Mont'e Scan: Effective shape and color digitization of cluttered 3D artworks

FABIO BETTIO, ALBERTO JASPE VILLANUEVA, EMILIO MERELLA, FABIO MARTON, and
ENRICO GOBBETTI, CRS4 Visual Computing, Italy
RUGGERO PINTUS, CRS4 Visual Computing, Italy and Yale University, USA

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 used to segment the range maps and the color photographs. This approach allows the removal of unwanted 3D and color data prior to the integration of acquired data in a 3D model. Sharp shadows generated by flash acquisition are easily handled by this masking process, and color deviations introduced by the flash light are corrected at the color blending step by taking into account the geometry of the object. The approach has been evaluated on a large scale acquisition campaign of the Mont'e Prama complex. This site contains an extraordinary collection of stone fragments from the Nuragic era, which depict small models of prehistoric nuraghe (cone-shaped stone towers), as well as larger-than-life archers, warriors, and boxers. The acquisition campaign has covered 37 statues mounted on metallic supports. Color and shape were acquired at a resolution of 0.25mm, which resulted in over 6200 range maps (about 1.3G valid samples) and 3817 photographs.

1. 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 stand to benefit particularly from this technological evolution. In fact, this technological progress is leading to the

possibility to construct accurate colored digital replicas not only for single digital objects but at a large scale. Accurate reconstructions built from objective measures have many applications, ranging from virtual restoration to visual communication.
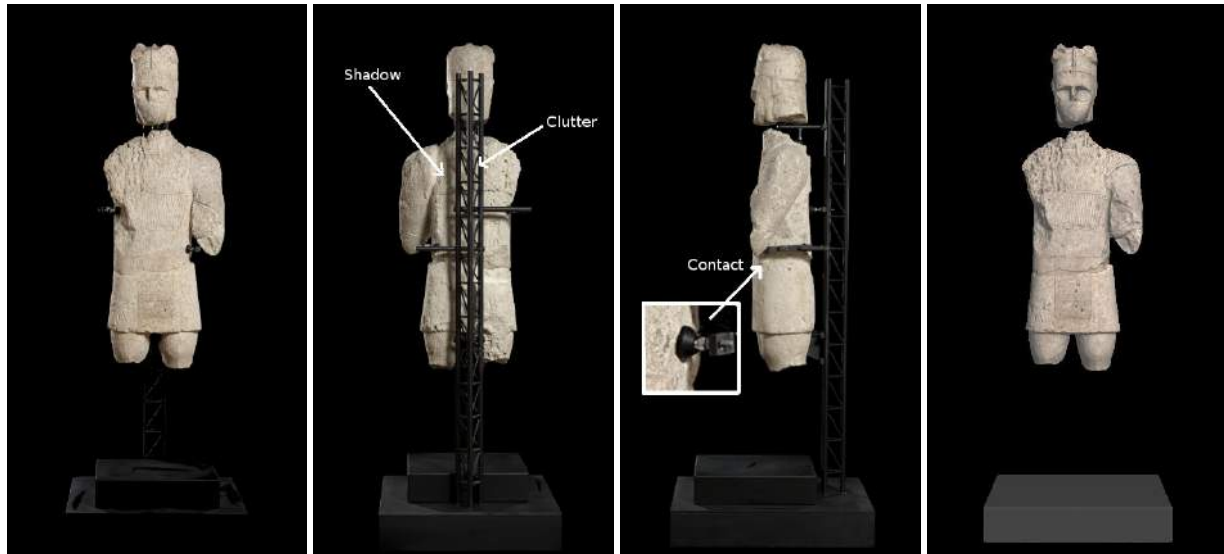

Fig. 1. **Reassembled Nuragic statue with supports and its virtual reconstruction.** The black support structure holds the fragments in the correct position, with minimal contact surface, avoiding pins and holes in the original material. A 360-degree view is possible, but color and shape capture is difficult because of clutter, occlusions, and shadows. The rightmost image depicts our 3D reconstruction. Photo courtesy of ArcheoCAOR.

The digitization approach most widely used today is a combination of laser scanning with digital photography. Using computational techniques, digital object surfaces are reconstructed from 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. Since early demonstrations of the complete modeling pipeline (e.g., [Bernardini and Rushmeier 2002; Levoy et al. 2000]), most of its components have reached sufficient maturity for adoption in a variety of application domains. This approach is particularly well suited to cultural heritage digitization, 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. Thus, in recent years research has focused on improving and automating the post-processing steps – for instance, leading to (semi-)automated scalable solutions for range-map alignment [Pingi et al. 2005], surface reconstruction from point clouds [Kazhdan et al. 2006; Manson et al. 2008; Cuccuru et al. 2009; Calakli and Taubin 2011], photo registration [Pintus et al. 2011c; Corsini et al. 2012], and color mapping [Callieri et al. 2008; Pintus et al. 2011b; 2011a]. Even though passive image-based methods have recently emerged as a viable (and low-cost) 3D reconstruction technology [Remondino 2011], the standard pipeline based on laser scanning or other active sensors still remain a widely used general-purpose approach, mainly because of the higher reliability for a wider variety settings (e.g., featureless surfaces) [Remondino 2011; Koutsoudis et al. 2014].

In this paper, we tackle the difficult problem of effectively adapting the 3D scanning pipeline to the acquisition of color and shape of 3D artworks on-site, 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 allows one to perform scanning directly on location and without moving the fragments, therefore enabling a completely contactless approach. On the other hand, the presence of the supporting structure typically generates shadow-related color artifacts, holes due to occlusion effects and extra geometry that must be removed. With the standard 3D scanning pipeline these issues lead to laborious and mostly manual post-processing steps – including cleaning the geometry and careful pixel masking (for more details see Sec. 2 with an overview of the related work).

Motivated by these issues, in this work we present a practical approach for improving the digitization of the shape and color of 3D artworks in a cluttered environment. While our methods are generally applicable, the work was spurred by our involvement in the *Digital Mont'e Prama* project (see Sec. 3), which included the fine-scale acquisition of 37 large statues and therefore required robust and scalable methods. Our main contributions presented in this article 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, infilling, color-correction, and color-blending pipeline, that works fully out-of-core without limits on model size and photo number;

—the evaluation of the method and tools in a *large-scale real-world application* involving a massive acquisition campaign covering 37 statues mounted on metallic supports, acquired at 0.25mm resolution, resulting in over 6200 range scans (approximately 1.3G valid samples) and 3817 10Mpixel photographs.

This work is an invited extended version of our *Digital Heritage 2013* contribution [Bettio et al. 2013]. In addition to supplying a more thorough exposition in this work, we also provide significant new material including an analysis of requirements, the description of an improved pipeline which performs color infilling and takes into account position and orientation of surfaces during color mapping, and the presentation of results on the complete Mont'e Prama dataset.

