

Chapter 17

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.

Keywords Image-to-geometry registration • Color mapping • Color blending • 3D reconstruction

17.1 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 noninvasive 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 3D digitization approach is a combination of an active scanning device (e.g., laser triangulation, structured light, time-of-flight, interference) and digital photography. By using computational techniques, digital surfaces are reconstructed from the

R. Pintus • E. Gobbetti (✉)
CRS4, Visual Computing Group, Cagliari, Italy
e-mail: ruggero.pintus@crs4.it; enrico.gobbetti@crs4.it

M. Callieri • M. Dellepiane
ISTI-CNR, Visual Computing Group, Pisa, Italy
e-mail: marco.callieri@isti.cnr.it; matteo.dellepiane@isti.cnr.it

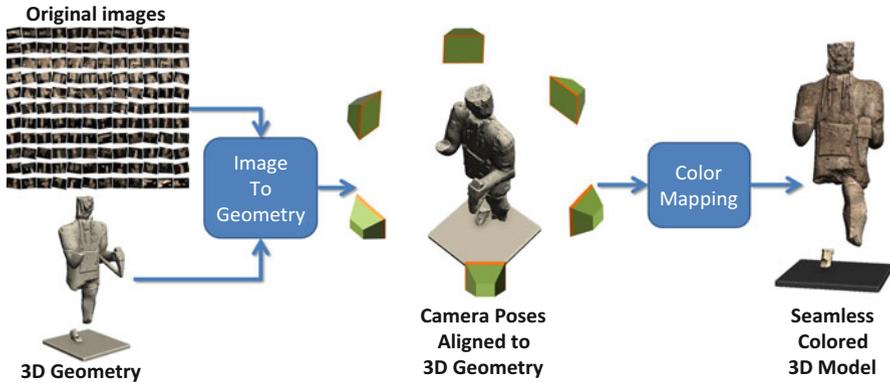


Fig. 17.1 Pipeline overview. The input data are a 3D geometry and a set of photos of the real object. First, an image-to-geometry alignment step registers the images to the 3D model, by producing a set of corresponding camera poses. By exploiting the resulting link between 3D points and pixels, a color mapping procedure produces a seamless color signal across the surface

scanner-generated range maps, and 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 projecting pixel values to the 3D surface using the recovered inverse projections (see Fig. 17.1). Since early demonstrations of the complete modeling pipeline (e.g., Bernardini and Rushmeier 2002; Levoy et al. 2000), 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 (Remondino 2011), the pipeline based on active geometric scanning sensors and separate color acquisition still remains a widely used general-purpose approach, mainly because of the higher flexibility and reliability for a wider variety of settings (e.g., featureless surfaces) (Remondino 2011; Koutsoudis et al. 2014).

As exemplified in Fig. 17.1, 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 the geometric data, in order to establish the mapping between appearance information and the 3D surface with 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 show that, although 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 the image and geometry side, and are mostly (and often completely) automatic, opening the way to interesting innovative applications.

17.2 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 Sect. 17.3) to transfer color information from the images to the 3D model.

The most straightforward way to get 2D image data mapped to a geometry is to fix a calibrated camera rigidly to an acquisition device. *Fixed-relative* (or *colocated*) methods (Pulli et al. 1998; Fruh and Zakhor 2001, 2003) 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. (2007) show that this kind of approach, in a hybrid framework that exploits other methods (e.g., *color-based*, see Sect. 17.2.3), can be applied in a number of interesting applications, because the color obtained in the acquisition phase can be used to map other photos with different appearances. 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 (Corsini et al. 2013)). There is often an interest in mapping multiple image datasets acquired in different moments to the same geometry, such as the two image sets acquired by a professional photographer before and after the restoration of Michelangelo's statue of David (Dellepiane et al. 2008). 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 thus be missed in some surface regions. Finally, although a *colocated* 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 use

multispectral (MS) acquisitions (e.g., far ultraviolet, far infrared thermal images). In contrast to a reflectance signal, which is typically from a visible wavelength, these signals might have really different appearance features; this inconsistency depends on variations of transmittance and absorption across the spectrum. The resulting images must be aligned to the geometry without exploiting the presence of the *colocated* camera.

All these reasons lead to a broad range of techniques to enable registration of independently acquired image and 3D datasets. The available solutions include manual registration methods (Sect. 17.2.1), feature-based approaches that match geometric and image features (Sect. 17.2.2), color-based techniques that exploit additional color signals from the 3D scanning device (Sect. 17.2.3), statistical approaches that rely on mutual information (Sect. 17.2.4), 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 (Sect. 17.2.5).

