Front. Comput. Sci., 2017, 11(2): 175–191
DOI 10.1007/s11704-016-5520-8

REVIEW ARTICLE

The role of prior in image based 3D modeling: a survey

Hao ZHU, Yongming NIE, Tao YUE, Xun CAO (✉)

Lab for Computational Imaging Technology and Engineering, School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China

Abstract The prior knowledge is the significant supplement to image-based 3D modeling algorithms for refining the fragile consistency-based stereo. In this paper, we review the image-based 3D modeling problem according to prior categories, i.e., classical priors and specific priors. The classical priors including smoothness, silhouette and illumination are well studied for improving the accuracy and robustness of the 3D reconstruction. In recent years, various specific priors which take advantage of Manhattan rule, geometry template and trained category features have been proposed to enhance the modeling performance. The advantages and limitations of both kinds of priors are discussed and evaluated in the paper. Finally, we discuss the trend and challenges of the prior studies in the future.

Keywords prior information, consistency-based stereo, smoothness, illumination, silhouette, specific prior

1 Introduction

3D reconstruction is intended to capture the shape and appearance of real objects, and can be widely used in a variety of fields, such as computer aided geometric design, computer graphics, computer animation, computer vision, medical imaging, computational science, virtual reality, digital media, 3D printing, etc. In the past decades, the 3D reconstruction technique draws widespread attentions and achieves a great progress. Among various 3D modeling approaches, image-based methods own their distinct characteristics: feasible input and convenient operation. With the development of reconstruction algorithms and imaging techniques, the state-of-the-art image-based methods update rapidly. The accuracy of some well-designed recovering systems approaches to that of the laser-scanning systems, while their computational speed can reach real-time level.

The photo-consistency measurements and reconstruction algorithms have been well studied. Several excellent surveys [1–3] have been proposed to review the early multi-view reconstruction methods. Nevertheless, there still remain a few tough problems in image-based modeling methods. The primary problem is the ambiguity in textureless regions, which makes the problem ill-posed and degrades the reconstruction results. Furthermore, different applications demand models with different peculiarity, and thus require different reconstruction approaches. For example, the high-quality rendering needs the integrated and colored mesh model, while the motion capture prefers the unambiguous model, which can be a rough mesh or even the sparse point cloud. Therefore, the additional depth cues, namely priors, are required to supplement the image-based modeling methods, and make the algorithms more practical, robust and distinctive. Different from the above-mentioned surveys, this paper concentrates on the studies and categories of the priors in the existing image-based modeling methods, and summarizes a comprehensive comparison of these priors, and analyzes the role of them in the image-based modeling algorithms. As the state-of-the-art has been improved a lot, we pay closer attention to recent well-performed priors in this paper.

In the rest of this article, we first browse the related methods and discuss the subsistent challenges in Section 2. Then, the priors are classified as common priors (discussed in Section 3) and specific priors (discussed in Section 4). We ana-

Received December 3, 2015; accepted June 16, 2016

E-mail: caoxun@nju.edu.cn

lyze the trend of the future research and make the conclusion in Sections 5 and 6, respectively.

2 Related work

2.1 3D modeling techniques

Generally speaking, 3D modeling methods could be classified into three categories: active methods, image-based methods (passive methods), and the fusion of active and passive methods, as shown in Fig. 1.

Passive modeling methods interfere with the object using laser, infrared ray or other mediums. The traditional laser-based 3D model scanners, e.g., Ref. [4], provide highly-accurate 3D models, however, sophisticated equipments which are expensive and hard to operate are required. Besides, the reconstruction process of the laser scanners is time-consuming and sensitive to calibration errors, which limits its utilization for amateur applications. In contrast, the Kinect [5,6] provides a more convenient and real-time way to acquire depth images, but the spatial resolution is lower than the corresponding intensity sensor, which makes it difficult to capture the details of scenes.

Unlike active method, image-based methods recover depth directly from the image sequences. The recording process of image-based method is more achievable, since only images are required, which can be captured by consumer-level cameras. In exchange, the post-reconstruction of image-based methods is complex and computational-consuming. With the development of the imaging technology, the resolution of consumer-level cameras has increased dramatically, and the noise could be effectively restricted, which makes it possible to approach the quality of laser scanners. Furthermore, the image-based methods capture the corresponding color or texture of the 3D model simultaneously, while the traditional 3D scanner needs an additional RGB sensor and a registration process to acquire the color and texture separately. Therefore, image-based modeling methods have shown their advantages to the traditional active 3D scanning techniques, and own extensive application prospects.

2.2 Image-based reconstruction methods

We consider the image-based modeling problem as a shape-from-image process which may consist of various approaches. The related domain includes following research points:

Photo-consistency based stereo is the most prevailing way in multi-view 3D reconstruction, as it solves the dense corresponding map between two calibrated images and generates the integrated shape. The main idea of the photo-consistency based stereo is to match every pixel of one image (known as the local image) to pixels of the other image (known as the reference image) according to the pixel consistency, and then compute the 3D position of each pixel in the local image according to the epipolar geometry. The stereo methods have been well studied over the last decades. The "Middlebury stereo" evaluation [7] has received over a hundred stereo matching algorithms. The main limitation of stereo is its fragile performance on textureless objects, we will explain this problem in detail in the defect analyzing part.