2. RELATED WORK

Our system extends and combines state-of-the-art results in a number of technological areas. In the following text we only discuss the approaches most closely related to our novel contributions, which focus on effective clutter removal. For more detailed information on the entire 3D scanning pipeline we refer the reader to the recent survey by Callieri et al. [2011]. Our approach assumes that color images are reliably registered to the 3D models. Obtaining this registration is an orthogonal problem for which a variety of solutions exists (e.g., [Borgeat et al. 2009; Pintus et al. 2011c; Corsini et al. 2012; Pintus and Gobbetti 2014]). The results in this work have been obtained with the approach of Pintus et al. [2014].
Color and geometry masking. Editing and cleaning the acquired 3D model is often the most time-consuming reconstruction task [Callieri et al. 2011]. 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 [Adan and Huber 2011] or walls vs. organic models for exterior scanning [Lafarge and Mallet 2012]). Manual editing is also typically employed in approaches that work on images and range maps. For instance, Farouk et al. [2003] embedded a simple image editor into the scanning GUI. We also follow the approach of working on 2D representations, but concentrate our efforts to reduce human interventions. Interactive 2D segmentation is a well-known research topic with several

state-of-the-art solutions that typically involve classification and/or editing of color image datasets (see well-established surveys [Zhang et al. 2008; McGuinness and O'Connor 2010]). In general, the aim of these techniques is to efficiently cope with the foreground/background extraction problem with the least possible user input. The simplest tool available is the Magic Wand in Adobe Photoshop 7 [Adobe Systems Inc. 2002]. The user selects a point and the software automatically computes a connected set of pixels that belong to the same region. Unfortunately, an acceptable segmentation is rarely achieved automatically since choosing the correct color or intensity tolerance value is a difficult or even impossible task. Many classic methods, such as intelligent scissors [Mortensen and Barrett 1999], active contours [Kass et al. 1988] and Bayes matting [Chuang et al. 2001], require a considerable degree of user input in order to achieve satisfactory results. More accurate approaches have been presented that solve the semi-automatic image segmentation problem by using Graph Cuts [Boykov and Jolly 2001]; 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 large number of extensions to the Graph Cuts methodology [Zeng et al. 2008; Xu et al. 2007], the *GrabCut* technique [Rother et al. 2004] combines a very straightforward manual operation, with a color modeling and an extra layer of (local) minimization to the Graph Cuts technique; this requires a small effort from the user but proves to be very robust in different segmentation scenarios. In this work we propose an adaptation of the *GrabCut* approach to the problems of editing point cloud geometries and pre-processing images for texture blending. In our approach, we perform a minimal user-assisted training on a small set of acquired range maps and images in order to automatically remove clutter from images and point clouds.

Color acquisition and blending. Most cultural heritage applications require the association of material properties to geometric reconstructions of the sampled artifact. While many methods exist for sampling Bidirectional Radiance Distribution Functions (BRDF) [Debevec et al. 2000; Lensch et al. 2003] in sophisticated environments with controlled lighting, the typical cultural heritage applications impose fast on-site acquisition and 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 use probes [Debevec 1998; Corsini et al. 2008]. However, these techniques are hard to use in practice, in typical museum settings with local lights. Dellepiane et al. [2009; 2010] proposed, instead, to use light from camera flashes. They propose to use the Flash Lighting Space Sampling (FLiSS) – a correction space where a correction matrix is associated to each point in the camera field of view. Nevertheless, this method requires a laborious calibration step. Given that medium- to high-end single-lens reflex (SLR) cameras support fairly uniform flash illumination and RAW data acquisition modes that produce images where each pixel value is proportional to incoming radiance [Kim et al. 2012], 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 effectively reduces calibration work. While our previously published results [Bettio et al. 2013] only employed a distance-based correction, in this work we employ a more complete correction that also takes into account surface orientation. The method is similar to the one originally used by Levoy et al. [2000], without the need of special fiber optic illuminators and of per-pixel color response calibration. Under our flash illumination, taken from relatively far from the statues (2.5m), the flash can be approximated as a point source and energy deposition on the statue is negligible compared to typical ambient lighting. In addition, while previous color blending pipelines worked on large triangulated surfaces [Levoy et al . 2000; Callieri et al. 2008] or single-resolution point-clouds [Pintus et al. 2011a], we blend images directly on multiresolution structures leading to increased scalability. The pipeline presented in our original work [Bettio et al. 2013] is also combined here with inpainting and infilling methods for constructing seamless models.

Our implementation is based on combining screened Poisson reconstruction [Kazhdan and Hoppe 2013] with an anisotropic color diffusion process [Wu et al. 2008] implemented in a multigrid framework. *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 [Wand et al. 2007; Wimmer and Scheiblauer 2006; Scheiblauer and Wimmer 2011] 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 estimates, which allows us keep to the structures more balanced.

3. CONTEXT AND METHOD OVERVIEW

The design of our method, which is of general use, has taken into account requirements gathered from domain experts in the context of a large scale project. Section 3.1 briefly illustrates the design process and the derived requirements, while Sec. 3.2 provides a general overview of our approach, justifying the design decisions in relation to the requirements.

Fig. 2. **Mont'e Prama Statues on display at the CRCBC exhibition hall.** Scanning was performed on-site.

3.1 Requirements

While our methods are of general use, our work is motivated by the *Digital Mont'e Prama* project, a collaborative effort between CRS4 and the *Soprintendenza per i Beni Archeologici per le Province di Cagliari ed Oristano* (ArcheoCAOR, the government department responsible for the archeological heritage of the provinces of Cagliari and Oristano), which aims to digitally document, archive, and present to the public the large and unique collection of pre-historic statues from the Mont'e Prama complex, including larger-than-life human figures and small models of prehistoric *nuraghe* (cone-shaped stone towers). The project covers aspects ranging from 3D digitization to visual exploration (see Balsa et al. [2014] and Marton et al. [2014] for more details on the visual exploration aspects).

The Mont'e Prama Collection. The Mont'e Prama complex is a large set of sandstone 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
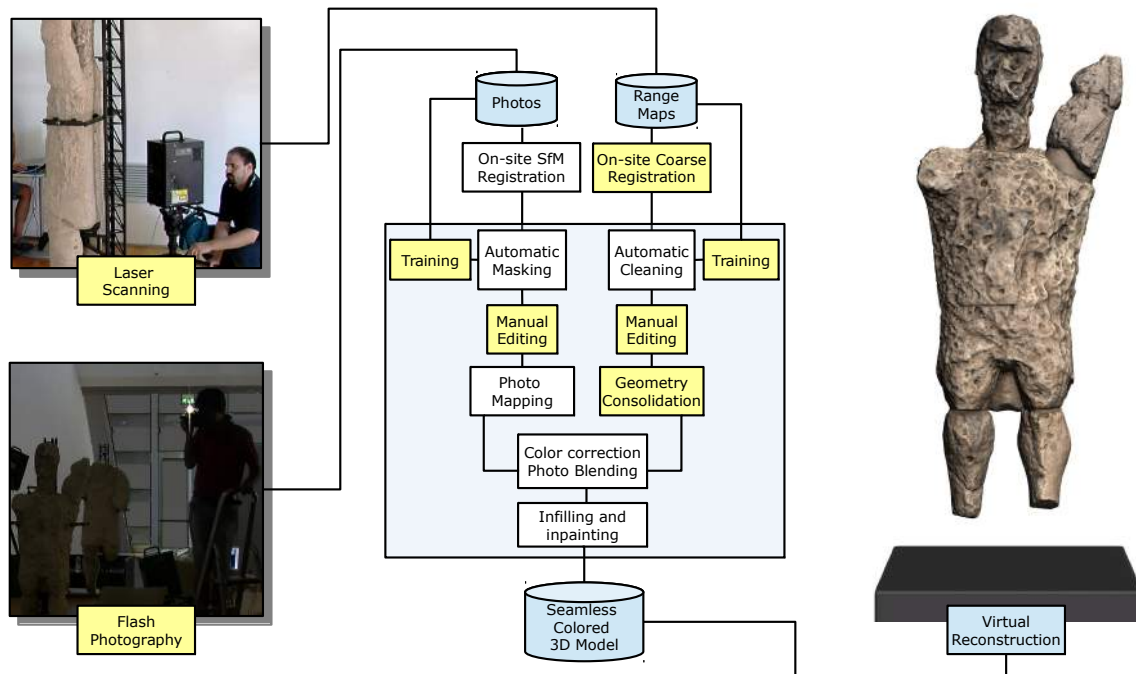
Fig. 3. **Pipeline.** We improve digitization of 3D artworks in a cluttered environment using 3D laser scanning and flash photography. Semi-automated methods are employed to generate masks to segment the 2D range maps and the color photographs, removing unwanted 3D and color data prior to 3D integration. 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. A final seamless model is created by combining Poisson reconstruction with anisotropic color diffusion. User-guided phases are highlighted in yellow.