17.2.1 *Manual Methods*

The most common pipeline for mapping photos onto a 3D geometry consists of asking the user to select correspondences manually 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's intrinsic and/or extrinsic parameters. Initially, the solutions proposed focus on the definition of the most usable user interfaces to facilitate this labor-intensive task (Dellepiane et al. 2008; Fruh and Zakhor 2003). Later, there was a focus on speeding up and reducing manual operations. Borgeat et al. (2009) developed an interface where possible feature matches were highlighted. Another more efficient strategy is to include not only 2D/3D pairs, but also links between pixels in different images (Franken et al. 2005). Moreover, in particular settings, careful planning of image capture might be a solution to speed up or mostly automate the registration (Matsushita and Kaneko 1999). In any case, although these semi-automatic methods are robust due to continuous support and check by humans on the registration process, they are hardly applicable if the image set starts to expand over a certain moderate cardinality (e.g., tens or hundreds of photographs).

17.2.2 *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 because the pure geometry and the object, which is a mix of shape and texture, have very different appearances. For this reason, the works in this area are typically tailored to models with specific geometric characteristics, such as points, lines, circles (Stamos et al. 2008), rectangles (Liu and Stamos 2005), edge intensities (Neugebauer and Klein 1999), or viewpoint-invariant patches (Wu et al. 2008). Their performances are mostly reliable with 3D models of buildings, which show very sharp edges that can be easily extracted by a geometric operator applied to the 3D data, and are a very clear gradient signal in the 2D images. Another class of approaches (Kaminsky et al. 2009; Stamos and Alien 2001) focuses 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. (2006) rely on the presence of clusters of vertical and horizontal lines, and they exploit the orthogonality constraints to drive the registration algorithm. They process the geometry to extract parallelepipeds, and they find correspondences between them and rectangles in the photographs. This method results in being 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 image-to-geometry alignment (Lowe 1991; Brunie et al. 1992). 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. (2000) present an efficient, hardware-accelerated implementation of such algorithms, and propose a robust similarity function. They also do not need any manual intervention.

Unfortunately, these techniques require the whole object to be visible in all input photographs. In addition, they moderately depend 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., a severe limitation for a large-scale target), reducing the spectrum of possible applications.

17.2.3 *Color-Based Methods*

Whereas *feature-based* methods try to extract a visual cue from the geometry comparable to the intensity or color signal in the photographs, *color-based* techniques exploit 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. (2007) and Sequeira and Gonçalves (2002) rely on the extraction of edges in the reflectance information. These data show characteristics similar to color images, because 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 signals. Yang et al. (2007) use a colocated setup that produces a set of colored range maps. Then they extract image descriptors from the color captured by the scanner and by a handheld camera. The correspondences between these two sources are employed to compute the extrinsic and intrinsic parameters of the handheld camera, solving the 2D/3D registration. Wu et al. (2008) 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.

17.2.4 Statistical Methods

Statistical image registration methods based on mutual information (MI) are extensively used in many applications. In the field of image-to-geometry registration, MI is capable of analyzing nonlinear correlations and photoconsistency 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 (1997) correlates the image gradient with a particular rendering of the surface normals. Cleju and Saupe (2007) 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. (2004) 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 (2006) 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. (2009) exploits the statistical dependency that in urban scenes correlates the LIDAR elevation with the optical appearance. They measure MI between LIDAR and the optical signal, and apply downhill simplex optimization to find camera poses. They initialize each camera pose from GPS data, and harness the power of graphics hardware to reduce registration times dramatically. In the field of CH, Corsini et al. (2009) take 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. (2010) strives to integrate, instead, the advantages of *statistical* and *feature-based* registration (Zheng et al. 2010) 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 global stochastic optimization. Recently, Dellepiane and Scopigno (2013) proposed a MI framework that focuses on registration refinement; they globally tune the camera poses by fitting the original color signal with the color mapping result, thus removing artifacts produced by small image misalignments (see also Sect. 17.3.2).

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

17.2.5 *Multiview Methods*

In previous sections the vast majority of the approaches already mentioned aim at mapping one single image at a time; conversely, *multiview* techniques exploit the intrinsic grouping nature of the input images and the relations among them. The fundamental insight that underlies the class of *multiview*-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 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, inasmuch as they do not require such strict constraints as others do. Compared to *feature-based* methods (Sect. 17.2.2) they do not rely on specific features such as circles, lines, or orthogonal planes. Furthermore, they do not try to find correspondences between images and geometry, but they first find matches between images (SfM step), which is a much faster and simpler task, and then solve a pairwise geometry alignment, which is a well-studied problem too. Apart from the completely automatic *multiview* solutions, compared to *statistical* approaches (Sect. 17.2.4), the *multiview* methods reduce manual intervention and completely avoid the use of any additional attributes of the geometrical primitives, which are not always accessible, such as reflectance, color from the acquisition device, LIDAR elevation, or probability of detection (see Sects. 17.2.3 and 17.2.4). The techniques in the *multiview* class just need a moderate presence of any type of texture and/or geometrical features, which is a normal nonlimiting requirement of the classical SfM algorithm.

