataset we estimate the Lambertian model

using the classic robust fitting and our guided version. In Table 1, we report the computational times together with the number of LS solve iterations, and we present the average, median, 1st and 3rd quartile of the angular error. We show how our method is capable of providing a relevant speed-up (an average of $\sim 10x$) without a significant loss in normal map accuracy. For display purposes, we show in Fig. 3 a visual comparison between the ground truth normal map of the dataset *Buddha* and those computed by the classic PS algorithm [47], the classic robust approach [52] and the proposed method. Again it provides a visual cue on how our speed-up strategy, compared to the classic robust method, does not affect the final reconstruction result. Low speed-ups are related to complex, still challenging objects, i.e., *Cow* and *Harvest*. Although our self-tuning adaptive method slightly increases the error for PS compared to the classic robust fitting, the increase is only $\sim 5\%$ on average, and results remain within the error range of specifically designed PS techniques, as shown in Fig 2. Our future work will concentrate on improving the estimation of the residual threshold.

Dataset	Time	# Solve	Avg.	Med.	1st Qr	3rd Qr.	Speed-up
Ball	2.8s/0.2s	2.5M/122K	2.0/2.1	2.1/2.1	1.5/1.6	2.6/2.6	$\sim 14x$
Cat	8.3s/0.5s	7.2M/345K	6.4/6.7	5.7/5.9	3.7/3.8	7.9/8.6	$\sim 16x$
Pot1	10.2s/0.7s	9.2M/481K	7.4/8.0	5.3/6.0	3.2/3.4	8.9/10.1	$\sim 14x$
Bear	7.7s/0.7s	6.6M/463K	5.3/5.5	4.2/4.4	2.5/2.6	6.7/7.0	$\sim 10x$
Pot2	5.9s/0.8s	5.6M/547K	11.8/12.7	9.6/10.7	5.5/5.6	16.4/18.7	$\sim 7x$
Buddha	8.3s/0.7s	7.1M/449K	9.0/9.4	7.3/7.7	4.5/4.6	10.9/12.0	$\sim 11x$
Goblet	4.7s/0.6s	4.2M/357K	12.9/14.3	11.2/11.9	6.8/7.1	16.7/20.0	$\sim 8x$
Reading	4.5s/0.7s	4.4M/461K	12.8/13.3	7.2/7.4	4.2/4.3	14.7/16.1	$\sim 7x$
Cow	4.4s/1.1s	4.2M/721K	21.3/24.0	21.9/26.1	13.2/14.9	29.2/33.4	$\sim 4x$
Harvest	9.3s/2.6s	8.8M/1.9M	24.3/25.2	18.5/19.6	8.0/8.6	34.5/35.9	$\sim 4x$

Table 1: **Lambertian Fitting.** Time and error statistics of the Lambertian Fitting with three unknown coefficients. Each cell compares the classic Robust method with our proposed Guided approach (R/G). The speed-up is $\sim 10x$ on average.

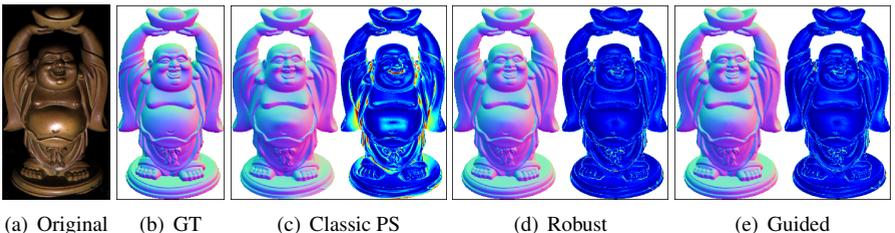


Figure 3: **Normal Maps.** Visual comparison between ground truth normal map (b) and those computed by the classic PS (c), the classic robust method (d), and our approach (e). We provide a significant speed-up ($\sim 11x$) with a preserved reconstruction quality.

4.2 Matte-model estimation for polynomial texture maps

Beside photometric stereo (or normal map) computation, our method is more general, since it aims at robustly fitting an appearance analytical model with a custom number of parameters. Testing the provided speed-up in the case of a model with only three coefficients is the worst case for us, not in terms of error comparison, but in terms of gain in computational time; this is because the more coefficients we have, the more the classic robust algorithm is

computationally heavy, and the more our algorithm can prove to be more efficient by providing a higher speed-up for it. For this reason, in Table 4(a) we show the time statistics related to the commonly used 6-coefficient polynomial fitting [14], together with the number of LS solves performed during the classic and guided iterative routine. In this case the speed-up is greater than $\sim 20x$ for most of all the datasets, and $\sim 30x$ on average. In Figure 4(b) and Figure 4(c) we visually compare the matte model relighting of the *Buddha* model obtained with the 6-coefficient polynomial computed respectively by the classic robust and our guided approach; we produce a very similar outcome while significantly breaking down the computational time.