models of prehistoric nuraghe. These can be traced to an as-yet undetermined period, which goes from the tenth to the seventh century BC. Restoration, carried out at the *Centro di Restauro e Conservazione dei Beni Culturali (CRCBC)* of Li Punti (Sassari, Italy) resulted in the partial reassembly of 25 human figures with height varying between 2 and 2.5 meters, and 13 approximately one-meter-sized nuraghe models. Following modern restoration criteria, reassembly was performed in a non-invasive way (no drilling or bolt insertions into the sculptures). Definitely joining fragments have been glued together using a water-soluble epoxy resin, and all the gaps on the resin-filled surface were covered with lime-mortar stucco. Custom external supports have been designed to sustain all the parts of a statue in order to ensure stability to all the components without the use of mechanical attachments, while minimizing contacts with the statue and maximizing visibility (see Fig. 1). All supports allow a 360 degree view of the statue. The project covers 37 out of the 38 reassembled statues – one nuraghe was excluded since it is small and still subject to further reassembling work (see Fig. 2 for the setup of the exhibition in the CRCBC hall and Fig. 9 for the final digitally reconstructed models).

Requirement Analysis. In order to design our acquisition and 3D reconstruction technique, we embarked in a participatory design process involving domain experts with the goal of collecting the detailed requirements of the application domain; the experts included two archaeologists from *ArcheoCAOR* and

two restoration experts from *CRCBC*. Additional requirements stem from our analysis of related work (see Sec. 2) and our own past experience developing capture, and reconstruction systems for architecture and cultural heritage [Cuccuru et al. 2009; Pintus et al. 2011a; Mura et al. 2013]. In the following text we describe the main requirements used for guiding the development process, briefly summarizing how they were derived:

R1. **On-site scanning in a limited time frame.** Scanning of the entire set of statues must be performed while the statues are on display in the CRCBC exhibition hall. All the scanning has to be performed during a time-frame of max 8h/day (early morning and late evening) and in a period of no more than two months.

R2. **No contact and no fragment motion.** The statues are made of relatively soft sandstone and are very sensitive to contact and erosion. Therefore, the acquisitions must be performed with the statues mounted on their supports and without any contact. The only motion possible is the sliding of the base of each fully reassembled statue (i.e., 2D rotation and translation on the exhibition floor). Fragments cannot be moved one relative to the other.

R3. **Macro- and micro-structural shape capture and reconstruction.** Like almost any decorated artifact of cultural heritage, the statues present information at multiple scales (global shape and carvings). Even the finest material micro-structure carries valuable information (e.g., on the artistic decorations, as well as on the carving process). For instance, almost all the fragments of the Mont'e Prama statues have millimeter-sized carvings on various parts, and thus require sub-millimetric model precision.

R4. **High-resolution color capture and reconstruction.** While no results of a human coloring process is currently visible on the statues, color information provides valuable information and is part of the aura of the objects. For instance, the fragments acquired in this project are often made of stones with different grains. In fact, the sandstone itself has a spatially varying texture, color inserts and characteristic patterns are due to organic and fossil inclusions as well as limestone deposits. In addition, traces of fire are visible on multiple fragments and lead to a characteristic dark-brown coloring. Moreover, surface finish is variable and subtly modifies stone color, while not sensibly modifying the reflectance, which remains extremely diffuse. In short, it is important to capture and reproduce color at a resolution comparable to the resolution of geometry capture.

R5. **Seamless virtual models for public presentation.** While the acquisition process should lead to the creation of models accurate enough for archival, study, and further restoration activities, one of the main motivations for the project is to use them for visual communication to scholars as well as to the public at large, both in museum settings and on the web. In order to preserve the aura of original artifacts, the 3D models should be at high resolution and seamless. The exostructures should not be present in these virtual models. Since the exostructure has contact points with the artwork, holes are unavoidable. However, for public presentation the model should be watertight, as visible holes would be confusing and unattractive. Since contact areas are small and placed in areas with lack of detail, smoothly infilling and inpainting these areas is considered acceptable for public display.

R6. **Low-labor approach.** To be applicable to this Mont'e Prama project, and to ensure wide adoption in other cultural heritage campaigns, the proposed method should be flexible and required a reasonably low amount of labor, to reduce the time and the costs required to produce the models.

3.2 Approach

Requirements **R1**-**R6** were taken as guidelines for our research, which resulted in the definition of a semi-automatic color and shape 3D scanning pipeline for the acquisition and reconstruction of cluttered

objects. The pipeline is an enhancement of the well-established pipeline based on laser scanning and digital photography. This makes the method general-purpose enough to be used in different settings by operators trained to use the classic robust approach based on 3D laser scanning. Figure 3 outlines the approach used, which consists of a short on-site phase and a subsequent, mostly automatic, off-site phase (**R1** and **R6**).

The only on-site operations are the acquisition of geometry and color, which are performed in a contact-less manner by only sliding and rotating the statue while it is mounted on its support (**R2**). The geometry acquisition operation is performed with a triangulation laser scanner (**R3**), which produces range and reflectance maps that are incrementally coarsely aligned during scanning in order to monitor 3D surface coverage. On the other hand, color is acquired in a dark environment by taking a collection of photographs with an uncalibrated camera while using the camera flash as the only source of light (**R1**, **R4** and **R6**). A Macbeth color checker, visible in at least one of the photographs, is used for post-process color calibration. Analogously to the geometry acquisition step, coverage is (optionally) checked on-site by coarsely aligning the photographs using a Structure-from-Motion (SfM) pipeline.

The remainder of the work can be performed off-site using semi-automatic (**R6**) geometry and color pipe-lines that communicate only at the final merging step. In order to remove geometric clutter (**R5**) the user manually segments a very small subset of the input range maps and produces a training dataset that is sufficient for the algorithm to automatically mask unwanted geometry. This step exploits the reflectance channel of the laser scanner. As commonly done for cultural heritage pipelines, the automatic masking can be in principle be revised by visually inspecting and optionally manually improving the segmentation, using the same tools employed for creating the example masks. Note that this step, in contrast to previous work [Farouk et al. 2003], is entirely optional (see Sec. 8 for an evaluation of the manual labor required). In order to create a clean 3D model, the masks are applied to all range maps, which are then finely registered with a global registration method and optionally edited manually for the finishing touch. The geometry reconstruction is then performed using Poisson reconstruction [Kazhdan et al. 2006], which takes care of infilling small holes that appears in the unscanned areas (i.e., where supports touch the surface of the scanned object) using a smoothness prior on the indicator function. The effect is similar to volumetric diffusion [Davis et al. 2002].

The color pipeline follows a similar work pattern. It begins 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 removing clutter (**R5**). The user then optionally performs a visual check and a manual refinement, then the masked photos – already coarsely aligned among themselves with SfM – are aligned with the geometry using the method by Pintus et al. [2014]. The photos are finally mapped to the surface by color-blended projection [Callieri et al 2008; Pintus et al. 2011a]. During the blending step colors are calibrated using a data extracted from the color checker and the differences in illumination caused by the flash used during photography are corrected using geometric information (**R5**). Finally, an anisotropic diffusion process [Wu et al. 2008] is employed to perform a conservative inpainting of the areas left without colors due to occlusions (**R5**).

It should be noted that the infilling and inpainting approaches employed in this work are minimalist (**R5**). We do not aim at reconstructing high frequency details in large areas. Instead, we just smoothly extend color and geometry from the neighborhood of holes to avoid the presence of confusing and unattractive holes for public display applications.

Details on semi-automatic geometry and color masking, as well as scalable data consolidation and color mapping are provided in the following sections. The corresponding phases are highlighted in yellow in Fig. 3.