One of the earliest works in this field was due to Zhao et al. (2005), which presented a way to register a video onto a point cloud. By using motion stereo, they computed 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 employed an ICP-based procedure to refine the registration between the sparse geometry 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 *multiview* processing step. For instance, Liu et al. (2006) extended their previous work on a *feature-based* 2D/3D mapping (Liu and Stamos 2005), by integrating a SfM geometry and an automatic 3D registration. They used a number of 3D scans to build a dense point cloud; in this step they inherited the *feature-based* approach of the previous work to find correspondences between 3D lines in the range images. Then, they computed a sparse geometry from the sequence of the input 2D photographs. They found the rotation, scale, and translation that minimized the distance between the two 3D shapes. Stamos et al. (2008) 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 *multiview* has not been completely exploited to solve a more general problem.

Inasmuch as relying on ordered sets of images is a too limiting constraint, such as in Zhao et al. (2005), a number of methods naturally arose that tried to deal with unordered sets of sufficiently overlapping photos. Moreover, acquisition of complex geometrical shapes inevitably leads to nonuniform image sampling; the capture must be performed adaptively, and the number and density of 2D frames per region depends on the presence and nature of surface occlusions. This kind of acquisition strategy eventually produces an unordered image set as input for the registration pipeline.

Li and Low (2009) register unordered image sequences from untracked and uncalibrated cameras onto 3D geometry. Their method is pure *multiview*. 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 a 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 reprojection error. A further manual tuning of a heuristic parameter is required to weight these two error terms properly. Similar semi-automatic approaches have been proposed by Banno and Ikeuchi (2010) and Pintus et al. (2011c). The first presents an accurate texturing method that estimates camera extrinsic parameters by computing a sparse geometry from a spherical image stereo dataset. They measure the distance between the sparse and dense geometry (from a range sensor) by taking into account the uncertainty distribution of 3D in the stereo data. After the rough manual alignment of a few

images to the dense geometry, Pintus et al. (2011c) propose an image-to-geometry optimization that in a nonrigid manner independently deforms the sparse 3D points together with the poses of each camera. The error metric relies only on image-based reprojection 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 require minimal user intervention to define an 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 that are capable of dealing with large input datasets and need neither an object-dependent parameter setting, nor, of course, a manual intervention.

Corsini et al. (2013) propose a novel multistage framework for a fully automatic 2D/3D global registration, shown in Fig. 17.2. After the SfM procedure, carried

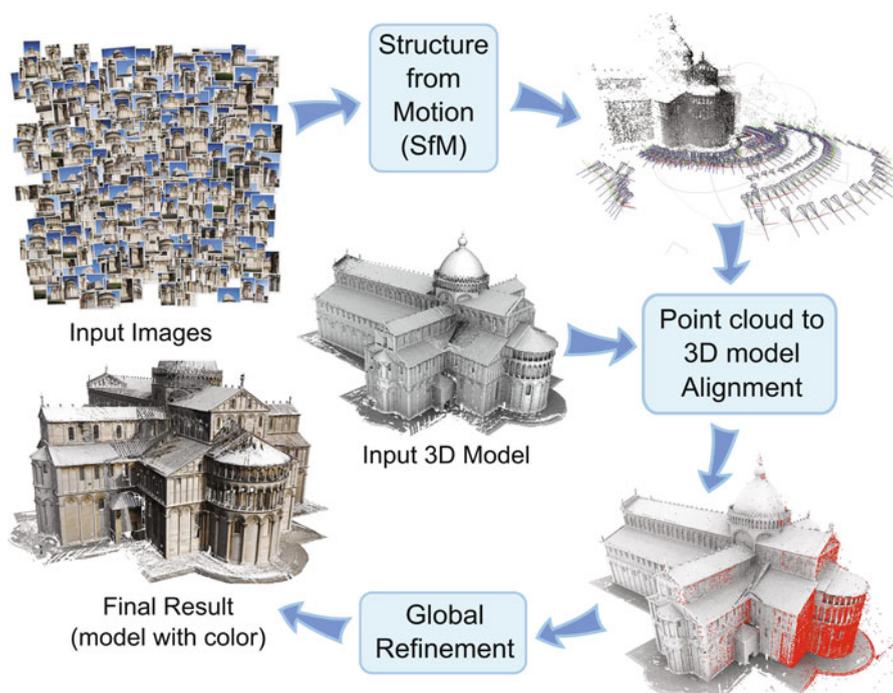


Fig. 17.2 The multi-image alignment proposed by Corsini et al. (2013), which uses an approach similar to Pintus et al. (2011c, 2015): The images are used to compute a sparse point cloud using SfM, and the cloud is then aligned using mutual information, thus obtaining an initial camera orientation that will then be refined using mutual information (Reprinted from Corsini et al. 2013 with kind permission from Springer Science+Business Media)

