Improving the digitization of shape and color of 3D artworks in a cluttered environment

Fabio Bettio*, Enrico Gobbetti*, Emilio Merella*, Ruggero Pintus*†

*CRS4 Visual Computing, Pula, Italy – http://www.crs4.it/vic/
†Yale University, New Haven, CT, USA – http://www.yale.edu/

Abstract—We propose an approach for improving the digitization of shape and color of 3D artworks in a cluttered environment using 3D laser scanning and flash photography. In order to separate clutter from acquired material, semi-automated methods are employed to generate masks for segment the 2D range maps and the color photographs, removing unwanted 3D and color data prior to 3D integration. Sharp shadows generated by flash acquisition are trivially handled by this masking process, and color deviations introduced by the flash light are corrected at color blending time by taking into account the object geometry. The approach has been applied to, and evaluated on, a large scale acquisition campaign of the Mont'e Prama complex, an extraordinary collection of stone fragments from the Nuragic era, depicting larger-than-life archers, warriors, boxers, as well as small models of prehistoric nuraghe (cone-shaped stone towers). The acquisition campaign has covered 36 statues mounted on metallic supports, acquired at 0.25mm resolution, resulting in over 6200 range scans (over 1.3G valid samples) and 3426 10Mpixel photographs.

I. Introduction

The increasing performance and proliferation of digital photography and 3D scanning devices is making it possible to acquire at reasonable costs very dense and accurate sampling of both geometric and optical surface properties of real objects. A wide variety of cultural heritage applications particularly benefits of this evolution, which makes it potentially possible to construct accurate colored digital replicas not for single digital objects but at a large scale. The wide availability of accurate reconstructions built from objective measures has many applications that range from virtual restoration to dissemination.

The most widely used approach today is a combination of (triangulation) laser scanning with digital photography. Digital object surfaces are reconstructed by the laser-scan-generated range maps, while the apparent color value sampled in digital photos is transferred by registering the photos with respect to the 3D model and mapping it to the 3D surface using the recovered inverse projections. This approach is particularly well suited to the cultural heritage domain, since scanning and photographic acquisition campaigns can be performed quickly and easily, without the need to move objects to specialized acquisition labs. The most costly and time consuming part of 3D reconstruction is thus moved to post-processing, which can be performed off-site. In recent years, research has thus focused on improving and automating the processing steps, leading, e.g., to (semi-)automated scalable solutions for range-map alignment [1], surface reconstruction from point clouds [2]–[5], photo registration [6], [7], and color mapping [8]–[10].

Fig. 1. **Reassembled Neolithic statue with supports.** The black support structures holds the fragments in the correct position, with minimal contact surface, avoiding pins and holes in the original material. A 360 degree view is possible, but scanning is difficult because of clutter and shadows.

In this paper, we improve the scanning and color mapping pipe-line for the important and difficult case of on-site scanning of 3D artworks in a cluttered environment. This case arises, for instance, when scanning restored and reassembled ancient statues in which (heavy) stone fragments are maintained in place by a custom exostructure (see Fig. 1 for an example). Digitizing statues without removing the supports enables scanning directly on location and without the need of moving the fragments. On the other hand, the presence of the supporting structure typically generates shadowing effects and extra acquired geometry. With the standard 3D scanning pipeline, this produces shadow-related color artifacts and leads to a heavy and mostly manual post-processing step that includes geometry cleaning and 2D pixels masking.

We propose a practical approach for improving the digitization of shape and color of 3D artworks in a cluttered environment. Geometry is acquired using a triangulation laser scanner, which produces range and reflectance maps, while color is acquired by taking a collection of photographs with an uncalibrated camera, using the camera flash as the only illumination. A semi-automatic method is employed to segment the 2D range maps and color photographs, separating clutter from the statue material. Starting from a small training set of manually segmented reflectance maps and photographs,

we learn a simple statistical description of the statue material, which is used to generate approximate segmentation masks for both range maps and color photographs. These masks, are then refined using graph-cut based segmentation, and used to remove the unwanted 3D and color data prior to manual editing. Sharp shadows created by flash acquisition are trivially handled by this masking process, and color deviations introduced by the flash light are corrected at color blending time by taking into account the object geometry. The quality and efficiency of our approach have been evaluated on a real world large scale digital reconstruction project concerning the Mont'e Prama complex, an extraordinary collection of stone fragments from the Nuragic era, depicting larger-than-life archers, warriors, boxers, as well as small models of prehistoric nuraghe (cone-shaped stone towers). The acquisition campaign, performed in a museum setting without removing the statues from their custom supports, has covered 36 statues, acquired at 0.25mm resolution, resulting in over 6200 range scans (over 1.3G valid samples) and 3426 10Mpixel photographs.

Contribution. Our main contributions are the following: an *easy-to-apply acquisition protocol* based on laser scanning and flash photography; a simple and practical *semi-automatic method for clutter removal and photo masking*; a *scalable implementation* of the entire masking, editing, and color-blending pipe-line, working fully out-of-core without limits on model size and photo number; an evaluation of a *large-scale real-workd application* of the method and tools on a massive acquisition campaign.

Advantages. Our method handles the difficult case of on-site acquisition of a cluttered environment, as exemplified by the case of statue fragments maintained in place by exostructures. Based on the standard combination of laser scanning and flash digital photography, it does not need complex illumination setups (just a dark room during color acquisition), reducing time required for color acquisition. Post-processing time is radically reduced with respect to current procedures, thanks to semi-automatic masking of color and geometry and scalable color-corrected color blending.