4. DATA ACQUISITION

While geometry acquisition is performed using the standard triangulation laser scanning approach, color acquisition is performed using an uncalibrated flash camera. In our context, flash illumination is a viable way to image the objects, as it provides us with sharp shadows together with information on the image-specific direction of the illumination. Since at color mapping-time the geometry of the image is known, we can correct each projected pixel according to the position of the surface on which it projects with respect to the camera and the flash's light, thus obtaining a reasonable approximation of the surface albedo (see Sec. 7). In addition, cluttering material – e.g., the supporting exostructure – generates sharp shadows which can be easily identified both by the masking process and by taking into account geometric occlusion in the color mapping process (see Sec. 5).

In contrast to previous work [Dellepiane et al. 2009; 2010; Levoy et al. 2000], we handle images directly in RAW format, which allows us to correct images without prior camera calibration. We experimentally measured that on a medium/high end camera, such as the Nikon D200 employed in our work, acquisition in RAW format produces images with pixel values proportional to the incoming radiance at medium illumination levels (as also verified elsewhere [Friedrich et al. 2010; Kim et al. 2012]), and that the flash emits a fairly uniform light within a reasonable working space.
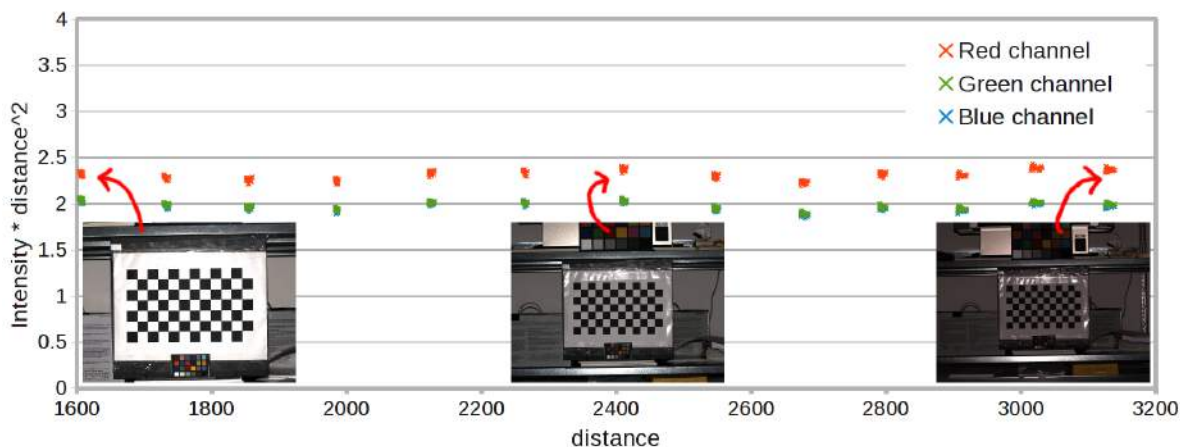

Fig. 4. **Flash illumination.** Using RAW camera data distance-based scaling provides a reasonable correction. Balance between color channels can then be ensured using color-checker-based calibration.

Sensor near-linearity has been verified by taking images of a checkerboard in a dark room with t=1/250s f/11.0+0.0, ISO 400. As shown in the graph in Fig. 4 values (measured on the white checkerboard squares) are proportional to $1/d^2$, where d is the distance from the flash light. Distance-based scaling can be thus exploited at color mapping time to provide reasonable correction, while balance between color channels can be ensured using color-checker-based calibration (see Sec. 7). The characteristics of flash illumination have been verified by taking a photograph of a white diffuse material at 2m using the same 50mm lens used for photographing the statues. After correcting for lighting angle and distance the illumination varies only by a maximum of 7.6% within the view frustum.

While more accurate results may be obtainable with calibration techniques, even the most accurate ones performed off-site [Levoy et al. 2000; Dellepiane et al. 2009; 2010; Friedrich et al. 2010] do not perfectly match local shading and illumination settings since, in particular, indirect illumination is not

taken into account and the photographed materials are (obviously) not perfect Lambertian scatterers. We thus consider this uncalibrated approach to be suitable for practical use. It should be noted that whenever needed these alternate techniques can be easily performed during post-processing using the same captured data. Indeed, the availability of RAW images in the captured database grants the ability to perform a variety of post-process enhancements [Kim et al. 2012].

5.   SEMI-AUTOMATIC GEOMETRY AND COLOR MASKING

Our masking process aims to separate the foreground geometry (the object to be modeled) from 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 a small set of examples (Sec. 5.1) we train a histogram-based classifier of the materials (Sec. 5.1), which is then refined by finding an optimal labeling of pixels using graph cuts (Sec. 5.3). Then, a final (optional) user-assisted revision can be performed using the same tool used for manual segmentation.

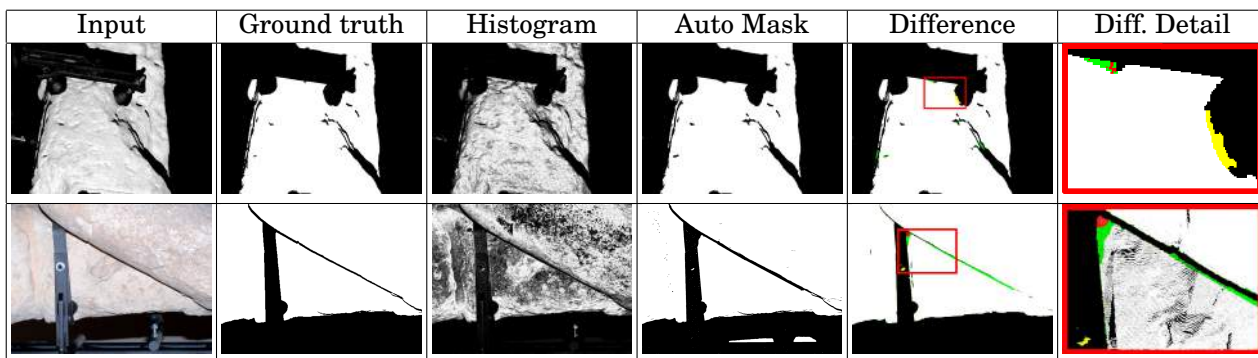| Input	Ground truth	Histogram	Auto Mask	Difference	Diff. Detail

Fig. 5.   **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.

5.1   Manual segmentation

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 the user to visually browse images/scans in the acquisition database, visually select a small subset (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 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, our tool supports interactive grab-cut segmentation [Rother et al. 2004] in which the user selects a bounding box of the foreground object to initialize the segmentation method.

5.2   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 the 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. On the other hand, for the color images 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, thus 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 a foreground probability.

It is worth noting that whether the histogram is computed from foreground or clutter data is not important; as long as the rest of the pipeline is consistent the only constraint is the aforementioned assumption that the two appearances are reasonably well separable.

5.3 Graph cut segmentation

As illustrated in Fig. 5, 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 [Ruzon and Tomasi 2000; Chuang et al. 2001].

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 [Rother et al. 2004] segmentation algorithm using a Gaussian Mixture Model with 5 components per region and estimating segmentation using min-cut. As illustrated in Fig. 5, column 4, the process produces tight and well regularized segmentation masks.

5.4 Morphological post-pass

Since masks must be conservative, especially at silhouette boundaries where a small misalignment is likely to occur, we found it useful to post-process the masks using morphological filters. After denoising the mask using a small median filter (5x5 in this work) we perform an erosion of the mask using an octagon kernel (4x4 in this case). This step has the effect of eliminating small isolated spots and to avoid being too close to the silhouettes, and the removal does not create problems given the large overlap of images that cover the model.

6. DATA CONSOLIDATION AND EDITING

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 the geometry and color information before further processing.

