Techniques for seamless color registration and mapping on dense 3D models

Ruggero Pintus, Enrico Gobbetti, Marco Callieri, and Matteo Dellepiane

Abstract Today's most widely used 3D digitization approach is a combination of active geometric sensing, mainly using laser scanning, with active or passive color sensing, mostly using digital photography. Producing a seamless colored object, starting from a geometric representation and a set of photographs, is a data fusion problem requiring effective solutions for image-to-geometry registration, and color mapping and blending. This chapter provides a brief survey of the state-of-the-art solutions, ranging from manual approaches to fully scalable automated methods.

Key words: Image-to-geometry registration, color mapping, color blending, 3D reconstruction

1 Introduction

In the Cultural Heritage (CH) domain, digital technologies are transforming the way researchers, archaeologists, and curators work. It is now widely acknowledged that accurate 3D digital models of cultural artifacts built from objective measures are becoming central to CH research, conservation, display and dissemination of knowledge, and have many applications, ranging from virtual restoration to visual communication.

A large variety of active and passive non-invasive systems exists for acquiring, at reasonable costs, a very dense and accurate sampling of both geometric and optical surface properties of real objects. However, today the most widely used

Ruggero Pintus (corresponding author) and Enrico Gobbetti
CRS4, Visual Computing Group, Cagliari, Italy, e-mail: ruggero.pintus@crs4.it, enrico.gobbetti@crs4.it

Marco Callieri and Matteo Dellepiane
ISTI-CNR, Visual Computing Group, Pisa, Italy, e-mail: marco.callieri@isti.cnr.it, matteo.dellepiane@isti.cnr.it

3D digitization approach is a combination of active a scanning device (e.g.,laser triangulation, structured light, time-of-flight, interference) and a digital photography. By using computational techniques, digital surfaces are reconstructed from the scanner-generated range maps, while the color (or material) value sampled in digital photos is transferred onto the surface by registering the photos with respect to the 3D model, and mapping it to the 3D surface using the recovered inverse projections (see Fig. 1). Since early demonstrations of the complete modeling pipeline (e.g., [BR02, LPC+00]), this approach has proven to be particularly well suited to CH digitization, because 3D scanning and photographic acquisition campaigns can be performed quickly and easily, without the need to move objects to specialized acquisition labs. Even though passive purely image-based methods have recently affirmed themselves as a viable (and low-cost) 3D reconstruction technology [Rem11], the pipeline based on active geometric scanning sensors and separate color acquisition still remain a widely used general-purpose approach, mainly because of the higher flexibility and reliability for a wider variety settings (e.g., featureless surfaces) [Rem11, KVI+14].

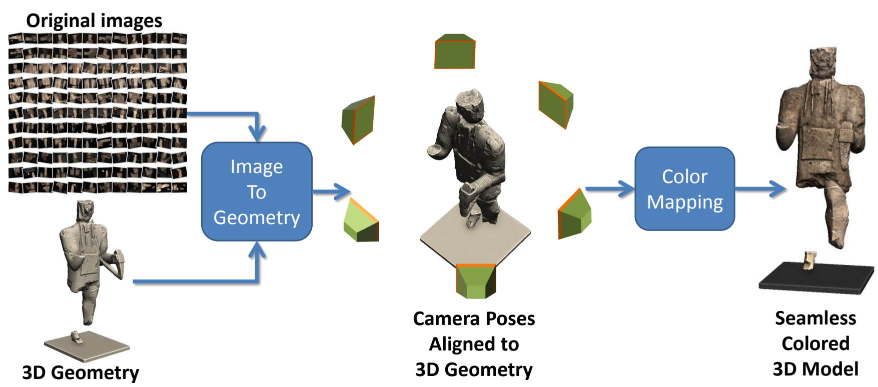

Fig. 1 Pipeline overview. The input data is a 3D geometry and a set of photos of the real object. First, an image-to-geometry alignment step registers the images to the 3D model, by producing a set of corresponding camera poses. By exploiting the resulting link between 3D points and pixels, a color mapping procedure produces a seamless color signal across the surface.

As exemplified in Fig. 1, producing seamless colored objects, starting from a geometric representation and a set of photographs, is a *data fusion* operation that requires the solution of two main problems. First, the color data must be *registered* with geometric data, in order to establish the mapping between appearance information and the 3D surface to which it must be associated. Then, once this registration is known, the appearance, possibly coming from different sources (images) must be *mapped* to the surface to produce a seamless color signal, where each surface point has the single associated color (or material) that best approximates the original one.

In this chapter, we provide a brief overview of the many solutions that have been proposed for effectively solving these two steps of the coloring problem. We will show that, while early works can deal only with small and medium datasets, and are often limited to specific object classes, recent approaches are scalable in terms of image and geometry side, and are mostly (and often completely) automatic, opening the way to interesting innovative applications.

2 Image-to-geometry registration

Image-to-geometry registration takes as input a dense 3D model, representing the shape of the imaged artifact, and a set of N photographs. Depending on the setting, the photographic dataset can cover the complete surface of the 3D object, only a part of it, or a larger area. The goal of registration is to find the intrinsic and extrinsic parameters of the imaging cameras in the reference frame of the 3D object. These parameters are required by the mapping step (see Sec. 3) to transfer color information from the images to the 3D model.

The most straightforward way to get a 2D image data mapped to a geometry is to rigidly fix a calibrated camera to an acquisition device. *Fixed-relative* (or *co-located*) methods [PARD$^+$98, FZ01, FZ03] assume that the pose of the camera is already known and relative to the position of the range scanner (or other source of geometric data). Hence, they have the color intrinsically calibrated with the 3D just after the acquisition, avoiding an explicit image-to-geometry registration task.

On one hand, studies such as Yang et al. [YBS07] show as this kind of approaches, in an hybrid framework that exploits other methods (e.g., *color-based*, see sec. 2.3), can be applied in a number of interesting applications, since the color obtained in the acquisition phase can be used to map other photos with different appearance. Moreover, some range scanners also acquire color information at a quality that might be high enough compared to the application needs. However, even if there are hardware solutions able to sample geometry and color altogether, the color quality and resolution of integrated solutions are often insufficient for CH (e.g., low resolution, poor sensitivity), and rigidly mounting off-axis high-quality cameras often leads to visibility problems. Moreover, in many cases, it is not possible to acquire the 3D and the images at the same time. Sometimes the light condition needed for the photographic dataset must be different from that used during the 3D scanning. For instance, some sites of CH interest can be captured only at night (e.g., Piazza della Signoria in Florence [CDG$^+$13]). There is often an interest in mapping multiple image datasets acquired in different moments to the same geometry, e.g., the two image sets acquired by a professional photographer before and after the restoration of the Michelangelo's David statue [DCPS08]. Moreover, the different parallax of the camera and the range sensor in *fixed-relative* setups possibly results in the presence of occlusions; the color will be thus missed in some surface regions. Finally, while a *co-located* camera might help in aligning additional color images, it will not work with a general 2D signal. In the last decade a vast number of applications in CH uses multi-spectral (MS) acquisitions (e.g., Far Ultra-Violet, Far Infra-Red Thermal images). In contrast to reflectance signal, which is typically from a visible wavelength, these signals might have really different appearance features; this inconsistency depends on variation of transmittance and absorption across the spectrum. The resulting images must be aligned to the geometry without exploiting the presence of the *co-located* camera.