Limitations. The method assumes that the statue material is easily separable from the unwanted support structure by analyzing reflectance and color, and can thus not be applied when the statue and the supporting structure are visually indistinguishable. However, It must be noted that this differentiation is enforced in modern restoration practices. The results presented in this paper take the assumption that the statue material is fairly diffuse and homogeneous, as is common for all ancient artifacts made of stone. This is not an intrinsic limitation of the method, which could also be applied, as is, if the supporting material has those characteristics just by reversing the masks. Regions of contact between the exostructure and the imaged object cannot obviously be recovered, since they are invisible to the imaging devices. This is common to all use-cases involving static supports. Since these parts are small, and generally uninteresting, infilling techniques are generally used. This is, however, orthogonal to our work.

II. RELATED WORK

Our system extends and combines state-of-the-art results in a number of technological areas. In the following, we only discuss the approaches most closely related to our novel contributions. We refer the reader to the recent survey of Callieri et al. [11] for up-to-date information on the entire 3D scanning pipeline.

Color and geometry masking. Editing and cleaning the acquired 3D model is often the most time-consuming reconstruction task [11]. While some techniques exist for semi-automatic 3D clutter removal in 3D scans, they are typically limited to well-defined situations (e.g., walls vs. furniture for interior scanning [12] or walls vs. organic models for exterior scanning [13]). Interactive 2D segmentation is a well known research topic with different state-of-the-art solutions that typically involve classification and/or editing of color image datasets (see well-established surveys [14], [15]). In general, the aim of these techniques is to efficiently cope with the foreground/background extraction problem at the cost of as less as possible user interactive effort. The most simple tool is Magic Wand of Adobe Photoshop 7 [16]. The user selects a point and the software automatically computes a connected set of pixels that belongs to the same region. Unfortunately, an acceptable segmentation is rarely achieved due to the fact that choosing the right color or intensity tolerance value is a hard or even impossible task. All earliest methods, as the intelligent scissors [17], active contours [18], and Bayes matting [19], require a considerable degree of user interaction in order to attain satisfactory results. More accurate approaches are presented that solve the semi-automatic image segmentation by using Graph Cuts [20]; here the user marks a small set of background and/or foreground pixels as seeds, and the algorithm propagates that information to the remaining image regions. Among the big number of extensions of the Graph Cuts methodology [21], [22], the *GrabCut* technique [23] combines a very straightforward manual operation, with a color modeling and an extra layer of (local) minimization to the Graph Cuts technique; this puts a light load on the user and proves to be very robust in different segmentation scenarios. Here we propose an adaptation of the *GrabCut* approach to the problems of point cloud geometry editing and pre-processing of images for texture blending. First we employ this method to perform a minimal user assisted training of the algorithm on a small set of acquired range maps and images, and then we combine it with a rough color segmentation in order to automatically mask the massive amount of 3D points and textures.

Color acquisition and blending. Most cultural heritage applications require to associate material properties to geometric reconstructions of the sampled artifact. While many works exist for sampling Bidirectional Radiance Distribution Functions (BRDF) [24], [25] using sophisticated controlled lighting environments, most cultural heritage applications impose fast on-site acquisition together with the use of low-cost and easy to use procedures and technologies. Color photography is the most common approach. Since removing lighting artifacts requires knowledge of the lighting environment, one approach is to employ specific techniques that uses probes [26], [27], which are, however, hard to use in practice in a museum setting with local lights. Dellepiane et al. [28], [29] proposed, instead to used camera flash light, employing the Flash Lighting Space Sampling (FLiSS), a correction space where a correction matrix is associated to each point in the camera field of view. The method requires a heavy calibration step. Starting from

the fact that medium- to high-end reflex cameras support fairly uniform flash illumination and RAW data acquisition modes that produce images where each pixel value is proportional to incoming radiance [30], we take the simpler approach of using a constant color balance correction for the entire set of photographs and apply a per-pixel intensity correction based on geometric principles. This approach reduces calibration needs. In addition, while previous color blending pipe-lines worked on large triangulated surfaces [8] or single-resolution point-clouds [10], we blend images directly on a multiresolution structures, leading to increased scalability.

Scalable editing and processing. Massive point-cloud processing and interactive point-cloud editing are required to produce high quality 3D models from an input set of registered photos and merged geometries. In this work, we represent geometric datasets as a forest of out-of-core octrees of point samples [31]–[33], and employ the same structure for all operations, including color blending and editing. While previous works split and refine nodes based on strict per-node sample budgets, our approach is based on local density estimating, which allows us to use more balanced structures.

III. TECHNIQUE OVERVIEW

Fig. 2 outlines the proposed approach, which consists in a on-site and an off-site phase.

The on-site operations are geometry and color acquisition. Geometry acquisition is performed with a (triangulation) laser scanner, incrementally coarsely aligning range maps in order to monitor 3D surface coverage. Color acquisition, is, instead, performed in a dark room using flash photography. A Macbeth color checker, visible in at least one of the photographs, is used for post-process color calibration. Similarly to what done for range maps, coverage is (optionally) checked by coarsely aligning the photographs using a SfM pipeline.