The remaining processing steps, optionally including color mapping (see Sec. 7), are performed by using an editable out-of-core structure based on a forest of octrees. The structure consists of 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 the system of Wand et al. [2007], the data points are stored unmodified in the leaf nodes of the octrees while inner nodes provide grid-quantization-based multiresolution representations. In contrast to previous work, we dynamically decide whether to subdivide or merge nodes based 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
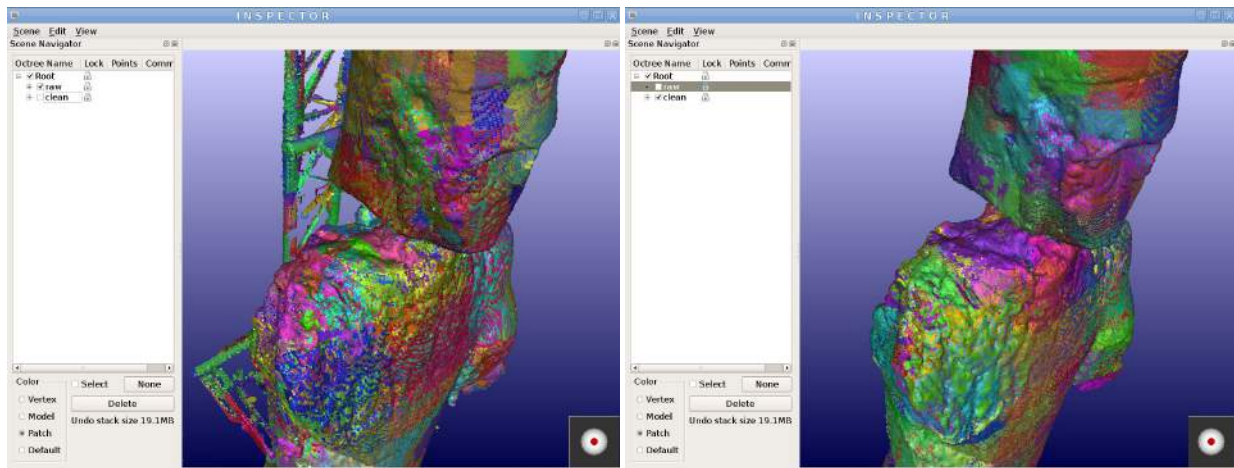
Fig. 6.   **Interactive inspection and editing.** Left: all captured range maps imported into the out-of-core scene structure without applying masks. Right: automatic masking removes most – if not all – clutter. Editing can thus be limited to fixing fine details.

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 for each one (see Fig. 6). The initial transformation of each range map is the coarse alignment transformation computed during the scanning operation, 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 and labeling semantically relevant parts). Alignment and editing operations are applied to the structure using an interactive editor. Alignment is performed by applying to (a subsets of) the range scans a global registration method based on the scalable approach by Pulli et al. [1999], using GPU-accelerated KNN to speed-up local pairwise ICP [Cayton 2012]. After satisfactory global alignment multiple octrees can optionally be merged together. Interactive editing is performed on the structure using a select-and-apply paradigm that supports undo history. At each application of a modifier (e.g., point deletion) modified samples are moved to a temporary out-of-core structure (a memory-mapped array) prior to modification. By associating the array to the undo list we are able to perform reversible editing.

The final colored point clouds can then be further elaborated to produce seamless surface models. To produce consolidated models represented as colored triangle meshes, the most common surface representation, there exist a number of state-of-the-art approaches [Cuccuru et al. 2009; Manson et al. 2008; Kazhdan et al. 2006]. We have adopted the recent screened Poisson surface reconstruction approach [Kazhdan and Hoppe 2013] which produces high-quality watertight reconstructions by incorporating input points as interpolating constraints, while reasonably infilling missing areas based on smoothness priors. Because the Poisson approach does not handle colored surfaces we incorporate color in a post-processing phase (see Sec. 7.1).

7. COLOR CORRECTION, MAPPING, AND INPAINTING

The color attribute is obtained first by projecting masked photos onto the 3D model reference frame and then performing seamless texture blending of those images onto the surface [Pintus et al. 2011c; Pintus et al. 2011a]. In contrast to previous work, we blend and map images directly to the out-of-core structure and perform color correction starting from captured RAW images during the mapping operation.

7.1 Streaming color mapping

Our streaming photo blending implementation closely follows our previous work [Pintus et al. 2011c; Pintus et al. 2011a], which we have extended to work on the multiresolution point cloud structure. We associate a blending weight to each point, which is initialized at zero. We then perform photo blending adding one image at a time. For each image, we start rendering the point cloud from the camera point of view, using a rendering method that performs a screen-space surface reconstruction and adapting the point cloud resolution to 1 projected point/pixel. We then estimate a per-pixel blending weight with screen-space GPU operations that take as input the depth buffer as well as the stencil masks (see Pintus et al. [Pintus et al. 2011a] for details). In a second pass on the point cloud, we update the point colors and weights contained in the visible samples of the leafs of the multiresolution structure. Once all images are updated we consolidate the structure recomputing bottom-up the colors and weights of inner node samples using averaging operations. As a result, the colored models are available in our out-of-core multiresolution point cloud structure for further editing.

In order to apply this same process to triangulated surfaces, such as those coming out of the Poisson reconstruction, we import the surface vertices in our octree, perform the mapping, and then map the color back to the triangulated surface. In this manner we can use the spatial partitioning structure for view-frustum and occlusion culling during mapping operations.

Fig. 7. **Color correction and relighting.** Left: original image under flash illumination; note the sharp shadows and uneven intensity. Center-left: projected color with distance-based correction and no synthetic illumination; notice that the flash highlight has been removed, but a darker shade is on the slanted surface. Center-right: projected color with distance-based and orientation-based correction and no synthetic illumination; note the even distribution and good approximation of surface albedo. Right: synthetically illuminated model based on recovered albedo using a different lighting setup.

7.2 Flash color correction

Color correction happens at color blending time during color mapping operations. At this phase of the processing the color mapping algorithm knows the color stored in the corresponding pixel of the RAW image (the *apparent color* $C_{(raw)}$), the camera parameters (camera intrinsic and extrinsic parameters as well as flash position), and the geometric information of the current sample (position and normal stored in the corresponding pixel of the frame buffers used to compute blending weights).

As we verified, the RAW data acquisition produces images where each pixel value is proportional to incoming radiance and the flash light is fairly uniform (see Sec. 4). Therefore, we apply a simple

4:14 • F. Bettio, A. Jaspe Villanueva, E. Merella, F. Marton, E. Gobbetti, and R. Pintus

color correction method based of first principles, similar to the original approach of Levoy et al. [2000], but without per-pixel calibration of flash illumination and camera response. The results presented in this paper assume that the imaged surface is a Lambertian scatterer so that the measured color, for a sufficiently distant illumination, can be approximated for each color channel i by

$$\mathbf{C}^{(i)}_{(raw)} \approx w^{(i)}_{(balance)} \frac{\bar{\mathbf{I}}_{(flash)}}{d^2} \mathbf{C}^{(i)}_{(surface)} (\mathbf{n} \cdot \mathbf{l})^+ \tag{1}$$

where $w^{(i)}_{(balance)}$ is the channel's scale factor used to achieve color balance, $\bar{\mathbf{I}}_{(flash)}$ is flash intensity, $\mathbf{C}^{(i)}_{(surface)}$ is the diffuse reflectance (albedo) of the colored surface sample, $\mathbf{n}$ is the surface sample normal, $\mathbf{l}$ is the flash light direction, and d is the distance of the surface sample from the flash light. As in standard settings, the color balance factors are recovered by taking a single image of a calibration target (Macbeth charts in our case) and using the same setting used for taking the photographs of the artifacts.