All these reasons lead to a broad range of techniques to enable registration of independently acquired image and 3D data sets. The available solutions include manual registration methods (Sec. 2.1), feature-based approaches that match geometric and image features (Sec. 2.2), color-based techniques that exploit additional color signals from the 3D scanning device (Sec. 2.3), statistical approaches that rely on mutual information (Sec. 2.4), as well as general techniques that convert the hard to solve 2D-3D registration task to a 2D-2D registration obtained via Structure-from-Motion (SfM) followed by a more viable 3D/3D alignment problem (Sec. 2.5).

2.1 Manual methods

The most commong pipeline for mapping photos onto a 3D geometry consists in asking the user to manually select correspondences between pixels in the 2D domain and points in 3D. Hence, the initial step will be a sequence of point-and-clicks with which the user feeds the algorithm. Then, an automatic process will take the burden of refining the initial, manual, rough alignment via an error minimization strategy, and it will finally estimate the camera intrinsic and/or extrinsic parameters. Initially, the solutions proposed focused on the definition of the most usable user interfaces to facilitate this labour-intensive task [DCPS08, FZ03]. Later, reduce focused on speeding-up and reducing manual operations. Borgeat et al. [BPB+09] developed an interface where possible feature matches are highlighted. Another more efficient strategy is to include not only 2D/3D pairs, but also links between pixels in different images [FDG+05]. Moreover, in particular settings, careful planning of image capture might be a solution to speed-up or mostly automate the registration [MK99]. In any case, although these semi-automatic methods are robust due to a continuous support and check by the human on the registration process, they are hardly applicable if the image set starts to grow high over a certain moderate cardinality (e.g., tens or hundreds of photographs).

2.2 Feature-based methods

Feature-based techniques extract visual descriptors from the 3D model and the photographs, and solve the image-to-geometry problem by finding correspondences between them. Although among this class of methods a number of completely automatic frameworks exist, this task is generally very complex, due to the fact that the pure geometry and the object, which is a mix of shape and texture, have very different appearance. For this reason, the works in this area are typically tailored to models with specific geometric characteristics, such as points, lines, circles [SLC+08], rectangles [LS05], edge intensities [NK99], or viewpoint-invariant patches [WCL+08]. Their performance are mostly reliable with 3D models of buildings, which show very sharp edges that can be easily extracted by geometric operator applied to the 3D data, and are a very clear gradient signal in the 2D images. Another class of approaches [KSSS09, SA01] focus on the alignment of maps and/or floor plans (which contain a number of straight lines) with both indoor and outdoor environments acquired with 3D LIDAR scanners, which exhibit several linear edges and planes. Also the relationship between these features could help in finding a good set of camera poses. Liu et al. [LSY+06] rely on the presence of clusters of vertical and horizontal lines, and they exploit the orthogonality constraints to drive the registration algorithm. On one hand they process the geometry to extract parallelepipeds, while they find correspondences between them and rectangles in the photographs. This method results to be very suitable for complex urban scenes.

A strong and robust feature that can be efficiently employed to find matches between images and shapes is the silhouette. A number of methods successfully use it to undertake the image-to-geometry alignment [Low91, BLS92]. The key idea is that silhouette contours are very fast and simple to extract both in the image and in the rendered input 3D model, and their appearance is very similar and independent of object texture. Lensch et al. [LHS00] present a efficient, hardware-accelerated implementation of such algorithms, and propose a robust similarity function. They don't need any manual intervention too.

Unfortunately, these techniques require the whole object to be visible in all input photographs. They also moderately depends on the accuracy of the algorithm for background and foreground separation, which sometimes is refined manually by the user. This latter element might pose some restrictions on the acquisition setup and size of the framed object (e.g., severe limitation for large scale target), reducing the spectrum of possible applications.

2.3 Color-based methods

While *feature-based* methods tries to extract from the geometry a visual cue comparable to the intensity or color signal in the photographs, *color-based* techniques exploits appearance data provided by the acquisition device and, hence, already attached to the geometric primitives (e.g., intensity, color, laser reflectance). In this scenario, the comparison of image descriptors and the computation of matches is naturally more efficient and reliable. The methods of Ikeuchi et al. [IOT+07] and Sequeira and Gonçalves [SG02] rely on the extraction of edges in the reflectance information. This data shows characteristics similar to color images, since these edges come from the discontinuities in the materials across the surface, and they are likely to produce gradients in the same regions both in the color and reflectance signal. Yang et al. [YBS07] uses a co-located setup, which produces a set of colored range maps. Then they extract image descriptors from the color captured by the scanner and by a hand-held camera. The correspondences between these two sources will be employed to compute the extrinsic and intrinsic parameters of the hand-held camera, solving the 2D/3D registration. Wu et al. [WCL+08] present a novel feature called VIP (Viewpoint Invariant Patch), and apply it to both the image-to-geometry problem, and to the registration of two shapes. They normalize the local orientation and viewpoint by performing local texture rectification, and by computing a dominant image gradient. Their method is a robust alignment strategy of images and scenes, even if they are seen from very different viewpoints, or they have a little overlap. This framework is particularly suitable for large-scale and complex scenes when a lot of 2D/3D or 3D/3D registrations have to be undertaken.

2.4 Statistical methods

Statistical image registration methods based on Mutual Information (MI) are extensively used in many application. In the field of image-to-geometry registration, MI is capable of analyzing non-linear correlations and photo-consistency between the intensity signal in the image and some measure present in the target surface, avoiding 3D descriptor extraction. The seminal work of Viola and Wells [VWI97] correlates the image gradient with a particular rendering of the surface normals. Cleju et al. [CS07] improve this work by a global approach based on a stochastic mathematical framework for joint optimization of multiple image-to-geometry registrations.

When available, an important signal for MI computation is the intensity of the reflected laser beam. Williams et al. [WLH+04] present an automatic registration that uses intensity comparisons between color images and geometry texture mapped with laser reflectance. They employ a gradient descent optimization to find the best solution that minimizes the alignment cost function. However, Hantak and Lastra [HL06] show how in some cases the metric used in this approach produces multiple minima within the search space, dropping down the confidence in spotting the right minimum. Hence, they explore different information metrics applied to different environments. This leads them to the conclusion that coupling these metrics with global minimization techniques produce in general better results at the cost of higher computational time. The method of Mastin et al. [MKF09] exploits the statistical dependency that in urban scene correlates the LIDAR elevation with the optical appearance. They measure MI between LIDAR and optical signal, and apply the downhill simplex optimization to find camera poses. They initialize each camera pose from GPS data, and harness the power of graphics hardware to dramatically reduce registration times. In the field of CH, Corsini et al. [CDPS09] take the Viola and Wells's idea of generating an intensity signal from normals, and extend that by including additional surface properties, such as ambient occlusion and reflection directions. The hybrid method of Zheng et al. [ZCS10] strives to integrate, instead, the advantages of *statistical* and *feature-based* registration [ZCS10] by projecting the 3D surface onto 2D normal images. This enables the extraction of local geodesic descriptors, which can be used to perform an initial image-to-image and image-to-geometry registration, which is then refined by using MI and a global stochastic optimization. Recently, Dellepiane and Scopigno [DS13] proposed a MI framework that focuses on registration refinement; they globally tune the camera poses by fitting original color signal with the color mapping result, thus removing artifacts produced by small image mis-alignments (see also Sec. 3.2).

Unfortunately, all these methods are robust only if they are given in advance a good enough initial camera pose(s). Otherwise, it is likely they will not converge to a right solution. Moreover, signals used for correlation are not always available as geometric attributes.