All the rest of the work can be performed off-site, using separate geometry and color pipe-lines, which communicate only at the final merging step. In order to remove geometric clutter, the user manually segments a very small subset of the input range maps, and produces a training dataset that allows the algorithm to automatically mask unwanted geometry. This step exploits the reflectance channel. The automatic masking can be (optionally) revised by visually inspecting it and manually improving the segmentation. Masks are then applied to all range maps, which are then finely registered with a global registration approach and manually edited for giving the final touches.

The color pipe-line is similar. Starting from the photographs in raw format, after the training performed by the user on a small subset of images, the algorithm automatically masks all the input photos. After a visual check and a manual refinement, the masked photos, already coarsely aligned among themselves with SfM, are aligned with the geometry using the approach of Pintus et al. [6]. The photos are finally blended to the surface by color-blended projection [8], [10]. During the blending step, colors are calibrated using a data extracted from the color checker, and differences in illumination caused by flash photography is corrected using geometric information.

Details on semi-automatic geometry and color masking, scalable data consolidation and color mapping are provided in the following sections. The corresponding phases are highlighted in yellow in Fig. 2.
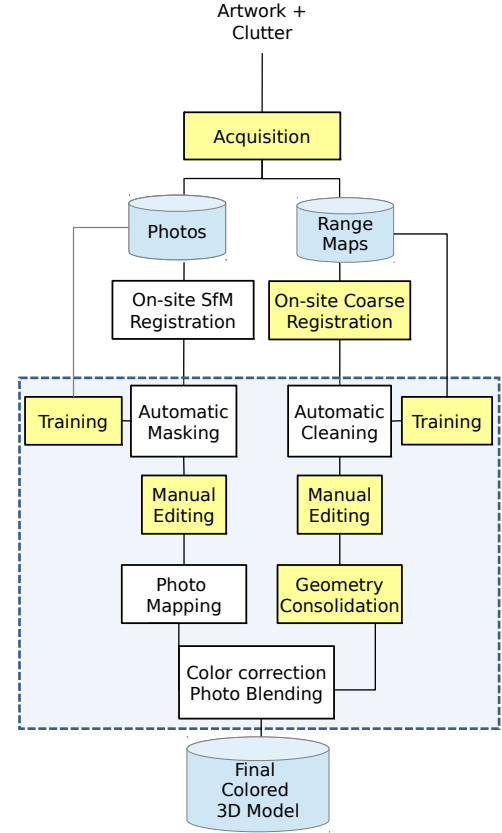

Fig. 2. **Pipeline.** We improve digitization of 3D artworks in a cluttered environment using 3D laser scanning and flash photography. Semi-automated methods are employed to generate masks for segment the 2D range maps and the color photographs, removing unwanted 3D and color data prior to 3D integration. Sharp shadows generated by flash acquisition are handled by the masking process, and color deviations introduced by the flash light are corrected at color blending time by taking into account object geometry. User-guided phases are highlighted in yellow.

IV. SEMI-AUTOMATIC GEOMETRY AND COLOR MASKING

Our masking process aims to separate the foreground geometry (the object to be modeled) by the cluttering data (in particular occluding objects), under the assumption of different appearances, as captured in the reflectance and color signals. Starting from a manual segmentation of small set of examples (Sec. IV-A), we learn a histogram-based classification of materials (Sec. IV-A), which is then refined by finding an optimal labeling of pixels using graph cuts (Sec. IV-C), before a final (optional) user-assisted revision, which can be performed using the same tool used for manual segmentation.

A. Manual segmentation

In order to perform the initial training, the user is provided with a custom segmentation tool, with the same interface for range maps and color images. The tool allows to visually browse images/scans in the acquisition database, visually select a small subset of them (typically, less than 5%), and draw a segmentation in the form of a binary mask, using white for foreground and black for background. The mask layer is

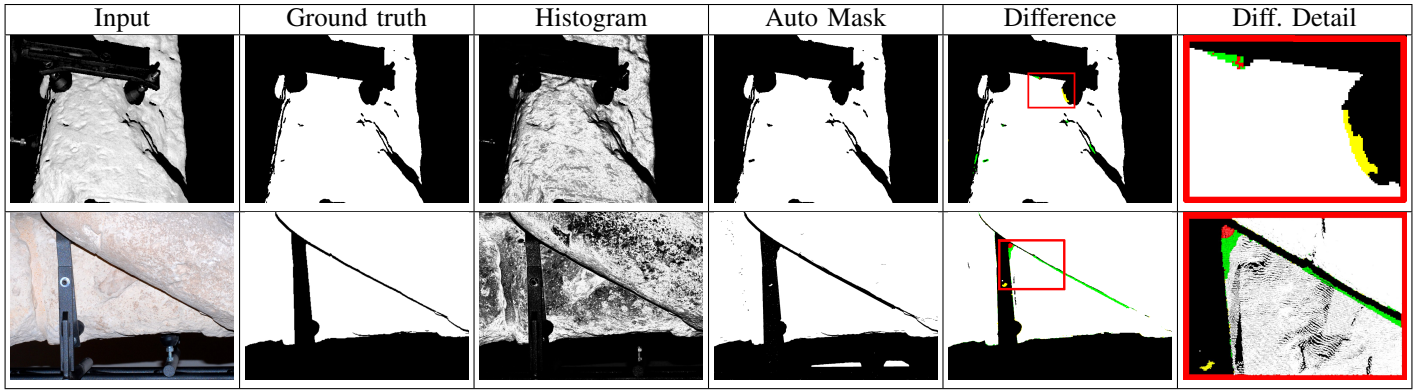| Input	Ground truth	Histogram	Auto Mask	Difference	Diff. Detail

Fig. 3. **Automatic masking.** Geometry (top) and color (bottom) results for a single image. From left to right: acquired reflectance/color image; user-generated ground truth masking; mask generated by histogram-based classification; final automatically generated mask; difference to ground truth; magnified region of difference image. In the difference image, black and white pixels are perfect matches, while yellow pixels are false positives, green pixels are false negative points on this image, and red pixels are real false negative points considering the entire dataset.

rendered on top of the image layer, and the user can vary the transparency of the mask to evaluate the masking results. In addition to using standard draw/erase brushes, we support interactive grab-cut segmentation [23], in which the user selects a bounding box of the foreground object to initialize the segmentation method.