Thus, to compute the color of the surface at color mapping time we consider a user-provided desired object distance d_o, using equation 1 to find

$$\mathbf{C}^{(i)}_{(mapped)} = \frac{d^2}{w^{(i)}_{(balance)} d_0^2 (\epsilon + (1 - \epsilon)\tilde{\mathbf{n}} \cdot \mathbf{l})} \mathbf{C}^{(i)}_{(raw)} \tag{2}$$

where ϵ is a small non-null value (0.1 in our case) and $\tilde{n}$ is the smoothed normal (obtained in screen-space with a 5x5 averaging filter. Normal smoothing and dot product offsetting are introduced to reduce the effect of possible over-corrections in the presence of a small misalignment – particularly at grazing angles. It should be noted that, since the Lambertian model does not take into account the roughness of the surface, under flash illumination it tends to over-shadow at grazing angles. As noted by Oren and Nayar [Oren and Nayar 1994], this effect is due to the fact that while the brightness of a Lambertian surface is independent of viewing direction, the brightness of a rough surface increases as the viewing direction approaches the light source direction. The small angular weight correction thus also contributes to reduce the boosting of colors near silhouettes.

Figure 7 shows how a single flash image introduces sharp shadows and uneven intensity based on distance and angle of incidence. Shadows are removed by the color masking process described in Sec. 5 as well as by shadow mapping during color projection. Distance-based correction removes flash highlights but still produces darker shades on slanted surfaces. On the other hand, combining distance-based and orientation-based correction produces a reasonable approximation of surface albedo, thereby enabling a seamless combination of multiple images without illumination-dependent coloring. The resulting colored model can thus be used for synthetic relighting.

7.3 Inpainting

Points of contacts between supports and status generate small holes in the geometry, as well as missing colors due to occlusions and shadows (seen in white in Fig. 8 left). In order to produce final colored watertight models – useful, e.g., for public presentations – it is important to smoothly reconstruct these missing areas. We took the conservative approach of only using smoothness priors to perform geometry infilling and color inpainting, rather than applying more invasive reconstruction methods based on – for instance – non-local cloning methods. This conservative approach has the advantage of not introducing spurious details, while repairing the surface enough to avoid the presence of distracting surface and color artifacts during virtual exploration.

Geometry infilling is simply achieved by applying a Poisson surface reconstruction method [Kazhdan and Hoppe 2013] to reasonably infill missing areas based on smoothness priors (see Fig. 8 center).

ACM Journal on Computing and Cultural Heritage, Vol. 8, No. 1, Article 4, Publication date: August 2014.

On the other hand, color inpainting uses an anisotropic color diffusion process [Wu et al. 2008] implemented in a multigrid framework. We employ a meshless approach that can be applied either to the vertices of the triangle mesh produced by Poisson reconstruction, or directly to a point cloud constructed from it. We assume that each color sample stores the accumulated color and weight coming from color blending. We first extract all points with a null weight, which are those requiring infilling. We then extract a neighbor graph for this point cloud (by edge connectivity when operating on a triangle mesh or by a k-nearest neighbor search, with k=8, when working on point clouds), growing the graph by one layer in order to include colored points in the neighborhood of holes. We then produce a hierarchy of simplified graphs using sequence coarsening operations on the neighbor graph, so that each level has only one quarter of the samples of the finer one. We stop simplification when the number of nodes is small enough (less than 1000 in this paper) or no more simplification edges exists. The graph is used to quickly compute anisotropic diffusion using a multigrid solver based on V-Cycle iterations. Boundary conditions are computed using the samples with non-zero weight that are included in the hierarchy. The anisotropic diffusion equations are then successively transferred to coarser grids by simple averaging and used in a coarse-to-fine error-correction scheme. Once the coarser grid is reached the problem is solved through Gauss-Seidel iterations and the coarse grid estimates of the residual error can be propagated down to the original grid and used to refine the solution. The cycle is repeated a few times until convergence (results in this paper use 10 V-Cycle iterations). As illustrated in Fig 8 right, color diffusion combined with watertight surface reconstruction successfully masks the color and geometry artifacts due to occlusion and shadows. It is important to note that the original colors and geometry are preserved in the database and that these extra colors can be easily removed from presentation when desired.

Fig. 8. **Geometry infilling and inpainting.** Left: the points of contact between the support and the statue generate small holes in the geometry as well as missing colors due to occlusions (in white). Middle: Poisson reconstruction smoothly infills holes. Right: color is diffused anisotropically for conservative inpainting.

8. IMPLEMENTATION AND RESULTS

We implemented the methods described in this paper in a C++ software library and system running on Linux. The out-of-core octree structure is implemented on top of *Berkeley DB 4.8.3*, while *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* [Snavely et al. 2008]. All tests were run on a PC with an 8-core Intel Core i7-3820 CPU (3.60GHz), 64GB RAM and an NVIDIA GTX680 graphics board.

Fig. 9. **Reconstructed Status of the Mont'e Prama complex.** Colored reconstructions of the 37 reassembled statues.

8.1 Acquisition

The scanning campaign covered 37 statues, which were scanned and photographed directly in the museum. Fig. 9 summarizes the reconstruction results. The geometry of all the statues was acquired at a resolution of 0.25mm using a Minolta Vivid 9i in tele mode, resulting in over 6200 640x480 range scans. The number or scans includes a few (wide) coarse scans which fully cover the statue, that were acquired to help with global scan registration. The scanning campaign produced over 1.3G valid position samples. Color was acquired with a Nikon D200 camera mounting a 50mm lens. All photos were taken with a flash in a dark room, with a shutter speed of 1/250s, aperture f/11.0+0.0, and ISO sensitivity 400. A total of 3817 10Mpixel photographs were produced. 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.

8.2 Automatic geometric masking

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

	Samples	(%)
Model points	51.4M	
Clutter points	790K	
False-Positives	240 (486)	0.03 (0.06)
False-Negatives	35757 (11219)	4.53 (1.42)
True False-Negatives	5639 (2746)	0.68 (0.35)

Fig. 10. **Evaluation of automatic geometric masking.** Results of manual segmentation of a single statue (*Guerriero3*) compared with the results produced by automatic masking. We report the number of range map samples labeled as model ("Model points") and clutter ("Clutter points") in the ground truth dataset, the samples erroneously labeled as statue ("False-positives") or clutter ("False-negatives") 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-negatives"). Values between parentheses compare the manually refined and the ground-truth datasets, instead of the purely automatic method. Percentages are computed with respect to the number of clutter points.

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 measured 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 minutes, which is relatively similar to the time required for 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.

The automatic segmentation process was started by manually segmenting 5 reflectance images using the same editor used for manual segmentation. This training set was used as input for the automatic classifier. The entire process took 9 minutes for the creation of the training set and 6 minutes for the automatic computation of the mask on an 8-core processor. The automatically generated masks were 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 only took 5 hours, excluding manual cleaning, and a total of 13.5 hours including the manual post-process cleanup: this result is a more than ten-fold speed-up with respect to the manual approaches.

The efficiency of the automatic masking method can be seen from the results presented in Fig. 10, which shows the results of the comparison tests between the automatically 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 acquire the geometry of same model 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 the points that were completely missed by the acquisition (*True false negative*) are only 0.68% of the total imaged clutter surface. This check was performed by searching in overlapping scans for samples within a radius of 1mm from each missing sample. Therefore, we can conclude that only a small portion of the surface is missed by the system. Further, Fig. 5 illustrates the position of the missing points; from the images it is easy to see that the points in question are often very sparse or represent small boundary area of the model. Thus, their overall effect on dataset quality is quite limited.

8.3 Automatic color masking

The quality and efficiency of the color masking process was evaluated in a manner analogous to the geometry masking procedure. 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 required 9 minutes to generate the training set, 15 minutes to automatically compute the masks on 8 CPU cores, plus a final 30 minutes for the optional manual post-process cleanup. The speed-up provided by our automated procedure is, again, substantial. The semi-automatic masking process for the entire set of statues only required a total of 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.