2.5 Multi-view methods

In previous sections the vast majority of already mentioned approaches aim at mapping one single image at a time; conversely, *multi-view* techniques exploit the intrinsic grouping nature of the input images and the relations among them. The fundamental insight that underlies the class of *multi-view* based methods is to transform the original image-to-geometry registration problem into an image-to-image registration step, followed by an alignment between two 3D shapes. In particular, recasting the 2D/3D matching problem into a 3D/3D alignment is made possible by using available and mature Structure from Motion (SfM) approaches, which take an image set and robustly generate a sparse point cloud of the framed object together with a structure of images aligned to it. Hence, registering the original, dense geometry with the one resulting from the SfM step will implicitly solve the image-to-geometry mapping. Methods in this class are more generally applicable, since they do not require such strict constraints as others did. Compared to *feature-based* methods (sec. 2.2) they do not rely on specific features as circles, lines, orthogonal planes. Further, they do not try to find correspondences between images and geometry, but they first find matches between images(SfM step), which is a much fast and simple task, and, then, solve a pairwise geometry alignment, which is a well studied problem too. Apart from the completely automatic *multi-view* solutions, compared to *statistical* approaches (sec. 2.4), the *multi-view* methods reduce manual intervention and completely avoid to use any additional attributes of the geometrical primitives, which are not always accessible, e.g., reflectance, color from the acquisition device, LIDAR elevation or probability of detection (see sec. 2.3 and sec. 2.4). The techniques in the *multi-view* class just need a moderate presence of any type of texture and/or geometrical features, which is a normal, non-limiting requirement of the classical SfM algorithm.

One of the earliest works in this field has been proposed by Zhao et al. [ZNH05], which present a way to register a video onto a point cloud. By using motion stereo, they compute a point cloud and a set of relative camera positions from a continuous (oblique) video. After the alignment of just two frames with a dense point cloud directly obtained from a range sensor, they employ an ICP-based procedure to refine the registration between the sparse and the target 3D models. This framework proved to be very efficient for coloring large-scale digital models in urban scenes. Unfortunately, it is limited to a dense and ordered (video) sequence of highly overlapping images. After this seminal contribution, some past methods have been updated in a hybrid framework that includes some advantages of a *multi-view* processing step. For instance, Liu et al. [LSY$^+$06] extends their previous work on a *feature-based* 2D/3D mapping [LS05], by integrating a SfM geometry and an automatic 3D registration. They use a bunch of 3D scans to build a dense point cloud; in this step they inherit the *feature-based* approach of the previous work to find correspondences between 3D lines in the range images. Then, they compute a sparse geometry from the sequence of the input 2D photographs. They find the rotation, scale and translation that minimize the distance between the two 3D shapes. Stamos et al. [SLC$^+$08] relaxed the line-based orthogonality constraint in order to apply the algorithm not only in urban scenes, but also in complex indoor architectures. Unfortunately, the limit of these hybrid systems to align only 3D models of architectural items remains, so that the spirit of *multi-view* has not been completely exploited to solve a more general problem.

Since relying on ordered sets of images is a too limiting constraint, as in Zhao et al. [ZNH05], a number of methods naturally arose that try to deal with unordered sets of sufficiently overlapping photos. Moreover, acquisition of complex geometrical shapes inevitably leads to a non-uniform image sampling; the capture must be performed adaptively, and the number and density of 2D frames per region depends on the presence and the nature of surface occlusions. This kind of acquisition strategy eventually produces an unordered image set as input for the registration pipeline.

Li and Low [LL09] register unordered image sequences from untracked and uncalibrated cameras onto 3D geometry. Their method is pure *multi-view*. However, they make use of projectors to illuminate the scene of indoor environments with a special light pattern, in order to increase the algorithm robustness in regions with low presence of image features. They manually define a rough registration between the model from SfM and the detailed one. In the refinement stage, they exploit plane segmentation of the input geometry, and employ a cost function that is a mix of object-space error (point-to-plane distance), and image-based re-projection error. A further manual tuning of a heuristic parameter is required to properly weight these two error terms. Similar semi-automatic approaches have been proposed by Banno and Ikeuchi [BI10] and Pintus et al. [PGC11c]. The first presents an accurate texturing method that estimates camera extrinsic parameters by computing a sparse geometry from spherical image stereo data set. They measure the distance between the sparse and dense geometry (from range sensor) by taking into account the uncertainty distribution of 3D in the stereo data. After the rough manual alignment of few images to the dense geometry, Pintus et al. [PGC11c] proposes an image-to-geometry optimization that deform in a non-rigid manner the sparse 3D points together with the poses of each camera independently. The error metric relies only on image-based re-projection error. On the other hand, their method is only applicable to uniformly sampled 3D models.

Due to the well-known problem that SfM techniques reconstruct a sparse 3D point cloud up to a scale factor, aligning it with the geometry from the range sensor requires an affine transformation. All the previous methods ask a minimal user intervention to define a approximate registration, which is the initial affine mapping condition for the final algorithm to converge. The lack of a completely automatic pipeline is the main drawback of these otherwise very robust techniques. To overcome this issue, recently, two completely automatic and orthogonal approaches have been presented, which are capable of dealing with large input datasets, and need neither a object dependent parameter setting, nor, of course, a manual intervention.

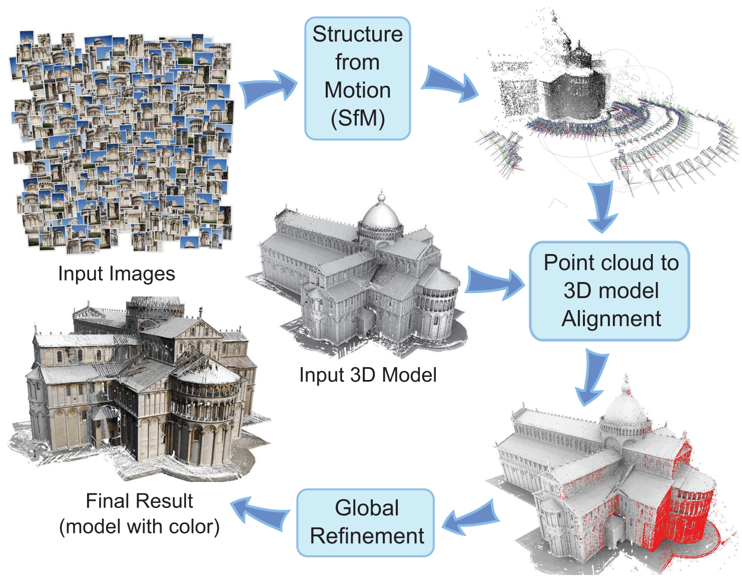

Fig. 2 The multi-image alignment proposed by Corsini et al. [CDG+13], which uses an approach similar to Pintus et al.: [PGC11c, PG15]: the images are used to compute a sparse point cloud using SfM, and the cloud is then aligned to the target geometry, thus obtaining an initial camera orientation, that will then be refined using mutual information. Reprinted from [CDG+13] (With kind permission from Springer Science+Business Media: International journal of computer vision, Fully automatic registration of image sets on approximate geometry, 102, 2013, 91-111, Corsini, M., Dellepiane, M., Ganovelli, F., Gherardi, R., Fusiello, A., Scopigno, R., Fig.1) [CDG+13]