B. Histogram-based classification

The user-selected small subset of manually masked images and range maps is used to learn a rough statistical distribution of pixel values that characterize foreground objects. For artifacts made of fairly uniform materials, e.g. stone sculptures, 3-4 range maps and 4-5 images are typically sufficient.

For range maps, we build a 1D histogram of reflectance values, quantized to 32 levels, by accumulating all pixels that were marked as foreground in the user-defined mask. For color images, instead, we use a 2D histogram based on hue and saturation, both quantized to 32 levels. Ignoring the value component is more robust to shading variation due to flash illumination and variable surface orientation. The process is repeated for all manually masked images, accumulating histogram values before a final normalization step.

The histogram computed on the training set can be used for a rough classification of range map/image pixels based on reflectance/color information. This classification is simply obtained by back-projecting each image pixel to the corresponding bin location, and interpreting the normalized histogram value as foreground probability.

It is worth to be noted that it is not important whether the histogram is computed from foreground or clutter data; if the rest of the pipeline is consistent, the only constraint is the already mentioned assumption that the two appearances are reasonably well separable.

C. Graph cut segmentation

As illustrated in Fig. 3, third column, the histogram-based classification is very noisy, but roughly succeeds in identifying the foreground pixels, which are generally marked with high probabilities. This justifies our use of histograms for the rough classification step rather than the more complex statistical representations typically used in soft segmentation [34], [35].

Segmentation is improved by using the rough histogram-based classification as starting point for an iterated graph cut process. We initially separate all pixels in two regions: probably foreground for those with normalized histogram value larger than 0.5, and probably background for the others. We then iteratively apply the GrabCut [23] segmentation algorithm, using a Gaussian Mixture Models with 5 components per region and estimating segmentation using min-cut. As illustrated in Fig. 3, column 4, the process produces tight and well regularized segmentation masks.

V. DATA CONSOLIDATION AND COLOR MAPPING

The final result of the automatic masking step is a mask image associated to each range map and color image. These masks are used for pre-filtering geometry and color information before further processing.
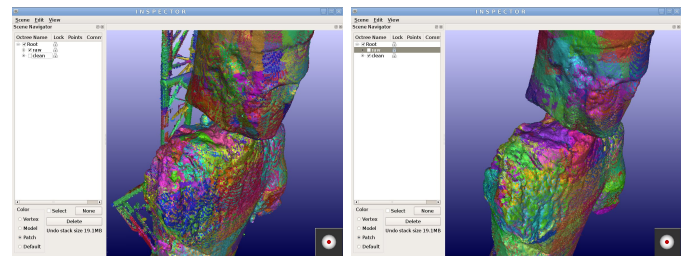

Fig. 4. **Interactive inspection and editing.** Left: all captured range maps imported into out-of-core scene structure without applying masks. Right: automatic masking removes most if not all clutter. Editing can thus be limited to fixing fine details.

A. Multiresolution octree for point-cloud management

All the rest of the processing, including color mapping, is performed by using an editable out-of-core structure based on a forest of octrees. The structure consists in a scene structure, hierarchically grouping models and associating to each group a rigid body transformation for positioning it with respect to its parent. The leafs of the scene structure are out-of-core octrees of points. Similarly to Wand et al.'s

system [31], data points are stored unmodified in the leaf nodes of the octrees, while inner provide grid-quantization-based multiresolution representations. In contrast to previous work, we dynamically decide whether to subdivide or merge nodes base on a local dynamic density estimation. In particular, at each insertion/deletion operation in a leaf node, we count how many grid cells are occupied in the quantization grid that would be associated to the node upon subdivision. If this number is larger than four times the number of points, we consider the node dense enough for splitting. The structure provides efficient rendering and allows for handling very large data sets using out-of-core storage, while supporting efficient online insertion, deletion and modification of point samples.