out similarly to Pintus et al. (2011c), the generated sparse point cloud is aligned with the dense model by using a modified version of the four-point congruent set (4PCS) (Aiger et al. 2008) algorithm. The 4PCS deals with rigid motion, however, they extend it to cope with both different scales and an unknown amount of overlapping regions. Finally, a global refinement strategy is used based on mutual information, which minimizes the color reprojection error of the images onto the geometry. The method of Pintus et al. (2015) 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 (Agus et al. 2014). They propose a robust statistics to manage outliers, and a strategy to estimate automatically a local adaptive tolerance to deal with nonuniformly sampled 3D geometry. As in Pintus et al. (2011c), they refine intrinsic and extrinsic parameters of each camera in a nonrigid manner, by employing a specialized sparse bundle adjustment (SBA) step and an image-based error metric. Compared to Corsini et al. (2013), they do not require heavy preprocessing to make the SfM sparse point cloud dense, because 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 (Agus et al. 2014).

These approaches based on fully automatic 2D/3D registration are more generally applicable and robust. Their main advantage is that they do not make any assumption tailored to the geometry type, size, or sampling, and are suitable to be applied to small, medium, and large objects of different natures. Because they are scalable, they remain 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 a few hand-picked correspondences between image and geometry, followed by local refinement.

17.3 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. Although there are many direct uses of a set of calibrated and oriented images, in most cases, the image alignment is used to map the color information onto the 3D geometry, thus obtaining a colored 3D model.

The recovered camera alignment data are, 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, because a simple inverse perspective projection will be enough to map the color onto the 3D geometry. The visibility problem determines what part of the surface is directly visible by the camera (and thus can be mapped from that specific photo) and 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; inasmuch as direct shadow mapping would not work, Pintus et al. (2011a) use 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. The basic inverse projection is still used to map color from the photos to the surface, however, the issue is now what happens when more than one input photo maps onto 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 to solve the “blending problem,” that is, combine the available contribution to have a more continuous color mapping. This topic is covered in Sect. 17.3.1.

A common problem of the color mapping process is the fact that, no matter how precise 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 are covered in Sect. 17.3.2.

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 artifacts (e.g., highlights) that change their relative position from one photo to another. Although some of these issues can be reduced or removed by the blending process, there are specific methods in the literature that can be applied to solve these situations (see Sect. 17.3.3).

17.3.1 Projecting and Blending

An issue that is orthogonal to the mapping method is where to store the mapped color: a family of solutions uses (or creates) a surface parameterization and stores the mapped color in the texture map; a different approach is to 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, such as in Debevec et al. (1998); or by selecting the relevant images for a specific view, and then mixing multiple images using a weighting scheme (Pulli et al. 1998), blending the different contributions in a continuous mapped image.

Although it is able to provide good results, view-dependent texture mapping is not used much 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 first 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 detail 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 the literature try to cover the spectrum between these two extremes, providing a 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 (Walkowski et al. 2008), 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 result in a natural and usable parameterization. Some mapping methods (Lensch et al. 2001; Rocchini et al. 2002; Callieri et al. 2002) build a parameterization 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. (2001) and Rocchini et al. (2002) filter the border between adjacent patches to reduce the discontinuity, whereas (Callieri et al. 2002) apply a texture-wide color/luminance correction to make the color uniform across the different photos.

Instead of using a simple “best contribution” criterion, but still wanting to use fragments of the original images directly, 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. (2008) 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 (2007) also

use this kind of MRF minimization, but at a per-triangle level, and Gal et al. (2010) also introduce a parameterization shifting order to correct 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 a smooth mapping simultaneously, but with sharp details. Baumberg (2002) separates the input images in two bands: in every given area of the 3D model all the low-frequency components are blended using weights, but only the “best” high-frequency details are mapped. This solution is also used by Allène et al. (2008) and by Pintus et al. (2011a) on point clouds.

Among the methods blending all available contributions, performing a weighted sum is the most common way to combine contributions (Pulli et al. 1998; Bernardini et al. 2001; Callieri et al. 2008; Pintus et al. 2011a,b; Bettio et al. 2015). All these researchers 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), or 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 that 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 (Chuang et al. 2009), taking into account the input image gradients, building a continuous yet nonblurred color field. Another interesting method is to use a super-resolution approach over the texture space (Goldluecke and Cremers 2009), resulting in a texture of the input photos that is more detailed.