Corsini et al. [CDG+13] propose a novel multi-stage framework for a fully automatic 2D/3D global registration, shown in Figure 2. After the SfM procedure, carried out similarly to Pintus et al. [PGC11c], the generated sparse point cloud is aligned with the dense model by using a modified version of the 4 Point Congruent Set (4PCS) [AMCO08] algorithm. While the 4PCS deals with rigid motion, they extend it to cope with both different scales and unknown amount of overlapping regions. Finally, a global refinement strategy is used based on mutual information, which minimizes the color re-projection error of the images onto the geometry. The method of Pintus et al. [PG15] is another completely automatic solution that performs a coarse image-to-geometry alignment by using a GPU-based global stochastic registration approach to find the optimal affine transformation between the sparse SfM 3D and the dense, original point cloud [AGJ+14]. They propose a robust statistics to manage outliers, and a strategy to automatically estimate a local, adaptive tolerance to deal with non-uniformly sampled 3D geometry. As in Pintus et al. [PGC11c], they refine intrinsic and extrinsic parameters of each camera in a non-rigid manner, by employing a specialized sparse bundle adjustment (SBA) step and a image-based error metric. Compared to Corsini et al. [CDG+13], they do not require a heavy preprocessing to densify the SfM sparse point cloud, since the global registration method employed in the coarse alignment step recovers the globally optimal scale, rotation, and translation alignment parameters using a stochastic algorithm for pairwise affine registration of the coarse point cloud coming from SfM with the fine point cloud representing the geometric model [AGJ+14].

These approaches based on fully automatic 2D/3D registration are more generally applicable and robust. Their main advantage is that they do not put any assumption tailored to the geometry type, size or sampling, being suitable to be applied to small, medium and big objects of different nature. Since they are scalable, they remains suitable even if the image set size grows, opening possibilities for new scenarios and more interesting applications of unattended image-to-geometry alignment and mapping (e.g., automatic web-based services). The most costly and difficult step is the automatic global registration, which can eventually be replaced, in difficult cases, by a simple manual initialization based on few hand-picked correspondences between image and geometry, followed by a local refinement.

3 Mapping Color to 3D geometries

In all the described cases, the result of the image registration algorithms is the calibration and orientation of one or more photos in the reference space of a 3D geometry. While there are many direct uses of a set of calibrated and oriented images, in most cases, the image alignment is is used to map the color information onto the 3D geometry, thus obtaining a colored 3D model.

The recovered camera alignment data is, basically, a perspective projection matrix. By using this perspective projection, it is easy to establish the inverse correlation between the color information on the photos and the 3D geometry. Using this inverse projection, the color can be mapped onto the 3D geometry: this principle is the basis for all the mapping methods.

The case of a single registered image is, then, trivial, since a simple inverse perspective projection will be enough to map the color onto the 3D geometry. The visibility problem, i.e. determine what part of the surface is directly visible by the camera (and thus can be mapped from that specific photo), can be solved numerically (precise but computationally cumbersome) or by exploiting the GPU, using the well-known shadow mapping algorithm or other specialized shaders. The visibility problem is more complex in the case of point-clouds; since a direct shadow mapping would not work, Pintus et al. [PGC11a] use a GPU accelerated screen-space operators, able to fill the sparse depth-map produced from the point cloud.

The problems arise when there is more than one photo that has to be mapped on the 3D surface. While the basic inverse projection is still used to map color from the photos to the surface, the issue is now what happens where more than one input photo maps on the same area of the 3D surface. It is necessary to determine, among the available registered photos mapping in a specific area of the 3D surface, which is the correct color information that has to be projected and/or solve the "blending problem", i.e. combine the available contribution to have a more continuous color mapping. This topic will be covered in Section 3.1.

A common problem of the color mapping process is the fact that, no matter how precise is the image-to-geometry alignment, it will never be perfect, or perfectly coherent between images. By applying local or local-global optimization of the cameras and projected color information it is possible to overcome the classical artifacts of ghosting and blurring. These methods will be covered in Section 3.2.

Finally, another problem in the color projection from multiple registered cameras is the presence of lighting and photometric incoherence between the input photos, or the presence of view-dependent artifact (like highlights) that changes their relative position from a photo to another. While some of these issues can be reduced or removed by the blending process, in literature there are specific methods which can be applied to solve these situations (see Section 3.3).

3.1 Projecting and blending

An issue that is orthogonal to the mapping method is where to store the mapped color: a family of solutions use (or creates) a surface parametrization and store the mapped color in the texture map; a different approach is encode color per-vertex; finally, some algorithms may work regardless of the final color encoding.

The earliest works dealing with the mapping and blending problem tried to avoid entirely the computation of a static color mapping, but opted for a view-dependent texture mapping, dynamically creating a color mapping as the object was rendered. This was done, for example, by using a per-triangle computation of the best photo, and perspective mapping directly on the graphics hardware, like Debevec et al. [DYB98]; or by selecting the relevant images for a specific view, and then mixing multiple images using a weighting scheme [PARD$^+$98], blending the different contribution in a continuous mapped image.

Although it is able to provide good results, view-dependent texture mapping is not much used anymore, and the more modern color mapping methods all rely on the creation of a static mapping.

In these early examples, however, it is already possible to see the two opposite strategies that can be used when multiple photos are projecting on the same element of the 3D geometry that has to be mapped (a triangle, in texture mapping, or a vertex for per-vertex color mapping or in point clouds):

- among the available contributions, the "best" photo for that specific element will be chosen, according to some criteria, and mapped on the surface

- all the available contributions will be blended, using some form of weighting, and the interpolated result will be mapped on the surface

The advantage of the fist strategy is that the detail of the mapping remains at the same level of the input photos, but there will be discontinuities (color shift and details misalignments) in the borders between areas assigned to different input photos. On the other hand, the color mapped using the blending strategy will produce a much smoother result, without discontinuities, but at the price of a less sharp appearance, and the presence of ghosting and blurring if the input photos are not perfectly aligned to the 3D geometry. The various methods available in literature try to cover the spectrum between these two extreme, providing solution that combines these two basic strategies.

The "best" contribution is one of the oldest strategies in color mapping, it is still used a lot in both open and commercial tools [WJP08], especially when the final result of the mapping has to be a texture map. As the registered photos define a perspective projection, in most cases, the projective transformations from the 3D surface to the input photos results in a natural and usable parametrization. Some mapping methods [LHS01, RCMS02, CCS02] build a parametrization for the 3D surface by dividing the triangulated mesh in patches, each one assigned to the "best" input image,parameterized using the perspective projection of that input image and packed in a texture map. Lensch et al. [LHS01] and Rocchini et al. [RCMS02] filter the border between adjacent patches, to reduce the discontinuity, while Callieri et al. [CCS02] apply a texture-wide color/luminance correction to uniform the color across the different photos.

Instead of using a simple "best contribution" criteria, but still wanting to directly use fragments of the original images, it is also possible to rely on existing image mosaicing and stitching optimization techniques, derived from the research on 2D image processing. Sinha et al. [SSS+08] use a graph-cut optimization using a Markov Random Field (MRF) minimization framework to define, at a per-texel level, the most suitable photographic source to minimize incoherence, and then apply a Poisson blending to clear the remaining discontinuities. Lempitsky and Ivanov [LI07] also use this kind of MRF minimization, but at a per-triangle level, while Gal et al. [GWO+10] also introduce a parametrization shifting order to corrects image registration errors.

A different strategy may be to separate the high and low frequencies of the input images, and use different strategies to combine different bands, to obtain simultaneously a smooth mapping, but with sharp details. Baumberg [Bau02] separates the input images in two bands: in every given area of the 3D model all the low-frequency components are blended using weights, while the only the "best" high-frequency details is mapped. This solution is also used by Allène et al. [APK08] and by Pintus et al. [PGC11a] on point-clouds.