After completing the masking process, we import into a scene structure all the range maps, using a separate octree per range map (see Fig. 4). The initial transformation of each range map is the coarse alignment transformation computed during the scanning campaign, while normals and local sampling space are computed by local differences. The hierarchical scene structure is exploited for grouping scans (e.g., separating low-res from hi-res scans, as well as labeling semantically relevant parts). Alignment and editing operations are applied to the structure using an interactive editor. Alignment is performed by applying to (subsets of) the range scans a global registration method based on the scalable approach of Pulli et al. [36], using GPU-accelerated KNN to speed-up local pairwise ICP [37]. After satisfactory global alignment, multiple octrees can optionally be merged together. Interactive editing is performed on the structure using a select-and-apply paradigm supporting undo history. At each application of a modifier (e.g., point deletion), modified samples are moved to a temporary out-of-core structure (an memory-mapped array) prior to modification. By associating the array to the undo list, we are able to perform reversible editing.
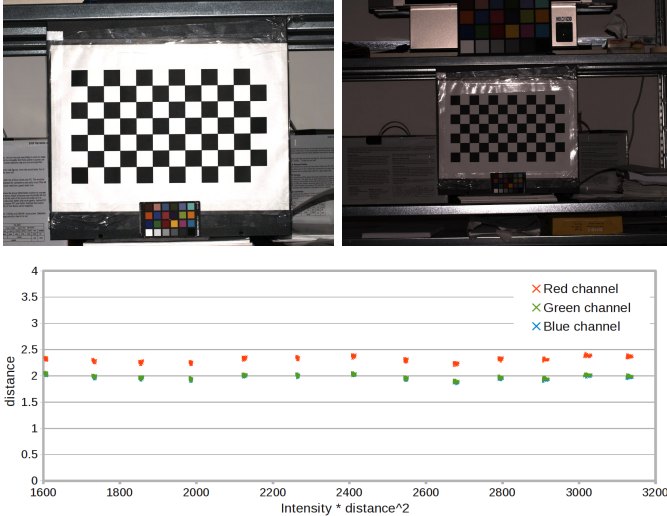

Fig. 5. **Flash illumination.** Top: two out of many calibration images taken at difference distances under flash illumination. Bottom: using RAW camera data, distance-based scaling provides a reasonable correction. Balance between color channels can then be ensured using color-checker-based calibration.

B. Color correction and blending

The color attribute is obtained first by mapping masked photo in the same 3D model reference frame [6], and then by

Fig. 6. **Color correction and masking.** Left: original image under flash illumination: note sharp shadows and uneven intensity; Right: projected images with corrected color and masked regions highlighted in red. Note how distance-based correction correctly equalizes shading and support as well as shadow areas are identified by the mask.

performing seamless texture blending of those images onto the surface [10]. In contrast to previous work, we blend and map images directly to the out-of-core multiresolution structure, and perform color correction starting from captured RAW images during the mapping operation.

In our context, flash illumination is a viable way to image the objects, as it provides us sharp shadows together with knowledge of per-image illumination direction. Since at color mapping-time the geometry of the image is known, we can correct each projected pixel according to the position of surface on which it projects with respect to the camera and the flash light. In addition, cluttering material, e.g., the supporting exostructure, generates sharp shadows, which can be easily identified both by the masking process and by shadow casting of the geometry.

In contrast to previous work [28], [29], we handle images directly in RAW format, which allows us to correct images also without prior camera calibration. We experimentally measured that, on a medium/high end camera, such as the Nikon D200, RAW data acquisition produces images where each pixel value is proportional to incoming radiance (as also verified elsewhere [30]), and that the flash light emits fairly uniform illumination within a large working space. Fig. 5 shows results of images of a checkerboard taken in a dark room with t=1/250s f/11.0+0.0, ISO 400. As shown in the graph on the right, values (measured on the white checkerboard squares) are proportional to $1/d^2$, where d is the distance from the flash light. It is thus reasonable to simply correct pixel intensities at color mapping time by applying a correction factor of $(d_i/d_0)^2$, where d_0 is the user-provided desired object distance, and d_i is the distance to the flash of the surface visible from the pixel. Fig 6 shows how distance-based correction correctly equalizes shading. In the same image, it can be noted that support as well as shadow areas are correctly identified by the mask. As in standard settings, color balance can be recovered by taking an image of a calibration target (color checker).

While more accurate results can possibly be obtained with calibration techniques, even the most accurate ones performed off-site [28], [29] do not perfectly match local illumination settings, since indirect illumination is not taken into account. We thus consider this uncalibrated approach to be of practical use. It should be noted that, whenever needed, these alternate techniques can be easily performed in a post-process from the same captured data. The availability of RAW images in the

captured database adds in addition the ability to perform a variety of post-process enhancements [30].

C. Surface reconstruction and filling

The final colored point cloud can be directly used for many applications. In order to produce consolidated models represented as colored triangle meshes, among the many state-of-the-art approaches [2]–[4], we adopt the smoothed signed distance technique [5], a recent variational formulation for the problem of reconstructing a watertight models that forces the implicit function to be a smooth approximation of the signed distance function to the surface. The method produces watertight surfaces and directly handles the color signal within the same framework. This is particularly interesting in our setting, since it smoothly fills the unsampled areas by diffusing both geometry and colors.

VI. IMPLEMENTATION AND RESULTS

We implemented on Linux the methods described in this paper in a C++ software library and system. The out-of-core octree structure is implemented on top of *Berkeley DB 4.8.3*, and *OpenMP* is used for parallelizing blending operations. The automatic masking subsystem is implemented on top of *OpenCV 2.4.3*. RAW color images from the camera are handled using the *dcraw 9.10* library. The SfM software used for image-to-image alignment is *Bundler 0.4.1* [38]. All tests were run on a PC with a 8 cores Intel Core i7-3820 CPU (3.60GHz), 64GB RAM, and a NVIDIA GTX680 graphics board.

A. The Mont'e Prama test case

We have tested our system with a variety of high resolution models and settings. In this paper, we discuss the results obtained during a very large scanning campaign of the Mont'e Prama complex, a large set of Neolithic stone sculptures created by the Nuragic civilization in Western Sardinia. More than 5000 sculpture fragments were recovered after four excavation campaigns carried out between 1975 and 1979.