Even though point clouds are widely used directly, without converting them to a triangulated surface, there are not many color mapping methods suitable for these kinds of data. Pintus et al. (2011a,b) work directly on the point clouds, mapping color to each vertex, using a weighted blending two-band strategy, able to work on very large datasets. Conversely, Arıkan et al. (2014) 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, and the image and mask dataset are stored in an out-of-core structure (e.g., Xu et al. 2010; Callieri et al. 2008). 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. (2008) and Bettio et al. (2013, 2015) 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. (2011b; 2011a) 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. 17.3). In



Fig. 17.3 *First row:* Image blending on massive point clouds, the case of Michelangelo's David statue (0.25 mm 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 the 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: Pintus et al. 2011b, Figs. 7, 8 and 9; Scopigno, R., Fig. 1)

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. 17.3, second row).

17.3.2 Fixing Misalignments

Color projection on 3D models is a 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, because 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 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 the 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 of achieving a perfect alignment, because the model is not able to describe the camera perfectly.
- The 3D model quality: 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 (2007) inserted a term in their minimization function that should reduce misalignment 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 (2013) 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 multiview stereo matching techniques. In this case, similarly to the work by Gal et al. (2010), 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 us to preserve detail. The work by Waechter et al. (2014) shows that this can be used to obtain large- scale 3D texturing.

In a more general case, geometry and images are acquired at a different moment, thus it could be harder to handle them. Pintus et al. (2011a) 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. (2008)

used it to fit textures to body models. They proposed a GPU implementation of the method that reaches real-time performances. Hence the algorithm can be employed not only in offline image-to-geometry registration frameworks, but also in image-based visualizations.

Dellepiane et al. (2012) 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 (2014) proposed a method to map color on data coming from depth cameras. The method put 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 was solved in an iterative way (by refining cameras and flow one at a time), and the results were more than satisfying.

17.3.3 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 onsite 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 (Xu et al. 2005; Guo et al. 2011). 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 (Frahm et al. 2010). This is not always possible, especially when high-quality images are needed. Hence, an alternative way is to have better knowledge of the lighting environment.

That being said, except for laboratory conditions and objects of small size, it’s rarely possible to set up 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 (Corsini et al. 2008). 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. (2012) used video sequences to estimate the illumination environment, and also extracted a simplified description of the material of an object.

Another source of light that could be possibly estimated is the sky. Haber et al. (2009) analyzed the sky and the appearance of an object to estimate the skydome environment and relight images. Dellepiane et al. (2010a) estimated the sun's position (in an automatic or semi-automatic way) to remove hard shadows.

Finally, an interesting example of a controlled but easy to use light setup is FISS (Dellepiane et al. 2010b), where the integrated flash device of the digital camera is used. This type of light has the advantage of being always at the same position w.r.t. the camera, and to provide dominant lighting in the scene. Hence, it is easy to detect and remove artifacts, and it is possible to obtain a partially calibrated color after a calibration step is 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 (Kim et al. 2012; Bettio et al. 2013, 2015) take the simpler approach of using 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.

17.4 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 many 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 because 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.

Although 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, whereas 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 into the loop in order to control the fine behavior of algorithms. However, as data size is increasing, because acquisition is 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, although early methods, especially for the mapping step, required global operation with all data loaded into memory, modern methods 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.

Although in this work we have focused mostly on the registration of color images onto the geometry, the techniques presented here are the basis for managing a number of multimodal fusion operations. Recent trends highlight in particular the need of 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 these more complex data, with limited modifications to implement the mapping and fusion (e.g., different descriptors for automatic alignment). On the other hand, sometimes microscale data acquired as 2D maps also provide information at the geometric level. For instance, reflectance transformation imaging acquisitions (Malzbender et al. 2001) make it possible to compute high-resolution normal maps, which, once registered, can be employed to refine the initial geometry (Nehab et al. 2005; Berkiten et al. 2014). This is usually done in a post-pass, however, 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

- Agus M, Gobbetti E, Villanueva AJ, Mura C, Pajarola R (2014) SOAR: Stochastic optimization for affine global point set registration. In: Proceedings of the 19th international workshop on vision, modeling and visualization (VMV), Oct 2014, Darmstadt, pp 103–110
- Aiger D, Mitra NJ, Cohen-Or D (2008) 4-points congruent sets for robust pairwise surface registration. *ACM Trans Graph (TOG)* 27:85
- Allène C, Philippe Pons J, Keriven R (2008) Seamless image-based texture atlases using multi-band blending. In: Proceeding of the ICPR. IEEE, pp 1–4