	Samples	(%)
Model points	220.5M	
Clutter points	12.1M	
False-Positive	381K (334K)	3.16 (2.77)
False-Negative	263K (253K)	2.18 (2.09)
True False-Negative	8642 (7725)	0.07 (0.06)

Fig. 11. **Evaluation of automatic color masking.** 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 points") and clutter ("Clutter points") in the ground truth dataset, the samples erroneously labeled as statue ("False-positives") or clutter ("False-negatives") by the automatic method, as well as the number of false negative points that really lead to missing data in the combined dataset ("True False-negatives"). Values between parentheses compare the manually refined and the ground-truth datasets, instead of the purely automatic method. Percentages are computed with respect to the number of clutter points.

As illustrated in the table in Fig. 11, the color masking procedures achieves results similar to those obtained by geometry masking. Again, about 95.0% of the samples are labeled correctly. In this case, false-positive samples are 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. Instead, false-negative points are 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 also very beneficial to the automatic masking process.

8.4 Consolidation and coloring

The generated geometry and color masks were used to create digital 3D models of the 37 statues (see Fig. 9). After cleaning, all models were imported into our system based on forests of octrees, which was used for all the 3D editing and color blending. We use lossless compression when storing our hierarchical database, thus achieving an average cost of 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 compared the performance of our system to the state-of-the-art streaming color blender [Pintus et al. 2011a]. Our pipeline required a total of 23 minutes for blending the *Guerriero3* statue; as we already mentioned, the pipeline works directly on the editable representation of the model and includes color correction for flash illumination. On the other hand, the streaming color blender required 2.5 minutes for pre-computing the Morton-ordered sample stream and the culling hierarchy, and 26 minutes for color blending. Therefore, the increased flexibility of our system does not introduce additional overhead in the form of processing time nor does it require additional temporary storage – all while supporting fast turnaround times during iterative editing sessions.

Flash color correction proved to be adequate in our evaluation. It produces visually appealing results without unwanted color variation and/or visible seams between acquisitions (see Fig. 7 for an example). It is important to note that, while no painting results are currently visible on the statues, including natural color considerably adds to the realism of the reconstruction, as demonstrated in Fig. 12.

Fig. 12. **Effect of color mapping.** From left to right: original photograph (boxer 16); virtual reconstruction without color; virtual reconstruction with color.

9. CONCLUSION

We have presented 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 was evaluated in

the context of a real-world large-scale digital reconstruction project concerning 37 statues mounted on supports. It proved capable of notably reducing on-site acquisition times and off-site editing times, while being able to produce good-quality results.

Our method is of general applicability and is able to handle the difficult case of on-site acquisition in a cluttered environment, as exemplified by the problem of acquiring models of statue fragments held in place by exostructures. The technique is based on the standard combination of laser scanning and flash digital photography. One of its main advantages is that it does not need complex illumination setups (just a dark room during color acquisition), thus reducing the time required for color acquisition. Further, it drastically reduces post-processing time when compared to current procedures, thanks to semi-automatic masking of color and geometry and scalable color-corrected color blending. The produced colored 3D models are the starting point for creating a large number of applications based on visual communication, including passive approaches (e.g., still images, video, computer animations) and interactive ones (e.g., multimedia books, web distribution, interactive navigation).

In its current implementation the method assumes that the statue material is easily separable from the unwanted support structure by analyzing reflectance and color, and cannot thus be successfully applied when the statue and the supporting structure are visually indistinguishable. However, 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 – a common case for all ancient artifacts made of stone. This is not an intrinsic limitation of the method, since it could also be applied, just by reversing the masks, if the supporting material is diffuse and homogeneous. Regions of contact between the exostructure and the imaged object cannot obviously be recovered, since they are invisible to the imaging devices. This problem is common to all use-cases involving static supports. Since these parts are small and generally uninteresting, infilling techniques are typically used. However, this issue is orthogonal to this discussion. In the pipeline presented by this work, infilling is performed in a conservative way based on smoothness priors. An interesting avenue of future work would be to efficiently combine the method with more elaborate geometry and color synthesis techniques.

Our work aimed at improving the effectiveness of the standard 3D scanning pipe-line in the case of cluttered 3D models. This pipeline, combining passive and active acquisition methods, is currently among the most commonly applied in the cultural heritage domain, due to the good reliability in a large variety of settings. An important avenue of future work is to evaluate whether alternative image-based approaches based on digital photography can effectively be adapted to the same difficult settings, e.g., by incorporating clutter analysis and flash light modeling in dense reconstruction pipelines.

REFERENCES

ADAN, A. AND HUBER, D. 2011. 3D reconstruction of interior wall surfaces under occlusion and clutter. In *Proc. 3DIMPVT*. 275–281.

ADOBE SYSTEMS INC. 2002. Adobe photoshop user guide.

BALSA RODRIGUEZ, M., AGUS, M., MARTON, F., AND GOBBETTI, E. 2014. HuMoRS: Huge models mobile rendering system. In *Proc. ACM Web3D International Symposium*. ACM Press, New York, NY, USA.

BERNARDINI, F. AND RUSHMEIER, H. 2002. The 3D model acquisition pipeline. In *Computer Graphics Forum*. Vol. 21. 149–172.

BETTIO, F., GOBBETTI, E., MERELLA, E., AND PINTUS, R. 2013. Improving the digitization of shape and color of 3d artworks in a cluttered environment. In *Proc. Digital Heritage*. 23–30.

BORGEAT, L., POIRIER, G., BERALDIN, A., GODIN, G., MASSICOTTE, P., AND PICARD, M. 2009. A framework for the registration of color images with 3d models. *Image Processing (ICIP) Proceedings*, 69–72.

BOYKOV, Y. Y. AND JOLLY, M.-P. 2001. Interactive graph cuts for optimal boundary & region segmentation of objects in N–D images. In *Proc. ICCV*. Vol. 1. 105–112.

CALAKLI, F. AND TAUBIN, G. 2011. SSD: Smooth signed distance surface reconstruction. In *Computer Graphics Forum*. Vol. 30. 1993–2002.

CALLIERI, M., CIGNONI, P., CORSINI, M., AND SCOPIGNO, R. 2008. Masked photo blending: Mapping dense photographic data set on high-resolution sampled 3D models. *Computers & Graphics 32,* 4, 464–473.

CALLIERI, M., DELLEPIANE, M., CIGNONI, P., AND SCOPIGNO, R. 2011. *Digital Imaging for Cultural Heritage Preservation: Analysis, Restoration, and Reconstruction of Ancient Artworks*. CRC Press, Chapter Processing sampled 3D data: reconstruction and visualization technologies, 69–99.

CAYTON, L. 2012. Accelerating nearest neighbor search on manycore systems. In *Proc. IEEE IPDPS*. 402–413.

CHUANG, Y.-Y., CURLESS, B., SALESIN, D. H., AND SZELISKI, R. 2001. A Bayesian approach to digital matting. In *Proc. CVPR*. 264–271.

CORSINI, M., CALLIERI, M., AND CIGNONI, P. 2008. Stereo light probe. *Computer Graphics Forum 27,* 2, 291–300.

CORSINI, M., DELLEPIANE, M., GANOVELLI, F., GHERARDI, R., FUSIELLO, A., AND SCOPIGNO, R. 2012. Fully automatic registration of image sets on approximate geometry. *IJCV*, 1–21.

CUCCURU, G., GOBBETTI, E., MARTON, F., PAJAROLA, R., AND PINTUS, R. 2009. Fast low-memory streaming MLS reconstruction of point-sampled surfaces. In *Graphics Interface*. 15–22.

DAVIS, J., MARSCHNER, S. R., GARR, M., AND LEVOY, M. 2002. Filling holes in complex surfaces using volumetric diffusion. In *Proc. 3DPVT*. 428–441.