According to the most recent estimates, the stone fragments came from a total of 44 statues depicting archers, boxers, warriors and models of prehistoric nuraghe (cone-shaped stone towers). These can be traced to an as-yet undefined period, which goes from the tenth to the seventh century BC. Restoration, carried out at the Centro di Restauro e Conservazione dei Beni Culturali of Li Punti (Sassari) resulted in the partial reassembly of 25 human figures with height varying between 2 and 2.5 meters, as well as of 13 one-meter-sized nuraghe models.

Following modern restoration criteria, reassembly has been performed in a non-invasive way (no drilling or bolt insertions into the sculptures). Definite joining fragments have been glued using a water soluble epoxy resin, and all the gaps on the surface filled with resin have been covered with lime-mortar stucco. Custom external supports have been designed to sustain all the parts of a statue in order to assure stability to all the components without the use of mechanical attachments. The support allows a 360 degree view of the statue.

The scanning campaign covered 36 reconstructed statues, which were scanned and photographed directly in the museum setting. Figure 7 summarizes the reconstruction results.

B. Acquisition

The geometry of all the statues has been acquired at 0.25mm resolution, resulting in over 6200 640x480 range scans using a Minolta Vivid 9i in tele mode. Geometry scanning resulted in over 1.3G valid position samples. Color was acquired with a Nikon D200 camera with a 50mm lens. All photos were taken with a flash light in a dark room, with a shutter speed of 1/250s, aperture f/11.0+0.0, and ISO sensitivity 400. Photo scans resulted in 3426 10Mpixel photographs. The on-site scanning campaign required 620 hours to complete for a team of two people, one camera, and one scanner. In practice, on-site time was reduced by parallelizing acquisition with two scanning teams working on two statues at a time. The acquisition time includes scanning sessions, flash photography sessions (in dark room), and coarse alignment of scans using our point cloud editor. Photo alignment using the SfM pipeline was performed after each flash acquisition session, and in parallel to the scanning session, in order to verify whether sufficient coverage had been reached. Average bundle adjustment time was of 2 hours/statue.

C. Automatic geometric masking

	Model pts	Clutter pts	False-Pos	False-Neg	True False-neg
Samples	51.4M	790K	240 (486)	35757 (11219)	5369 (2746)
(%)			0.03 (0.06)	4.53 (1.42)	0.68 (0.35)

Fig. 8. **Automatic geometric masking evaluation.** Results of manual segmentation of a single statue (*Guerriero3*) compared with automatic masking results. We report the number of range map samples labeled as model ("Model pts") and clutter ("Clutter pts") in the ground truth dataset, the samples erroneously labeled as statue ("False-positive") or clutter ("False-negative") in the automatic method, as well as the number of false negative points that really lead to missing data in the combined dataset ("True False-negative"). Values between parentheses compare the cleaned-up and the ground-truth dataset, instead of the purely automatic method. Percentages are computed with respect to the number of clutter points.

The quality and efficiency of our automatic geometric masking process has been extensively evaluated on a selected dataset that was also manually segmented to create a ground-truth result. The selected statue, *Guerriero3* (depicted in Fig. 1), is composed by 226 range maps (54 of which containing clutter data).

Each ground-truth mask was created manually from the reflectance channel of the acquired range map using our interactive mask editor. An experienced user took about 330 minutes to complete the manual segmentation process for the entire statue. For the sake of completeness, we also evaluated the time required to remove clutter data from the 3D dataset by direct 3D point cloud editing, as done in typical scanning pipelines. Using our out-of-core point cloud editor, this operation was completed by an experienced user in about 300 min, which is not so different from the time required by the manual 2D segmentation approach. By taking into account the relative complexity of the other statues, we can estimate a total time of about 130-150 man hours for the manual cleaning of the entire collection of statues.

Fig. 7. **Mont'e Prama complex.** From left to right: full set of reconstructed statues; original image of one statue (warrior 3); reconstructed model; closeup on head of reconstructed model; closeup on eye of reconstructed model.

The automatic segmentation process started by manually segmenting 5 reflectance images using the same editor used for manual segmentation. This training set was used as input to the automatic classifier. The entire process took 9 minutes for the training set creation and 6 minutes for automatic mask computation on 8 cores. The automatically generated masks where then manually verified and retouched using our system. This additional step, which is optional, took about 30 minutes. Applying the automatic process to the entire statue collection took just 5 hours excluding manual cleaning and 13.5 hours with the manual post-process cleanup, leading to a more than ten-fold speed-up with respect to the manual approaches.

The efficiency of the automatic masking method is evaluated in the table in Fig. 8, which shows the results of the comparison tests between the automatic segmented masks (with and without post-process manual cleaning) and the ground-truth dataset.

More than 95.0% of the clutter samples are correctly labeled. False-positive samples represent extra points which can be easily identified and removed from the automated masks via 2D editing and they are only about 0.03% of the total clutter in the ground-truth dataset. False-negative points represent statue samples that have been erroneously masked; they are about 4.5% (1.4% in the clean-up dataset) of the total clutter in the ground-truth dataset. Since overlapping range maps typically acquires the same geometric region from multiple points of view, a false-negative sample is not a problem if its value is correctly classified in at least one mask covering the same area. By taking into account this fact, we verified that these really missing points (*True false negative*) are only 0.68% of the total imaged clutter surface. This check has been performed by looking in overlapping scans for samples within a radius of 1mm from each missing sample. As illustrated in Fig. 3, the missing points are, in addition, often very sparse or represent small boundary area and, thus, their effect on dataset quality is pretty limited.