- Arikan M, Preiner R, Scheiblaue C, Jeschke S, Wimmer M (2014) Large-scale point-cloud visualization through localized textured surface reconstruction. *IEEE Trans Vis Comput Graph* 20(9):1280–1292
- Banno A, Ikeuchi K (2010) Omnidirectional texturing based on robust 3D registration through Euclidean reconstruction from two spherical images. *Comput Vis Image Underst* 114(4):491–499
- Baumberg A (2002) Blending images for texturing 3D models. In: *Proceeding of the BMVC, Citeseer*, vol 3, p 5
- Berkiten S, Fan X, Rusinkiewicz S (2014) Merge2-3D: combining multiple normal maps with 3D surfaces. In: *2nd international conference on 3D vision*, Dec 2014, vol 1. IEEE, Los Alamitos/Washington/Tokyo, pp 440–447
- Bettio F, Gobbetti E, Merella E, Pintus R (2013) Improving the digitization of shape and color of 3D artworks in a cluttered environment. In: *Proceedings of the digital heritage*, Oct 2013. IEEE, Piscataway, pp 23–30. Best Paper Award
- Bettio F, Villanueva AJ, Merella E, Marton F, Gobbetti E, Pintus R (2015) Mont'e Scan: effective shape and color digitization of cluttered 3D artworks. *J Comput Cult Herit (JOCCH)* 8(1):4:1–4:23
- Bernardini F, Martin IM, Rushmeier H (2001) High-quality texture reconstruction from multiple scans. *IEEE Trans Vis Comput Graph* 7(4):318–332
- Bernardini F, Rushmeier H (2002) The 3D model acquisition pipeline. *Comput Gr Forum* 21:149–172
- Borgeat L, Poirier G, Beraldin A, Godin G, Massicotte P, Picard M (2009) A framework for the registration of color images with 3D models. In: *Proceedings of the 16th international conference on image processing (ICIP'09)*, IEEE Press, Piscataway, NJ, USA, pp 69–72
- Brunie L, Lavallée S, Szeliski R (1992) Using force fields derived from 3D distance maps for inferring the attitude of a 3D rigid object. In: *Computer Vision ECCV'92*. Springer, pp 670–675
- Callieri M, Cignoni P, Corsini M, Scopigno R (2008) Masked photo blending: mapping dense photographic data set on high-resolution sampled 3D models. *Comput Graph* 32(4):464–473
- Callieri M, Cignoni P, Scopigno R (2002) Reconstructing textured meshes from multiple range RGB maps. In: *VMV, Erlangen*, pp 419–426
- Chuang M, Luo L, Brown BJ, Rusinkiewicz S, Kazhdan M (2009) Estimating the laplace-beltrami operator by restricting 3D functions. *Comput Gr Forum* 28:1475–1484. Wiley Online Library
- Cleju I, Saupe D (2007) Stochastic optimization of multiple texture registration using mutual information. Springer, Konstanz
- Corsini M, Callieri M, Cignoni P (2008) Stereo light probe. *Comput Gr Forum* 27:291–300. Wiley Online Library
- Corsini M, Dellepiane M, Ganovelli F, Gherardi R, Fusiello A, Scopigno R (2013) Fully automatic registration of image sets on approximate geometry. *Int J Comput Vis* 102(1–3):91–111
- Corsini M, Dellepiane M, Ponchio F, Scopigno R (2009) Image-to-geometry registration: a mutual information method exploiting illumination-related geometric properties. *Comput Gr Forum* 28:1755–1764. Wiley Online Library
- Debevec P, Yizhou Yu, Borshukov G (1998) Efficient view-dependent image-based rendering with projective texture-mapping. Springer, Berkeley
- Dellepiane M, Benedetti L, Scopigno R (2010) Removing shadows for color projection using sun position estimation. In: *Proceedings of the 11th international conference on virtual reality, archaeology and cultural heritage*. Eurographics Association, pp 55–62
- Dellepiane M, Callieri M, Corsini M, Cignoni P, Scopigno R (2010) Improved color acquisition and mapping on 3D models via flash-based photography. *J Comput Cult Herit (JOCCH)* 2(4):9
- Dellepiane M, Callieri M, Ponchio F, Scopigno R (2008) Mapping highly detailed colour information on extremely dense 3D models: the case of David's restoration. *Comput Gr Forum* 27:2178–2187. Wiley Online Library