DEBEVEC, P. 1998. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. SIGGRAPH*. 189–198.

DEBEVEC, P., HAWKINS, T., TCHOU, C., DUIKER, H.-P., SAROKIN, W., AND SAGAR, M. 2000. Acquiring the reflectance field of a human face. In *Proc. SIGGRAPH*. 145–156.

DELLEPIANE, M., CALLIERI, M., CORSINI, M., CIGNONI, P., AND SCOPIGNO, R. 2009. Flash lighting space sampling. In *Computer Vision/Computer Graphics Collaboration Techniques*. 217–229.

DELLEPIANE, M., CALLIERI, M., CORSINI, M., CIGNONI, P., AND SCOPIGNO, R. 2010. Improved color acquisition and mapping on 3D models via flash-based photography. *ACM JOCCH 2,* 4, Article 9.

FAROUK, M., EL-RIFAI, I., EL-TAYAR, S., EL-SHISHINY, H., HOSNY, M., EL-RAYES, M., GOMES, J., GIORDANO, F., RUSHMEIER, H. E., BERNARDINI, F., ET AL. 2003. Scanning and processing 3d objects for web display. In *3DIM*. 310–317.

FRIEDRICH, D., BRAUERS, J., BELL, A. A., AND AACH, T. 2010. Towards fully automated precise measurement of camera transfer functions. In *IEEE Southwest Symposium on Image Analysis & Interpretation*. IEEE, 149–152.

KASS, M., WITKIN, A., AND TERZOPOULOS, D. 1988. Snakes: Active contour models. *IJCV 1,* 4, 321–331.

KAZHDAN, M., BOLITHO, M., AND HOPPE, H. 2006. Poisson surface reconstruction. In *Proc. SGP*. 61–70.

KAZHDAN, M. AND HOPPE, H. 2013. Screened poisson surface reconstruction. *ACM Trans. Graph 32,* 1.

KIM, S., LIN, H., LU, Z., SUESSTRUNK, S., LIN, S., AND BROWN, M. 2012. A new in-camera imaging model for color computer vision and its application. *IEEE Trans. PAMI 34,* 12, 2289–2302.

KOUTSOUDIS, A., VIDMAR, B., IOANNAKIS, G., ARNAOUTOGLOU, F., PAVLIDIS, G., AND CHAMZAS, C. 2014. Multi-image 3d reconstruction data evaluation. *Journal of Cultural Heritage 15,* 1, 73 – 79.

LAFARGE, F. AND MALLET, C. 2012. Creating large-scale city models from 3D point clouds: a robust approach with hybrid representation. *IJCV 99,* 1, 69–85.

LENSCH, H. P. A., KAUTZ, J., GOESELE, M., HEIDRICH, W., AND SEIDEL, H.-P. 2003. Image-based reconstruction of spatial appearance and geometric detail. *ACM TOG 22,* 2, 234–257.

LEVOY, M., PULLI, K., CURLESS, B., RUSINKIEWICZ, S., KOLLER, D., PEREIRA, L., GINZTON, M., ANDERSON, S., DAVIS, J., GINSBERG, J., ET AL. 2000. The digital michelangelo project: 3d scanning of large statues. In *Proc. SIGGRAPH*. 131–144.

MANSON, J., PETROVA, G., AND SCHAEFER, S. 2008. Streaming surface reconstruction using wavelets. In *Computer Graphics Forum*. Vol. 27. 1411–1420.

MARTON, F., BALSA RODRIGUEZ, M., BETTIO, F., AGUS, M., JASPE VILLANUEVA, A., AND GOBBETTI, E. 2014. IsoCam: Interactive visual exploration of massive cultural heritage models on large projection setups. *ACM Journal on Computing and Cultural Heritage 7,* 2, Article 12.

MCGUINNESS, K. AND O'CONNOR, N. E. 2010. A comparative evaluation of interactive segmentation algorithms. *Pattern Recognition 43,* 2, 434–444.

MORTENSEN, E. N. AND BARRETT, W. A. 1999. Toboggan-based intelligent scissors with a four-parameter edge model. In *Proc. CVPR*. Vol. 2.

MURA, C., MATTAUSCH, O., JASPE VILLANUEVA, A., GOBBETTI, E., AND PAJAROLA, R. 2013. Robust reconstruction of interior building structures with multiple rooms under clutter and occlusions. In *Proc. 13th International Conference on Computer-Aided Design and Computer Graphics*. 52–59.

OREN, M. AND NAYAR, S. K. 1994. Generalization of lambert's reflectance model. In *Proc. SIGGRAPH*. ACM, 239–246.

PINGI, P., FASANO, A., CIGNONI, P., MONTANI, C., AND SCOPIGNO, R. 2005. Exploiting the scanning sequence for automatic registration of large sets of range maps. *Computer Graphics Forum 24,* 3, 517–526.

PINTUS, R. AND GOBBETTI, E. 2014. A fast and robust framework for semi-automatic and automatic registration of photographs to 3d geometry. *ACM Journal on Computing and Cultural Heritage*. To appear.

PINTUS, R., GOBBETTI, E., AND CALLIERI, M. 2011a. Fast low-memory seamless photo blending on massive point clouds using a streaming framework. *ACM JOCCH 4,* 2, Article 6.

PINTUS, R., GOBBETTI, E., AND CALLIERI, M. 2011b. A streaming framework for seamless detailed photo blending on massive point clouds. In *Proc. Eurographics Area Papers*. 25–32.

PINTUS, R., GOBBETTI, E., AND COMBET, R. 2011c. Fast and robust semi-automatic registration of photographs to 3D geometry. In *Proc. VAST*. 9–16.

PULLI, K. 1999. Multiview registration for large data sets. In *Proc. 3D Digital Imaging and Modeling*. 160–168.

REMONDINO, F. 2011. Heritage recording and 3d modeling with photogrammetry and 3d scanning. *Remote Sensing 3,* 6, 1104–1138.

ROTHER, C., KOLMOGOROV, V., AND BLAKE, A. 2004. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM TOG*. Vol. 23. 309–314.

RUZON, M. A. AND TOMASI, C. 2000. Alpha estimation in natural images. In *Proc. CVPR*. 18–25.

SCHEIBLAUER, C. AND WIMMER, M. 2011. Out-of-core selection and editing of huge point clouds. *Computers & Graphics 35,* 2, 342–351.

SNAVELY, N., SEITZ, S. M., AND SZELISKI, R. 2008. Modeling the world from internet photo collections. *IJCV 80,* 2.

WAND, M., BERNER, A., BOKELOH, M., FLECK, A., HOFFMANN, M., JENKE, P., MAIER, B., STANEKER, D., AND SCHILLING, A. 2007. Interactive editing of large point clouds. In *Proc. SPBG*. 37–46.

WIMMER, M. AND SCHEIBLAUER, C. 2006. Instant points: Fast rendering of unprocessed point clouds. In *Proc. SPBG*. 129–137.

WU, C., DENG, J., AND CHEN, F. 2008. Diffusion equations over arbitrary triangulated surfaces for filtering and texture applications. *Visualization and Computer Graphics, IEEE Transactions on 14,* 3, 666–679.

XU, N., AHUJA, N., AND BANSAL, R. 2007. Object segmentation using graph cuts based active contours. *Computer Vision and Image Understanding 107,* 3, 210–224.

ZENG, Y., SAMARAS, D., CHEN, W., AND PENG, Q. 2008. Topology cuts: A novel min-cut/max-flow algorithm for topology preserving segmentation in N–D images. *Computer vision and image understanding 112,* 1, 81–90.

ZHANG, H., FRITTS, J. E., AND GOLDMAN, S. A. 2008. Image segmentation evaluation: A survey of unsupervised methods. *Computer Vision and Image Understanding 110,* 2, 260–280.