D. Automatic color masking

The quality and efficiency of the color masking process has been evaluated similarly to what done for the geometry. The selected statue, "Guerriero3" (depicted in Fig. 1), was imaged by 68 photographs (33 of which containing clutter data). Manually masking the images took 181 minutes, while applying the automated process took 9 minutes to generate the training set, 15 minutes to automatically computing the masks on 8 cores, plus other 30 minutes for the optional post-process cleanup. Speed-up is, again, substantial. This allowed us to

	Model pts	Clutter pts	False-Pos	False-Neg	True False-neg
Samples	220.5M	12.1M	381K (334K)	263K (253K)	8642 (7725)
(%)			3.16 (2.77)	2.18 (2.09)	0.07 (0.06)

Fig. 9. **Automatic color masking evaluation.** Results of manual segmentation of a single statue (*Guerriero3*) compared with automatic masking results. We report the number of colored samples labeled as model ("Model pts") and clutter ("Clutter pts") in the ground truth dataset, the samples erroneously labeled as statue ("False-positive") or clutter ("False-negative") in the automatic method, as well as the number of false negative points that really lead to missing data in the combined dataset ("True False-negative"). Values between parentheses compare the cleaned-up and the ground-truth dataset, instead of the purely automatic method. Percentages are computed with respect to the number of clutter points.

perform masking for the entire set of statues in only 41 hours (17 hours without the post-process cleaning). By taking into account the relative complexity of the other statues, we can estimate a total time of about 145 man hours for the manual cleaning of the entire collection of statues.

As illustrated in the table in Fig. 9, we obtain similar results in the color masking than those obtained for geometry masking. Again, about 95.0% of the samples are correctly labeled. False-positive samples, represent in the color case points where clutter color could potentially leak to geometry areas. These represent about 3% of the clutter area, i.e., below 0.2% of the model area. False-negative points represent, instead, statue samples that do not receive color by a given image since they have been erroneously masked; they are about 2.2% (2.1% in the cleaned-up dataset) of the total clutter in the ground-truth dataset, but reduce to negligible amounts when considering overlapping photographs. This is because of the large overlap between photos and the concentration of false negative in thin boundary areas covered from other angles. Sampling redundancy, required for alignment purposes, is thus largely beneficial also for automatic masking.

E. Consolidation and coloring

The generated geometry and color masks have been used to create digital 3D models for the 36 statues (see Fig. 7). All models have been imported after cleaning in our system based on forests of octrees, which has been used for all the 3D editing and color blending. Using lossless compression, we store our hierarchical database at an average cost about 38B/sample, with per sample positions, normals, radii, colors, and blending weights (including database overhead). Disk footprints for our multiresolution editable representation are thus similar to storing single-resolution uncompressed data. We

have compared the performance of our system with respect to the state-of-the-art streaming color blender of [10]. The total blending time of *Guerriero3* has been of 23 minutes with our pipeline, which works directly on the editable representation and includes flash illumination color correction, while the streaming color blender took 2.5 minutes for pre-computing the Morton-ordered sample stream and the culling hierarchy, and 26 minutes for color blending. The increased flexibility of our system does not, thus, introduce additional overhead, while not requiring additional temporary storage and supporting fast turnaround times during iterative editing sessions. Flash color correction proved to be adequate and produces visually appealing results without unwanted color variation and/or visible seams between acquisition (see Fig. 7 for an example).

VII. Conclusion

We have proposed an approach for improving the digitization of shape and color of 3D artworks in a cluttered environment using 3D laser scanning and flash photography. The method, applied on a real-world large scale digital reconstruction project concerning 36 statues mounted on supports, has proven capable of notably reducing on-site acquisition times and off-site editing times, while being able to produce good-quality results.

References

[1] P. Pingi, A. Fasano, P. Cignoni, C. Montani, and R. Scopigno, "Exploiting the scanning sequence for automatic registration of large sets of range maps," *Computer Graphics Forum*, vol. 24, no. 3, pp. 517–526, 2005.

[2] M. Kazhdan, M. Bolitho, and H. Hoppe, "Poisson surface reconstruction," in *Proc. SGP*, 2006, pp. 61–70.

[3] J. Manson, G. Petrova, and S. Schaefer, "Streaming surface reconstruction using wavelets," in *Computer Graphics Forum*, vol. 27, no. 5, 2008, pp. 1411–1420.

[4] G. Cuccuru, E. Gobbetti, F. Marton, R. Pajarola, and R. Pintus, "Fast low-memory streaming MLS reconstruction of point-sampled surfaces," in *Graphics Interface*, May 2009, pp. 15–22.

[5] F. Calakli and G. Taubin, "SSD: Smooth signed distance surface reconstruction," in *Computer Graphics Forum*, vol. 30, no. 7, 2011, pp. 1993–2002.

[6] R. Pintus, E. Gobbetti, and R. Combet, "Fast and robust semi-automatic registration of photographs to 3D geometry," in *Proc. VAST*, 2011, pp. 9–16.

[7] M. Corsini, M. Dellepiane, F. Ganovelli, R. Gherardi, A. Fusiello, and R. Scopigno, "Fully automatic registration of image sets on approximate geometry," *IJCV*, pp. 1–21, 2012.