- Dellepiane M, Marroquim R, Callieri M, Cignoni P, Scopigno R (2012) Flow-based local optimization for image-to-geometry projection. *IEEE Trans Vis Comput Graph* 18(3):463–474
- Dellepiane M, Scopigno R (2013) Global refinement of image-to-geometry registration for color projection on 3D models. In: *Digital heritage international congress*. The Eurographics Association, pp 39–46
- Eisemann M, De Decker B, Magnor M, Bekaert P, De Aguiar E, Ahmed N, Theobalt C, Sellent A (2008) Floating textures. *Comput Gr Forum* 27:409–418. Wiley Online Library
- Franken T, Dellepiane M, Ganovelli F, Cignoni P, Montani C, Scopigno R (2005) Minimizing user intervention in registering 2D images to 3D models. *Vis Commun* 21(8–10):619–628
- Frahm JM, Fite-Georgel P, Gallup D, Johnson T, Raguram R, Wu C, Jen Y-H, Dunn E, Clipp B, Lazebnik S et al (2010) Building Rome on a cloudless day. In: *Computer vision–ECCV 2010*. Springer, pp 368–381
- Fruh C, Zakhor A (2001) Fast 3D model generation in urban environments. In: *Proceedings of the MFI*. IEEE, pp 165–170
- Fruh C, Zakhor A (2003) Constructing 3D city models by merging aerial and ground views. *Comput Gr Appl IEEE* 23(6):52–61
- Gal R, Wexler Y, Ofek E, Hoppe H, Cohen-Or D (2010) Seamless montage for texturing models. *Comput Gr Forum* 29:479–486. Wiley Online Library
- Goldluecke B, Cremers D (2009) Superresolution texture maps for multiview reconstruction. In: *Proceedings of the ICCV*. IEEE, pp 1677–1684
- Guo R, Dai Q, Hoiem D (2011) Single-image shadow detection and removal using paired regions. In: *Proceedings of the CVPR*. IEEE, pp 2033–2040
- Haber T, Fuchs C, Bekaer P, Seidel H-P, Goesele M, Lensch HP et al (2009) Relighting objects from image collections. In: *Proceedings of the CVPR*. IEEE, pp 627–634
- Hantak C, Lastra A (2006) Metrics and optimization techniques for registration of color to laser range scans. In: *International symposium on 3D data processing, visualization, and transmission, third*. IEEE, pp 551–558
- Ikeuchi K, Oishi T, Takamatsu J, Sagawa R, Nakazawa A, Kurazume R, Nishino K, Kamakura M, Okamoto Y (2007) The Great Buddha project: digitally archiving, restoring, and analyzing cultural heritage objects. *Int J Comput Vis* 75(1):189–208
- Kaminsky RS, Snavely N, Seitz SM, Szeliski R (2009) Alignment of 3D point clouds to overhead images. In: *Proceedings of the CVPR workshops*. IEEE, pp 63–70
- Koutsoudis A, Vidmar B, Ioannakis G, Arnaoutoglou F, Pavlidis G, Chamzas C (2014) Multi-image 3D reconstruction data evaluation. *J Cult Herit* 15(1):73–79
- Kim SJ, Lin HT, Lu Z, Suesstrunk S, Lin S, Brown MS (2012) A new in-camera imaging model for color computer vision and its application. *IEEE Trans PAMI* 34(12):2289–2302
- Lensch H, Heidrich W, Seidel H-P (2000) Automated texture registration and stitching for real world models. In: *Proceedings of the pacific graphics*. IEEE, pp 317–452
- Lensch HPA, Heidrich W, Seidel H-P (2001) A silhouette-based algorithm for texture registration and stitching. *Graph Model* 63(4):245–262
- Lempitsky V, Ivanov D (2007) Seamless Mosaicing of image-based texture maps. In: *Proceedings of the CVPR*. IEEE, pp 1–6
- Leroy 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: *Proceedings of the SIGGRAPH*. ACM, New York, pp 131–144
- Li Y, Low K-L (2009) Automatic registration of color images to 3D geometry. In: *Proceedings of the CGI*. CGS, pp 21–28
- Liu L, Stamos I (2005) Automatic 3D to 2D registration for the photorealistic rendering of urban scenes. In: *Proceedings of the CVPR*. IEEE, vol 2, pp 137–143
- Liu L, Stamos I, Gene Yu, Wolberg G, Zokai S (2006) Multiview geometry for texture mapping 2D images onto 3D range data. In: *Proceedings of the CVPR*. IEEE, vol 2, pp 2293–2300
- Lowe DG (1991) Fitting parameterized three-dimensional models to images. *IEEE Trans Pattern Anal Mach Intell* 13(5):441–450