Dataset	Time	# Solve	Speed-up
Ball	48.4s/0.8s	21.3M/311K	$\sim 60x$
Cat	137.6s/2.3s	60.9M/891K	$\sim 60x$
Pot1	173.8s/2.8s	78.2M/1.1M	$\sim 63x$
Bear	129.5s/3.3s	55.8M/1.3M	$\sim 39x$
Pot2	104.0s/3.7s	47.9M/1.4M	$\sim 28x$
Buddha	138.9s/4.0s	60.7M/1.5M	$\sim 35x$
Goblet	80.5s/4.1s	35.4M/1.5M	$\sim 19x$
Reading	77.8s/3.9s	36.6M/1.5M	$\sim 20x$
Cow	77.5s/16.1s	35.6M/6.1M	$\sim 5x$
Harvest	159.9s/25.5s	74.6M/10.5M	$\sim 6x$

(a) Speed-up



(b) Robust

(c) Guided

Figure 4: **6-Coefficient Fitting.** Time statistics of the 6-coefficient polynomial fitting (a). Each cell compares the classic Robust method with our proposed Guided approach (R/G). The fitted coefficients from the robust and the guided algorithms produce the same relighting results (b)(c), while providing a $\sim 35x$ speed-up.

5 Conclusions

We have presented a new, guided multivariate robust regression to extract the matte component of a spatially-varying material from a multi-light data. We have proved how our similarity-driven flood-fill strategy is capable of exploiting knowledge of already computed pixels to accelerate per-pixel convergence and to produce a significant speed-up (on average greater than $\sim 10x$) of the overall computation. We have presented a series of testbeds with different reflectance behaviors, and proved that our acceleration strategy maintains state-of-the-art performances. Open problems and challenges still remain, mostly related to complex BRDFs. Future works will include the study of different fitting models beside polynomials (e.g., non-linear models such as rational functions), as well as an investigation on different strategies to select initial seeds, and to compute a better estimation of convergence thresholds to further increase speed-up while decreasing exit residuals.

Acknowledgments. This work was partially supported by the Scan4Reco project funded by European Union’s Horizon 2020 Framework Programme for Research and Innovation under grant agreement no 665091. We also acknowledge the contribution of Sardinian Regional Authorities under projects VIGEC and Vis&VideoLab.

References

- [1] Jens Ackermann, Michael Goesele, et al. A survey of photometric stereo techniques. *Foundations and Trends® in Computer Graphics and Vision*, 9(3-4):149–254, 2015.
- [2] Neil Alldrin, Todd Zickler, and David Kriegman. Photometric stereo with non-parametric and spatially-varying reflectance. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [3] Svetlana Barsky and Maria Petrou. The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(10):1239–1252, 2003.
- [4] Manmohan Chandraker, Sameer Agarwal, and David Kriegman. Shadowcuts: Photometric stereo with shadows. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [5] Tat-Jun Chin, Jin Yu, and David Suter. Accelerated hypothesis generation for multistruature data via preference analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(4):625–638, 2012.
- [6] Ondrej Chum and Jiri Matas. Matching with prosac-progressive sample consensus. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 220–226. IEEE, 2005.
- [7] Irina Ciortan, Ruggero Pintus, Giacomo Marchioro, Claudia Daffara, Andrea Giachetti, and Enrico Gobbetti. A practical reflectance transformation imaging pipeline for surface characterization in cultural heritage. In *The 13th Eurographics Workshop on Graphics and Cultural Heritage*, October 2016.
- [8] E North Coleman and Ramesh Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *Computer graphics and image processing*, 18(4):309–328, 1982.
- [9] Julie Dorsey, Holly Rushmeier, and François Sillion. *Digital modeling of material appearance*. Morgan Kaufmann, 2010.
- [10] Mark S Drew, Yacov Hel-Or, Tom Malzbender, and Nasim Hajari. Robust estimation of surface properties and interpolation of shadow/specularity components. *Image and Vision Computing*, 30(4):317–331, 2012.
- [11] Graeme Earl, Kirk Martinez, and Tom Malzbender. Archaeological applications of polynomial texture mapping: analysis, conservation and representation. *Journal of Archaeological Science*, 37(8):2040–2050, 2010.
- [12] Pavlos S Efrimidis and Paul G Spirakis. Weighted random sampling with a reservoir. *Information Processing Letters*, 97(5):181–185, 2006.
- [13] Victor Fragoso and Matthew Turk. Swigs: A swift guided sampling method. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2770–2777, 2013.