Among the methods blending all available contributions, performing a weighed sum is the most used way to combine contributions [PARD+98, BMR01, CCCS08, PGC11a, PGC11b, BJM+15]. All use weighted blending. The weights used in these and other works are many, and are related to different aspects of the dataset. They may be related to geometry (depth discontinuities), camera (image sharpness, low quality on borders, image discontinuities), geometry-camera relationship (orthogonality of the photo with respect to the surface, distance from the object). A useful feature of weighting schemes is that it is easy to add new metrics which are specific for a given dataset or acquisition/ reconstruction technology.

In order to combine the different contributions in a single mapping, blending is simple and effective, but prone to blurring. To obtain a more continuous but sharp mapping, it is possible to solve a Poisson equation over the 3D surface [CLB+09], taking in account the input image gradients, building a continuous yet non-blurred color field. Another interesting method is to use a super-resolution approach over the texture space [GC09], resulting in a texture which is more detailed of the input photos.

Even though point-clouds are widely used directly, without converting them to a triangulated surface, there are not many color mapping methods suitable for this kind of data. Pintus et al. [PGC11b, PGC11a] work directly on the point-clouds, mapping color to each vertex, using a weighted blending 2-band strategy, able to work on very large dataset. Conversely, Arikan et al. [APS+14] use the point-clouds to create a series of local meshes, defined by the areas framed by the registered photos, which are then properly stitched during rendering.

Many of the discussed techniques assume that both geometry and images fit in memory, which severely limits the scalability of the methods. Some methods assume, instead, that the 3D model is very small and fits in memory, while the image and mask dataset are stored in out-of-core structure (e.g., [XLL+10, CCCS08]). The approaches that scale to large amounts of color and geometric data, are typically based on some variation of the weighted blending approaches. Callieri et al. [CCCS08] and Bettio et al. [BGMP13, BJM+15] deal with massive models by using two out-of-core structures, one for the model and one for images and masks. By contrast, Pintus et al. [PGC11b, PGC11a] use a streaming framework for efficiently blending images on massive point clouds in a low-memory setting and through memory-coherent operations, providing unprecedented performance (fig. 3). In addition, the latter approach supports adaptive geometry refinement as a function of image content, which is achieved during the blending phase by selectively refining the point cloud in areas with high-frequency image contents (fig. 3, second row).

3.2 Fixing misalignments

Color projection on 3D models is straightforward and convenient way to obtain a visually pleasing result starting from a simple input. Nevertheless, annoying artifacts may appear on the model, regardless of the mapping strategy chosen.

One of major issues is due to the misalignments of images that have to be projected: this leads to the aliasing/blurring of fine details of models, since the images are simply projecting the same data on slightly different positions of the surface. Misalignments are less visible in the case of methods where a single image is chosen for projection: misalignments may be visible in the borders between adjacent patches. Unfortunately, the selection of a single image for each portion of a surface usually works well on

Fig. 3 *First row*: image blending on massive point clouds, the case of the Michelangelo's David statue (0.25mm dataset). It is a 470 million point 3D model with color coming from a 548 Mpixel dataset (67 photos). *Second row*: the effect of the image based adaptive point cloud refinement on final color signal; we present two images of the final colored model, and a comparison between the image blending on the original point cloud and the image blending on the adaptively refined geometry. (With kind permission from Eurographics Association: Eurographics 2011 Area Papers, A Streaming Framework for Seamless Detailed Photo Blending on Massive Point Clouds, 25-32, 2011, Pintus R., Gobbetti E. Callieri M., Fig. 7,8,9) Scopigno, R., fig.1) [PGC11b]

simple cases, but when the number of images rises, it leads to very fragmented color, where artifacts are visible again.

The misalignment issues are only partially inherent to the registration and projection methods. They usually come from the quality of input data, in particular:

- The camera parameters: it's important to stress that image registration is obtained by applying a camera model to fit real cameras. This means that in particular cases there will be no possibility to achieve a perfect alignment, since the model is not able to perfectly describe the camera.
- The quality of 3D model: if the geometric representation of the object is not accurate, a perfect alignment will not be possible
- The quality of images: similarly as above, if the quality of 2D data is low, even semi-automatic methods may fail. Moreover, low quality images may strongly concur in blurred and aliased details.

Given that the input data are not perfect, fixing the misalignments may not be an issue related to a better estimation of camera parameters. Some methods take this into account to get the best color: for example Lempitsky and Ivanov [LI07] inserted in their minimization function a term that should reduce misalignments artifacts. All these methods are based on a quite strong assumption: that the 3D model is accurate enough to allow an accurate color projection.

More recently, Dellepiane and Scopigno [DS13] proposed a global refinement method to reduce misalignments. The result of any image-to-geometry registration method can be the input of this approach. The method is based on mutual information. The camera parameters of each image are refined by trying to fit the color projected by all the other images on the 3D model. In this way the error is globally distributed, and fine details are better preserved. This works regardless of the quality of the 3D model, and only camera parameters are changed.

The main way to achieve an optimized alignment, though, is to be able to modify all the elements of the projection dataset. In the following, we review the recently proposed methods.

The quality of color projection can be improved by modifying all the elements of the projection set: camera parameters, geometry, images. Recently the problem came out after the rise of multi-view-stereo matching techniques. In this case, similarly to the work by Gal et al. [GWO^{+}10], the big advantage comes from the fact that geometry and color are produced starting from the same input: the images. Given this, the geometry can be modified to better fit the images. Alternatively, taking into account the matching among images enables to preserve detail. The work by Waechter et al. [WMG14] shows that this can be used to obtain large scale 3D texturing.

In a more general case, geometry and images are acquired in a different moment, so it could be harder to handle them. Pintus et al [PGC11a] used a frequency analysis and the increase in data density to improve the projection on point clouds. In the case of triangulated surfaces, a possibility is to warp images so that they will project the same color information in the same part of the surface. Eisemann et al. [EDDM^{+}08] used it to fit textures to body models. They propose a GPU implementation of the method that reaches real-time performances. Hence the algorithm can be employed not only in off-line image-to-geometry registration frameworks, but also in image-based visualizations.

Dellepiane et al. [DMC^{+}12] extended the concept to more complex cases: for each portion of the surface, an image is chosen as the “leading” one. All the other images are warped using optical flow, in order to project the same color detail as the leading one. The border between adjacent parts is handled by mixing the warping in order to avoid discontinuities. This approach is able to put together the advantages of the “best photo” approach (high quality fine details) with the advantages of blending approaches (coherent color, no discontinuities).

More recently, Zhou and Koltun [ZK14] proposed a method to map color on data coming from depth cameras. The method puts together the camera parameters refinement and the flow based approach in a unique system to be optimized. Given the complexity of the problem, the system is solved in an iterative way (by refining cameras and flow one at a time), but the results are more than satisfying.

3.3 Fixing color and lighting issues

The other issue when dealing with color mapping from images is that the color is projected “as it is”, without any previous knowledge of illumination conditions. This leads to the fact that illumination artifacts (highlights, shadows) are projected on the model.

The proper way to be able to remove artifacts would be to acquire the color under controlled light conditions (i.e. by knowing the position and intensity of the main sources of light in the scene). This is quite impractical during an on-site acquisition, or when working outdoors.

A possible approach is to analyze the images to detect and remove artifacts, but the purely image based approaches are prone to errors and false positives [XYWZ05, GDH11]. Another possible approach is to take advantage of the fact that more than one image will cover each part of the surface. In this case, some artifacts can be detected as outliers, and removed. These methods work reliably only when several images of the same portion are available [FFGG+10]. This is not always possible, especially when high quality images are needed. Hence, an alternative way is to have a better knowledge of the lighting environment.