[8] M. Callieri, P. Cignoni, M. Corsini, and R. Scopigno, "Masked photo blending: Mapping dense photographic data set on high-resolution sampled 3D models," *Computers & Graphics*, vol. 32, no. 4, pp. 464–473, 2008.

[9] R. Pintus, E. Gobbetti, and M. Callieri, "A streaming framework for seamless detailed photo blending on massive point clouds," in *Proc. Eurographics Area Papers*, 2011, pp. 25–32.

[10] ——, "Fast low-memory seamless photo blending on massive point clouds using a streaming framework," *ACM JOCCH*, vol. 4, no. 2, p. Article 6, 2011.

[11] M. Callieri, M. Dellepiane, P. Cignoni, and R. Scopigno, *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*. CRC Press, 2011, ch. Processing sampled 3D data: reconstruction and visualization technologies, pp. 69–99.

[12] A. Adan and D. Huber, "3D reconstruction of interior wall surfaces under occlusion and clutter," in *Proc. 3DIMPVT*, 2011, pp. 275–281.

[13] F. Lafarge and C. Mallet, "Creating large-scale city models from 3D point clouds: a robust approach with hybrid representation," *IJCV*, vol. 99, no. 1, pp. 69–85, 2012.

[14] H. Zhang, J. E. Fritts, and S. A. Goldman, "Image segmentation evaluation: A survey of unsupervised methods," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 260–280, 2008.

[15] K. McGuinness and N. E. O'Connor, "A comparative evaluation of interactive segmentation algorithms," *Pattern Recognition*, vol. 43, no. 2, pp. 434–444, 2010.

[16] Adobe Systems Inc., "Adobe photoshop user guide," 2002.

[17] E. N. Mortensen and W. A. Barrett, "Toboggan-based intelligent scissors with a four-parameter edge model," in *Proc. CVPR*, vol. 2, 1999.

[18] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *IJCV*, vol. 1, no. 4, pp. 321–331, 1988.

[19] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A Bayesian approach to digital matting," in *Proc. CVPR*, vol. 2, 2001, pp. II–264.

[20] Y. Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N–D images," in *Proc. ICCV*, vol. 1, 2001, pp. 105–112.

[21] Y. Zeng, D. Samaras, W. Chen, and Q. Peng, "Topology cuts: A novel min-cut/max-flow algorithm for topology preserving segmentation in N–D images," *Computer vision and image understanding*, vol. 112, no. 1, pp. 81–90, 2008.

[22] N. Xu, N. Ahuja, and R. Bansal, "Object segmentation using graph cuts based active contours," *Computer Vision and Image Understanding*, vol. 107, no. 3, pp. 210–224, 2007.

[23] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," in *ACM TOG*, vol. 23, no. 3, 2004, pp. 309–314.

[24] P. Debevec, T. Hawkins, C. Tchou, H.-P. Duiker, W. Sarokin, and M. Sagar, "Acquiring the reflectance field of a human face," in *Proc. SIGGRAPH*, 2000, pp. 145–156.

[25] H. P. A. Lensch, J. Kautz, M. Goesele, W. Heidrich, and H.-P. Seidel, "Image-based reconstruction of spatial appearance and geometric detail," *ACM TOG*, vol. 22, no. 2, pp. 234–257, 2003.

[26] P. Debevec, "Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proc. SIGGRAPH*, 1998, pp. 189–198.

[27] M. Corsini, M. Callieri, and P. Cignoni, "Stereo light probe," *Computer Graphics Forum*, vol. 27, no. 2, pp. 291–300, 2008.

[28] M. Dellepiane, M. Callieri, M. Corsini, P. Cignoni, and R. Scopigno, "Flash lighting space sampling," in *Computer Vision/Computer Graphics Collaboration Techniques*, 2009, pp. 217–229.

[29] ——, "Improved color acquisition and mapping on 3D models via flash-based photography," *ACM JOCCH*, vol. 2, no. 4, p. Article 9, 2010.

[30] S. Kim, H. Lin, Z. Lu, S. Suesstrunk, S. Lin, and M. Brown, "A new in-camera imaging model for color computer vision and its application," *IEEE Trans. PAMI*, vol. 34, no. 12, pp. 2289–2302, 2012.

[31] M. Wand, A. Berner, M. Bokeloh, A. Fleck, M. Hoffmann, P. Jenke, B. Maier, D. Staneker, and A. Schilling, "Interactive editing of large point clouds," in *Proc. SPBG*, 2007, pp. 37–46.

[32] M. Wimmer and C. Scheiblauer, "Instant points: Fast rendering of unprocessed point clouds," in *Proc. SPBG*, 2006, pp. 129–137.

[33] C. Scheiblauer and M. Wimmer, "Out-of-core selection and editing of huge point clouds," *Computers & Graphics*, vol. 35, no. 2, pp. 342–351, 2011.

[34] M. A. Ruzon and C. Tomasi, "Alpha estimation in natural images," in *Proc. CVPR*, 2000, pp. 18–25.

[35] Y.-Y. Chuang, B. Curless, D. H. Salesin, and R. Szeliski, "A bayesian approach to digital matting," in *Proc. CVPR*, 2001, pp. 264–271.

[36] K. Pulli, "Multiview registration for large data sets," in *Proc. 3D Digital Imaging and Modeling*, 1999, pp. 160–168.

[37] L. Cayton, "Accelerating nearest neighbor search on manycore systems," in *Proc. IEEE IPDPS*, 2012, pp. 402–413.

[38] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," *IJCV*, vol. 80, no. 2, 2008.