- Malzbender T, Gelb D, Wolters H (2001) Polynomial texture maps. In: Proceedings of the 28th annual conference on computer graphics and interactive techniques. ACM, pp 519–528
- Mastin A, Kepner J, Fisher J (2009) Automatic registration of LIDAR and optical images of urban scenes. In: Proceedings of the CVPR. IEEE, pp 2639–2646
- Matsushita K, Kaneko T (1999) Efficient and handy texture mapping on 3D surfaces. *Comput Gr Forum*, Blackwell Publishers, Oxford, UK, vol 18, pp 349–358
- Nehab D, Rusinkiewicz S, Davis J, Ramamoorthi R (2005) Efficiently combining positions and normals for precise 3D geometry. *ACM Trans Graph (TOG)* 24(3):536–543
- Neugebauer PJ, Klein K (1999) Texturing 3D models of real world objects from multiple unregistered photographic views. *Comput Gr Forum* 18:245–256. Wiley Online Library
- Palma G, Callieri M, Dellepiane M, Scopigno R (2012) A statistical method for SVBRDF approximation from video sequences in general lighting conditions. *Comput Gr Forum* 31:1491–1500. Wiley Online Library
- Pintus R, Gobbetti E (2015) A fast and robust framework for semi-automatic and automatic registration of photographs to 3D geometry. *J Comput Cult Herit (JOCCH)* 7(4):23:1–23:23
- Pintus R, Gobbetti E, Callieri M (2011a) Fast low-memory seamless photo blending on massive point clouds using a streaming framework. *J Comput Cult Herit (JOCCH)* 4(2):6
- Pintus R, Gobbetti E, Callieri M (2011b) A streaming framework for seamless detailed photo blending on massive point clouds. In: Proceedings of the eurographics area papers. pp 25–32
- Pintus R, Gobbetti E, Combet R (2011c) Fast and robust semi-automatic registration of photographs to 3D geometry. In: The 12th international symposium on virtual reality, archaeology and cultural heritage Oct 2011. Eurographics Association, pp 9–16
- Pulll K, Abi-Rached H, Duchamp T, Shapiro LG, Stuetzle W (1998) Acquisition and visualization of colored 3D objects. In: Proceedings of the ICPR. IEEE, pp 11–15
- Remondino F (2011) Heritage recording and 3D modeling with photogrammetry and 3D scanning. *Remote Sens* 3(6):1104–1138
- Rocchini C, Cignoni P, Montani C, Scopigno R (2002) Acquiring, stitching and blending diffuse appearance attributes on 3D models. *Vis Comput* 18(3):186–204
- Sequeira V, Gonçalves GMJ (2002) 3D reality modelling: photo-realistic 3D models of real world scenes. In: Proceedings of the 3DPVT. IEEE, pp 776–783
- Sinha SN, Steedly D, Szeliski R, Agrawala M, Pollefeys M (2008) Interactive 3D architectural modeling from unordered photo collections. *ACM Trans Gr (TOG)* 27:159
- Stamos I, Alien PK (2001) Automatic registration of 2-D with 3-D imagery in urban environments. In: Proceedings of the ICCV. IEEE, vol 2, pp 731–736
- Stamos I, Liu L, Chen C, Wolberg G, Yu G, Zokai S (2008) Integrating automated range registration with multiview geometry for the photorealistic modeling of large-scale scenes. *Int J Comput Vis* 78(2–3):237–260
- Viola P, Wells WM III (1997) Alignment by maximization of mutual information. *Int J Comput Vis* 24(2):137–154
- Waechter M, Moehrl N, Goesele M (2014) Let there be color! large-scale texturing of 3D reconstructions. In: Computer vision—ECCV 2014. Springer, pp 836–850
- Walkowski F, Johnston RA, Price NB (2008) Texture mapping for the Fastscan hand-held laser scanner. In: Proceedings of the IVCNZ. IEEE, pp 1–6
- Williams N, Low K-L, Hantak C, Pollefeys M, Lastra A (2004) Automatic image alignment for 3D environment modeling. In: Proceedings of the computer graphics and image processing. IEEE, pp 388–395
- Wu C, Clipp B, Xiaowei Li, Frahm J-M, Pollefeys M (2008) 3D model matching with viewpoint-invariant patches (VIP). In: Proceedings of the CVPR. IEEE, pp 1–8
- Xu S-C, Ye X, Wu Y, Zhang S (2005) Highlight detection and removal based on chromaticity. In: Image analysis and recognition. Springer, pp 199–206
- Xu L, Li E, Li J, Chen Y, Zhang Y (2010) A general texture mapping framework for image-based 3D modeling. In: Proceedings of the international conference on image processing, Hong kong, pp 2713–2716

- Yang G, Becker J, Stewart CV (2007) Estimating the location of a camera with respect to a 3D model. In: Proceedings of the 3DIM. IEEE, pp 159–166
- Zhao W, Nister D, Hsu S (2005) Alignment of continuous video onto 3D point clouds. *IEEE Trans Pattern Anal Mach Intell* 27(8):1305–1318
- Zheng H, Cleju I, Saupe D (2010) Highly-automatic mi based multiple 2D/3D image registration using self-initialized geodesic feature correspondences. In: *Computer vision–ACCV 2009*. Springer, pp 426–435
- Zhou Q-Y, Koltun V (2014) Color map optimization for 3D reconstruction with consumer depth cameras. *ACM Trans Gr (TOG)* 33(4):155