That being said, except for laboratory conditions and object of small size, it’s rarely possible to setup a controlled light environment. Alternatively, it’s possible to try to estimate it: once the 3D model, the camera parameters and the light positions are known, the lighting artifacts can be easily detected and removed. The lighting environment can be acquired using ad-hoc devices [CCC08]. An alternative solution can be to use the object itself as a light probe. The analysis of the behavior of artifacts (i.e. the change in position of highlights) can be used to estimate the position of the main light sources in the scene. Palma et al. [PCDS12] used video sequences to estimate the illumination environment, and also extract a simplified description of the material of an object.

Another source of light that could be possibly estimated is the sky. Haber et al. [HFB+09] analyzed the sky and the appearance of an object to estimate the sky-dome environment, and relight images. Dellepiane et al. [DBS10] estimated the sun position (in an automatic or semi-automatic way) to remove hard shadows.

Finally, an interesting example of controlled but easy to use light setup is FlISS [DCC+10], where the integrated flash light of the digital camera was used. This type of light has the advantage of being always at the same position w.r.t. the camera, and to provide a dominant lighting in the scene. Hence, it’easy to detect and remove artifacts, and it’s possible to obtain a partially calibrated color after a calibration step to be performed once-in-a-lifetime. Starting from the consideration 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 [KLL+12], Bettio et al. [BGMP13, BJM+15] 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, taking into account distance and surface orientation.

4 Conclusions

We have provided an overview of many different approaches that have been proposed for effectively solving image-to-geometry registration and color mapping and blending. Despite the fact that in recent years a lot of powerful pure vision-based methods have been developed and widely used to reconstruct both geometry and color, the interest in aligning 2D maps onto 3D surfaces remains very high. This is due to the fact that different types of signals, or data acquired at different times, still need to be registered. Effectively solving 2D to 3D mapping and fusion is therefore of high importance for many application domains, including cultural heritage.

While the various methods can be mixed and matched to create a pipeline for aligning and a separate one for producing seamless colored objects after alignment, we have seen that solutions of these two problems are often intertwined. In particular, a recent trend has been to perform fine registration steps during the color blending passes to produce seamless models. Moreover, while early techniques were mostly manual, most modern methods tend towards semi-automatic or automatic behaviors.

In CH practice it is however still common to introduce a human in the loop, in order to control the fine behavior of algorithms. However, as data size is increasing, since acquisition becoming fast and inexpensive, this fine control is increasingly becoming impractical, and we expect that automated methods will increasingly become standard even in this application field. The increasing size of datasets is also forcing algorithm development to tackle scalability issues using specialized techniques. For this reason, while early methods, especially for the mapping step, required global operation with all data loaded into memory, modern method strive to schedule localized operations, so as to minimize peak memory needs and I/O bandwidth.

The availability of modern fully scalable and mostly automatic techniques opens the door to innovative applications. In particular, cloud solutions based on web services are emerging for performing heavy operations, with lean user interfaces often working on mobile graphics devices.

While in this work we have focused mostly on the registration of color images onto the geometry, the technique presented here are the basis for managing a number of multimodal fusion operations. Recent trends highlight in particular the need of a enhancing geometric scans not only with the local body color of an object, but also with a detailed representation of the material and the appearance properties. In many cases, the rationale of the techniques reviewed here is still valid to deal with this more complex data, with limited modifications to implement the mapping and fusion (e.g., different descriptors for automatic alignment). On the other hand, sometimes micro-scale data acquired as 2D maps is also providing information at the geometric level. For instance, Reflectance Transformation Imaging acquisitions [MGW01] make it is possible to compute high resolution normal maps, which, once registered, can be employed to refine the initial geometry [NRDR05, BFR14]. While this is usually done in a post-pass, a stronger coupling of the various phases of registration, mapping, and enhancement could lead to better and faster results, especially when dealing with massive amounts of data. An interesting avenue of future work would thus be to exploit the power of both approaches into a common scalable framework, in order to create highly detailed objects with higher resolution geometries combined with detailed digital models of materials.

Acknowledgements This work is partially supported by the EU H2020 Program under the Scan4Reco project (665091) funded under H2020-EU.3.6.3. We also acknowledge the contribution of Sardinian Regional authorities for Visual Computing activities (projects VIGEC and HELIOS).

References

AGJ+14. Marco Agus, Enrico Gobbetti, Alberto Jaspe Villanueva, Claudio Mura, and Renato Pajarola. Soar: Stochastic optimization for affine global point set registration. In *Proc. 19th International Workshop on Vision, Modeling and Visualization (VMV)*, pages 103–110, October 2014.

AMCO08. Dror Aiger, Niloy J Mitra, and Daniel Cohen-Or. 4-points congruent sets for robust pairwise surface registration. In *ACM Transactions on Graphics (TOG)*, volume 27, page 85. ACM, 2008.

APK08. Cédric Allène, Jean-Philippe Pons, and Renaud Keriven. Seamless image-based texture atlases using multi-band blending. In *Proc. ICPR*, pages 1–4. IEEE, 2008.

APS+14. Murat Arikan, Reinhold Preiner, Claus Scheiblauer, Sabina Jeschke, and Manuel Wimmer. Large-scale point-cloud visualization through localized textured surface reconstruction. *Visualization and Computer Graphics, IEEE Transactions on*, 20(9):1280–1292, 2014.

Bau02. Adam Baumberg. Blending images for texturing 3d models. In *BMVC*, volume 3, page 5. Citeseer, 2002.

BFR14. S. Berkiten, X. Fan, and S. Rusinkiewicz. Merge2-3D: Combining multiple normal maps with 3d surfaces. In *2nd International Conference on 3D Vision*, volume 1, pages 440–447, Dec 2014.

BGMP13. Fabio Bettio, Enrico Gobbetti, Emilio Merella, and Ruggero Pintus. Improving the digitization of shape and color of 3d artworks in a cluttered environment. In *Proc. Digital Heritage*, pages 23–30, October 2013. Best Paper Award.

BI10. Atsuhiko Banno and Katsushi Ikeuchi. Omnidirectional texturing based on robust 3d registration through euclidean reconstruction from two spherical images. *Computer Vision and Image Understanding*, 114(4):491–499, 2010.

BJM+15. Fabio Bettio, Alberto Jaspe Villanueva, Emilio Merella, Fabio Marton, Enrico Gobbetti, and Ruggero Pintus. Mont'e Scan: Effective shape and color digitization of cluttered 3d artworks. *Journal on Computing and Cultural Heritage (JOCCH)*, 8(1):4:1–4:23, 2015.

BLS92. Lionel Brunie, Stéphane Lavallée, and Richard Szeliski. Using force fields derived from 3d distance maps for inferring the attitude of a 3d rigid object. In *Computer VisionECCV'92*, pages 670–675. Springer, 1992.

BMR01. Fausto Bernardini, Ioana M Martin, and Holly Rushmeier. High-quality texture reconstruction from multiple scans. *Visualization and Computer Graphics, IEEE Transactions on*, 7(4):318–332, 2001.

BPB+09. Louis Borgeat, Guillaume Poirier, Angelo Beraldin, Guy Godin, Philippe Massicotte, and Michel Picard. A framework for the registration of color images with 3d models. *Image Processing (ICIP) Proceedings*, pages 69–72, 2009.