- [14] Pascal Gautron, Jaroslav Krivanek, Sumanta N Pattanaik, and Kadi Bouatouch. A novel hemispherical basis for accurate and efficient rendering. *Rendering Techniques*, 2004: 321–330, 2004.
- [15] Andrea Giachetti, Irina Ciortan, Claudia Daffara, Giacomo Marchioro, Ruggero Pintus, and Enrico Gobbetti. A novel framework for highlight reflectance transformation imaging. *Computer Vision and Image Understanding*, 2017. To appear.
- [16] Dan B Goldman, Brian Curless, Aaron Hertzmann, and Steven M Seitz. Shape and spatially-varying brdfs from photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1060–1071, 2010.
- [17] Sven Havemann, Volker Settgast, Dieter Fellner, Geert Willems, Luc Van Gool, Gero Müller, Marting Schneider, and Reinhard Klein. The presentation of cultural heritage models in epoch. In *Open Digital Cultural Heritage Systems Conference*, page 32, 2008.
- [18] Tomoaki Higo, Yasuyuki Matsushita, and Katsushi Ikeuchi. Consensus photometric stereo. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1157–1164. IEEE, 2010.
- [19] Satoshi Ikehata and Kiyoharu Aizawa. Photometric stereo using constrained bivariate regression for general isotropic surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2179–2186, 2014.
- [20] Satoshi Ikehata, David Wipf, Yasuyuki Matsushita, and Kiyoharu Aizawa. Robust photometric stereo using sparse regression. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 318–325. IEEE, 2012.
- [21] Carme Julia, Felipe Lumbreras, and Angel D Sappa. A factorization-based approach to photometric stereo. *International Journal of Imaging Systems and Technology*, 21(1): 115–119, 2011.
- [22] Siddharth Kherada, Prateek Pandey, and Anoop M Namboodiri. Improving realism of 3d texture using component based modeling. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pages 41–47. IEEE, 2012.
- [23] Gudrun J Klinker, Steven A Shafer, and Takeo Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4(1):7–38, 1990.
- [24] Sanjeev J Koppal and Srinivasa G Narasimhan. Clustering appearance for scene analysis. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1323–1330. IEEE, 2006.
- [25] Thomas Malzbender, Bennett Wilburn, Dan Gelb, and Bill Ambrisco. Surface enhancement using real-time photometric stereo and reflectance transformation. *Rendering techniques*, 2006:17th, 2006.
- [26] Tom Malzbender, Dan Gelb, and Hans Wolters. Polynomial texture maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 519–528. ACM, 2001.