BR02. Fausto Bernardini and Holly Rushmeier. The 3D model acquisition pipeline. In *Computer Graphics Forum*, volume 21, pages 149–172, 2002.

CCC08. Massimiliano Corsini, Marco Callieri, and Paolo Cignoni. Stereo light probe. In *Computer Graphics Forum*, volume 27, pages 291–300. Wiley Online Library, 2008.

CCCS08. Marco Callieri, Paolo Cignoni, Massimiliano Corsini, and Roberto Scopigno. Masked photo blending: Mapping dense photographic data set on high-resolution sampled 3d models. *Computers & Graphics*, 32(4):464–473, 2008.

CCS02. Marco Callieri, Paolo Cignoni, and Roberto Scopigno. Reconstructing textured meshes from multiple range rgb maps. In *VMV*, pages 419–426, 2002.

CDG+13. Massimiliano Corsini, Matteo Dellepiane, Fabio Ganovelli, Riccardo Gherardi, Andrea Fusiello, and Roberto Scopigno. Fully automatic registration of image sets on approximate geometry. *International journal of computer vision*, 102(1-3):91–111, 2013.

CDPS09. Massimiliano Corsini, Matteo Dellepiane, Federico Ponchio, and Roberto Scopigno. Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties. In *Computer Graphics Forum*, volume 28, pages 1755–1764. Wiley Online Library, 2009.

CLB+09. Ming Chuang, Linjie Luo, Benedict J Brown, Szymon Rusinkiewicz, and Michael Kazhdan. Estimating the laplace-beltrami operator by restricting 3d functions. In *Computer Graphics Forum*, volume 28, pages 1475–1484. Wiley Online Library, 2009.

CS07. Ioan Cleju and Dietmar Saupe. *Stochastic optimization of multiple texture registration using mutual information*. Springer, 2007.

DBS10. Matteo Dellepiane, Luca Benedetti, and Roberto Scopigno. Removing shadows for color projection using sun position estimation. In *Proc. 11th International conference on Virtual Reality, Archaeology and Cultural Heritage*, pages 55–62. Eurographics Association, 2010.

DCC+10. Matteo Dellepiane, Marco Callieri, Massimiliano Corsini, Paolo Cignoni, and Roberto Scopigno. Improved color acquisition and mapping on 3d models via flash-based photography. *Journal on Computing and Cultural Heritage (JOCCH)*, 2(4):9, 2010.

DCPS08. Matteo Dellepiane, Marco Callieri, Federico Ponchio, and Roberto Scopigno. Mapping highly detailed colour information on extremely dense 3d models: the case of david's restoration. In *Computer Graphics Forum*, volume 27, pages 2178–2187. Wiley Online Library, 2008.

DMC+12. Matteo Dellepiane, Ricardo Marroquim, Marco Callieri, Paolo Cignoni, and Roberto Scopigno. Flow-based local optimization for image-to-geometry projection. *Visualization and Computer Graphics, IEEE Transactions on*, 18(3):463–474, 2012.

DS13. Matteo Dellepiane and Roberto Scopigno. Global refinement of image-to-geometry registration for color projection on 3d models. In *Digital Heritage International Congress*, pages 39–46. The Eurographics Association, 2013.

DYB98. Paul Debevec, Yizhou Yu, and George Borshukov. *Efficient view-dependent image-based rendering with projective texture-mapping*. Springer, 1998.

EDDM+08. Martin Eisemann, Bert De Decker, Marcus Magnor, Philippe Bekaert, Edilson De Aguiar, Naveed Ahmed, Christian Theobalt, and Anita Sellent. Floating textures. In *Computer Graphics Forum*, volume 27, pages 409–418. Wiley Online Library, 2008.

FDG+05. Thomas Franken, Matteo Dellepiane, Fabio Ganovelli, Paolo Cignoni, Claudio Montani, and Roberto Scopigno. Minimizing user intervention in registering 2d images to 3d models. *The Visual Computer*, 21(8-10):619–628, 2005.

FFGG+10. Jan-Michael Frahm, Pierre Fite-Georgel, David Gallup, Tim Johnson, Rahul Raguram, Changchang Wu, Yi-Hung Jen, Enrique Dunn, Brian Clipp, Svetlana Lazebnik, et al. Building rome on a cloudless day. In *Computer Vision–ECCV 2010*, pages 368–381. Springer, 2010.

FZ01. Christian Fruh and Avideh Zakhor. Fast 3d model generation in urban environments. In *Proc. MFI*, pages 165–170. IEEE, 2001.

FZ03. Christian Fruh and Avideh Zakhor. Constructing 3d city models by merging aerial and ground views. *Computer Graphics and Applications, IEEE*, 23(6):52–61, 2003.

GC09. Bastian Goldluecke and Daniel Cremers. Superresolution texture maps for multiview reconstruction. In *Proc. ICCV*, pages 1677–1684. IEEE, 2009.

GDH11. Ruiqi Guo, Qieyun Dai, and Derek Hoiem. Single-image shadow detection and removal using paired regions. In *Proc. CVPR*, pages 2033–2040. IEEE, 2011.

GWO$^+$10. Ran Gal, Yonatan Wexler, Eyal Ofek, Hugues Hoppe, and Daniel Cohen-Or. Seamless montage for texturing models. In *Computer Graphics Forum*, volume 29, pages 479–486. Wiley Online Library, 2010.

HFB$^+$09. Tom Haber, Christian Fuchs, Philippe Bekaer, Hans-Peter Seidel, Michael Goesele, Hendrik P Lensch, et al. Relighting objects from image collections. In *Proc. CVPR*, pages 627–634. IEEE, 2009.

HL06. Chad Hantak and Anselmo Lastra. Metrics and optimization techniques for registration of color to laser range scans. In *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, pages 551–558. IEEE, 2006.

IOT$^+$07. Katsushi Ikeuchi, Takeshi Oishi, Jun Takamatsu, Ryusuke Sagawa, Atsushi Nakazawa, Ryo Kurazume, Ko Nishino, Mawo Kamakura, and Yasuhide Okamoto. The great buddha project: Digitally archiving, restoring, and analyzing cultural heritage objects. *International Journal of Computer Vision*, 75(1):189–208, 2007.

KLL$^+$12. Seon Joo Kim, Hai Ting Lin, Zheng Lu, Sabine Suesstrunk, S. Lin, and M. S. Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Trans. PAMI*, 34(12):2289–2302, 2012.

KSSS09. Ryan S Kaminsky, Noah Snavely, Steven M Seitz, and Richard Szeliski. Alignment of 3d point clouds to overhead images. In *Proc. CVPR Workshops*, pages 63–70. IEEE, 2009.

KVI$^+$14. Anestis Koutsoudis, Bla Vidmar, George Ioannakis, Fotis Arnaoutoglou, George Pavlidis, and Christodoulos Chamzas. Multi-image 3d reconstruction data evaluation. *Journal of Cultural Heritage*, 15(1):73 – 79, 2014.

LHS00. Hendrik Lensch, Wolfgang Heidrich, and H-P Seidel. Automated texture registration and stitching for real world models. In *Proc. Pacific Graphics*, pages 317–452. IEEE, 2000.

LHS01. Hendrik PA Lensch, Wolfgang Heidrich, and Hans-Peter Seidel. A silhouette-based algorithm for texture registration and stitching. *Graphical Models*, 63(4):245–262, 2001.

LI07. Victor Lempitsky and Denis Ivanov. Seamless mosaicing of image-based texture maps. In *Proc. CVPR*, pages 1–6. IEEE, 2007.

LL09. Yunzhen Li and Kok-Lim Low. Automatic registration of color images to 3d geometry. In *Proc. CGI*, pages 21–28. CGS, 2009.

Low91. David G. Lowe. Fitting parameterized three-dimensional models to images. *IEEE transactions on pattern analysis and machine intelligence*, 13(5):441–450, 1991.

LPC$^+$00. Marc Levoy, Kari Pulli, Brian Curless, Szymon Rusinkiewicz, David Koller, Lucas Pereira, Matt Ginzton, Sean Anderson, James Davis, Jeremy Ginsberg, et al. The digital michelangelo project: 3d scanning of large statues. In *Proc. SIGGRAPH*, pages 131–144, 2000.

LS05. Lingyun Liu and Ioannis Stamos. Automatic 3d to 2d registration for the photorealistic rendering of urban scenes. In *Proc. CVPR*, volume 2, pages 137–143. IEEE, 2005.

LSY$^+$06. Lingyun Liu, Ioannis Stamos, Gene Yu, George Wolberg, and Siavash Zokai. Multi-view geometry for texture mapping 2d images onto 3d range data. In *Proc. CVPR*, volume 2, pages 2293–2300. IEEE, 2006.

MGW01. Tom Malzbender, Dan Gelb, and Hans Wolters. Polynomial texture maps. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 519–528. ACM, 2001.

MK99. Kenji Matsushita and Toyohisa Kaneko. Efficient and handy texture mapping on 3d surfaces. In *Computer Graphics Forum*, volume 18, pages 349–358. Wiley Online Library, 1999.

MKF09. Andrew Mastin, Jeremy Kepner, and J Fisher. Automatic registration of lidar and optical images of urban scenes. In *Proc. CVPR*, pages 2639–2646. IEEE, 2009.

NK99. Peter J Neugebauer and Konrad Klein. Texturing 3d models of real world objects from multiple unregistered photographic views. In *Computer Graphics Forum*, volume 18, pages 245–256. Wiley Online Library, 1999.

NRDR05. Diego Nehab, Szymon Rusinkiewicz, James Davis, and Ravi Ramamoorthi. Efficiently combining positions and normals for precise 3d geometry. *ACM transactions on graphics (TOG)*, 24(3):536–543, 2005.

PARD$^+$98. Kari Pulli, Habib Abi-Rached, Tom Duchamp, Linda G Shapiro, and Werner Stuetzle. Acquisition and visualization of colored 3d objects. In *Proc. ICPR*, pages 11–15. IEEE, 1998.

PCDS12. Gianpaolo Palma, Marco Callieri, Matteo Dellepiane, and Roberto Scopigno. A statistical method for svbrdf approximation from video sequences in general lighting conditions. In *Computer Graphics Forum*, volume 31, pages 1491–1500. Wiley Online Library, 2012.

PG15. Ruggero Pintus and Enrico Gobbetti. A fast and robust framework for semi-automatic and automatic registration of photographs to 3d geometry. *Journal on Computing and Cultural Heritage (JOCCH)*, 7(4):23:1–23:23, February 2015.

PGC11a. Ruggero Pintus, Enrico Gobbetti, and Marco Callieri. Fast low-memory seamless photo blending on massive point clouds using a streaming framework. *Journal on Computing and Cultural Heritage (JOCCH)*, 4(2):6, 2011.

PGC11b. Ruggero Pintus, Enrico Gobbetti, and Marco Callieri. A streaming framework for seamless detailed photo blending on massive point clouds. In *Proc. Eurographics Area Papers*, pages 25–32, 2011.

PGC11c. Ruggero Pintus, Enrico Gobbetti, and Roberto Combet. Fast and robust semi-automatic registration of photographs to 3d geometry. In *The 12th International Symposium on Virtual Reality, Archaeology and Cultural Heritage*, pages 9–16, October 2011.

RCMS02. Claudio Rocchini, Paolo Cignoni, Claudio Montani, and Roberto Scopigno. Acquiring, stitching and blending diffuse appearance attributes on 3d models. *The Visual Computer*, 18(3):186–204, 2002.

Rem11. Fabio Remondino. Heritage recording and 3d modeling with photogrammetry and 3d scanning. *Remote Sensing*, 3(6):1104–1138, 2011.

SA01. Ioannis Stamos and PK Alien. Automatic registration of 2-d with 3-d imagery in urban environments. In *Proc. ICCV*, volume 2, pages 731–736. IEEE, 2001.

SG02. Vítor Sequeira and João GM Gonçalves. 3d reality modelling: Photo-realistic 3d models of real world scenes. In *Proc. 3DPVT*, pages 776–783. IEEE, 2002.

SLC$^+$08. Ioannis Stamos, Lingyun Liu, Chao Chen, George Wolberg, Gene Yu, and Siavash Zokai. Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *International Journal of Computer Vision*, 78(2-3):237–260, 2008.

SSS$^+$08. Sudipta N Sinha, Drew Steedly, Richard Szeliski, Maneesh Agrawala, and Marc Pollefeys. Interactive 3d architectural modeling from unordered photo collections. In *ACM Transactions on Graphics (TOG)*, volume 27, page 159. ACM, 2008.

VWI97. Paul Viola and William M Wells III. Alignment by maximization of mutual information. *International journal of computer vision*, 24(2):137–154, 1997.

WCL$^+$08. Changchang Wu, Brian Clipp, Xiaowei Li, J-M Frahm, and Marc Pollefeys. 3d model matching with viewpoint-invariant patches (vip). In *Proc. CVPR*, pages 1–8. IEEE, 2008.

WJP08. F Walkowski, RA Johnston, and NB Price. Texture mapping for the fastscan hand-held laser scanner. In *Proc. IVCNZ*, pages 1–6. IEEE, 2008.

WLH$^+$04. Nathaniel Williams, K-L Low, Chad Hantak, Marc Pollefeys, and Anselmo Lastra. Automatic image alignment for 3d environment modeling. In *Proc. Computer Graphics and Image Processing*, pages 388–395. IEEE, 2004.

WMG14. Michael Waechter, Nils Moehrle, and Michael Goesele. Let there be color! large-scale texturing of 3d reconstructions. In *Computer Vision–ECCV 2014*, pages 836–850. Springer, 2014.

XLL+10. Lin Xu, Eric Li, Jianguo Li, Yurong Chen, and Yimin Zhang. A general texture mapping framework for image-based 3d modeling. In *ICIP*, pages 2713 –2716, 2010.

XYWZ05. Shu-Chang Xu, Xiuzi Ye, Yin Wu, and Sanyuan Zhang. Highlight detection and removal based on chromaticity. In *Image Analysis and Recognition*, pages 199–206. Springer, 2005.

YBS07. Gehua Yang, Jacob Becker, and Charles V Stewart. Estimating the location of a camera with respect to a 3d model. In *Proc. 3DIM*, pages 159–166. IEEE, 2007.

ZCS10. Hongwei Zheng, Ioan Cleju, and Dietmar Saupe. Highly-automatic mi based multiple 2d/3d image registration using self-initialized geodesic feature correspondences. In *Computer Vision–ACCV 2009*, pages 426–435. Springer, 2010.

ZK14. Qian-Yi Zhou and Vladlen Koltun. Color map optimization for 3d reconstruction with consumer depth cameras. *ACM Transactions on Graphics (TOG)*, 33(4):155, 2014.

ZNH05. Wenyi Zhao, David Nister, and Steve Hsu. Alignment of continuous video onto 3d point clouds. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(8):1305–1318, 2005.