- [27] Paul McIlroy, Edward Rosten, Simon Taylor, and Tom Drummond. Deterministic sample consensus with multiple match hypotheses. In *BMVC*, pages 1–11, 2010.
- [28] Mark Mudge, Jean-Pierre Voutaz, Carla Schroer, and Marlin Lum. Reflection transformation imaging and virtual representations of coins from the hospice of the grand st. bernard. In *VAST*, volume 6, pages 29–40, 2005.
- [29] Mark Mudge, Tom Malzbender, Alan Chalmers, Roberto Scopigno, James Davis, Oliver Wang, Prabath Gunawardane, Michael Ashley, Martin Doerr, Alberto Proenca, et al. Image-based empirical information acquisition, scientific reliability, and long-term digital preservation for the natural sciences and cultural heritage. In *Eurographics (Tutorials)*, 2008.
- [30] Ruggero Pintus, Kazim Pal, Ying Yang, Tim Weyrich, Enrico Gobbetti, and Holly Rushmeier. A survey of geometric analysis in cultural heritage. *Computer Graphics Forum*, 35(1):4–31, 2016.
- [31] Gilles Pitard, Gaëtan Le Goïc, Hugues Favrelière, Serge Samper, Simon-Frédéric Desage, and Maurice Pillet. Discrete modal decomposition for surface appearance modelling and rendering. In *SPIE Optical Metrology*, pages 952523–952523. International Society for Optics and Photonics, 2015.
- [32] Peter J Rousseeuw and Annick M Leroy. *Robust regression and outlier detection*, volume 589. John wiley & sons, 2005.
- [33] Holly Rushmeier, Gabriel Taubin, and André Guézic. Applying shape from lighting variation to bump map capture. In *Rendering Techniques 97*, pages 35–44. Springer, 1997.
- [34] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Elevation angle from reflectance monotonicity: Photometric stereo for general isotropic reflectances. *Computer Vision—ECCV 2012*, pages 455–468, 2012.
- [35] Boxin Shi, Ping Tan, Yasuyuki Matsushita, and Katsushi Ikeuchi. Bi-polynomial modeling of low-frequency reflectances. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(6):1078–1091, 2014.
- [36] Boxin Shi, Zhe Wu, Zhipeng Mo, Dinglong Duan, Sai-Kit Yeung, and Ping Tan. A benchmark dataset and evaluation for non-lambertian and uncalibrated photometric stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3707–3716, 2016.
- [37] Fredric Solomon and Katsushi Ikeuchi. Extracting the shape and roughness of specular lobe objects using four light photometric stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):449–454, 1996.
- [38] Jiuai Sun, Melvyn Smith, Lyndon Smith, Sagar Midha, and Jeff Bamber. Object surface recovery using a multi-light photometric stereo technique for non-lambertian surfaces subject to shadows and specularities. *Image and Vision Computing*, 25(7):1050–1057, 2007.

- [39] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [40] Kam-Lun Tang, Chi-Keung Tang, and Tien-Tsin Wong. Dense photometric stereo using tensorial belief propagation. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 132–139. IEEE, 2005.
- [41] Ben Tordoff and David W Murray. Guided sampling and consensus for motion estimation. In *European conference on computer vision*, pages 82–96. Springer, 2002.
- [42] Philip H. S. Torr and Colin Davidson. Impsac: Synthesis of importance sampling and random sample consensus. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(3):354–364, 2003.
- [43] Frank Verbiest and Luc Van Gool. Photometric stereo with coherent outlier handling and confidence estimation. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.
- [44] Oliver Wang, Prabath Gunawardane, Steve Scher, and James Davis. Material classification using brdf slices. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 2805–2811. IEEE, 2009.
- [45] Andreas Wenger, Andrew Gardner, Chris Tchou, Jonas Unger, Tim Hawkins, and Paul Debevec. Performance relighting and reflectance transformation with time-multiplexed illumination. In *ACM Transactions on Graphics (TOG)*, volume 24, pages 756–764. ACM, 2005.
- [46] Geert Willems, Frank Verbiest, Wim Moreau, Hendrik Hameeuw, Karel Van Lerberghe, and Luc Van Gool. Easy and cost-effective cuneiform digitizing. In *The 6th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST 2005)*, pages 73–80. Eurographics Assoc., 2005.
- [47] Robert J Woodham. Photometric method for determining surface orientation from multiple images. *Optical engineering*, 19(1):191139–191139, 1980.
- [48] Lun Wu, Arvind Ganesh, Boxin Shi, Yasuyuki Matsushita, Yongtian Wang, and Yi Ma. Robust photometric stereo via low-rank matrix completion and recovery. In *Asian Conference on Computer Vision*, pages 703–717. Springer, 2010.
- [49] Tai-Pang Wu, Kam-Lun Tang, Chi-Keung Tang, and Tien-Tsin Wong. Dense photometric stereo: A markov random field approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(11):1830–1846, 2006.
- [50] Qingxiong Yang and Narendra Ahuja. Surface reflectance and normal estimation from photometric stereo. *Computer Vision and Image Understanding*, 116(7):793–802, 2012.
- [51] Yizhou Yu and Johnny T Chang. Shadow graphs and 3d texture reconstruction. *International Journal of Computer Vision*, 62(1-2):35–60, 2005.

- [52] Alan Yuille and Daniel Snow. Shape and albedo from multiple images using integrability. In *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*, pages 158–164. IEEE, 1997.
- [53] Mingjing Zhang and Mark S Drew. Robust luminance and chromaticity for matte regression in polynomial texture mapping. In *European Conference on Computer Vision*, pages 360–369. Springer, 2012.
- [54] Mingjing Zhang and Mark S Drew. Efficient robust image interpolation and surface properties using polynomial texture mapping. *EURASIP Journal on Image and Video Processing*, 2014(1):1–19, 2014.