Multi-view 3D reconstruction generates the 3D model from several images. The multi-view system usually adopts an end-to-end pipeline with following steps: camera calibration (one-off phase if using fixed camera setup), depth estimation in each image, fusion, and finally postprocessing like meshing or coloring. Different from the photo-consistency based stereo methods which are widely used in the depth estimation, the multi-view 3D reconstruction methods pay more attention to the multi-view fusion and optimizing of the final model. Several surveys [1–3] have proposed thorough evaluations and taxonomies on multi-view reconstruction methods, we will further extend their review work and focus on the role of priors in recent state-of-the-art methods.

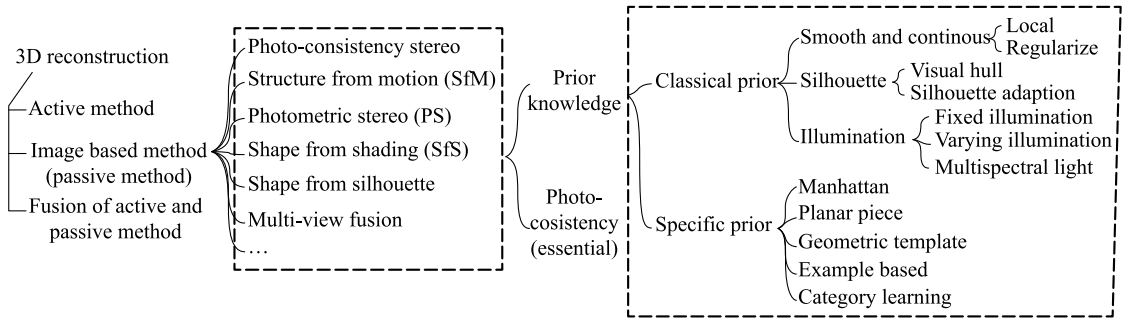


Fig. 1 Relational graph (prior information has played an important role in image-based 3D modeling method to supplement the fragile consistency-based reconstruction)

Structure-from-motion (SFM) is the process of estimating three-dimensional structures from two-dimensional image sequences, which is reviewed in detail in the survey [8,9]. Compared with the above research points, SFM focuses on the estimation of camera parameters, including the camera trajectory, focal length, distortion, etc. In the classical SFM pipeline, the image features, e.g., SIFT [10], are firstly extracted from the images, then the feature correspondences between image pairs are computed. After that, the trajectories of the features are filtered and then used to reconstruct their 3D positions and the camera's motion. The SFM can incrementally and real-timely generate a sparse point cloud from the feature correspondences, and produce the 3D model after subsequent process. The SFM algorithms have been successfully adapted in the **simultaneous localization and mapping (SLAM)** [11] systems, which can construct or update a map of an unknown environment while simultaneously keeping track of an agent's location within it. The state-of-the-art SFM systems can achieve real-time and dense tracking SLAM using inexpensive cameras [12,13], even for the dynamic scenes [14].

Shape-from-shading (SFS) [15] and **photometric stereo** [16] both are 3D reconstruction algorithms that utilize illumination priors. SFS technique was first presented by Horn in the early 1970s, it aims at recovering the light source and the surface shape from a single image. A common assumption in the SFS problem is the Lambertian model, which considers that the intensity of each pixel depends on the direction of the incident light and the surface normal. Under this assumption, the problem remains to be ill-posed as the surface shape and light direction contains overmuch unknowns. Photometric stereo extends SFS problem into multiple illumination, and estimates the surface normals of objects by observing that object under different lighting conditions. Generally speaking, photometric stereo achieves more robust and accurate result at the expense of more complex setups.

Besides the above topics, there are some other studies on image-based reconstruction. **Shape-from-template** [17] reconstructs the shape of a deformable surface from a single image and a known 3D template. **Shape-from-silhouette** [18], a.k.a. **Visual Hull**, reconstructs the 3D model of an object from multiple silhouette images. The silhouette could be segmented easily from the RGB image with pure color background or from multi-view images using multi-view segmentation algorithms [19].

Some of the aforementioned studies focus on a single type prior, and others integrate more than one prior. In this paper, we review the priors of the existing 3D modeling methods, as shown in Fig. 1. By categorizing the priors, we discuss the widely used priors in detail, and summarize the research development and future trend of the image-based 3D reconstruction in the following sections.

2.3 Necessity of prior knowledge

Consistency-based stereo is significant in the universal image-based 3D modeling method, since it reveals the dense 3D geometry in general case. However, the consistency-based stereo is unreliable mainly in the following three cases:

1) Textureless or texture-repetitive object In textureless region, the intensities of adjacent pixels are resemble, which makes it almost impossible to distinguish the right matches. Therefore, the depth estimation of textureless regions is unreliable and leads to the unfaithful results. In a similar way, texture-repetitive region is very likely to mislead the stereo matcher and results in the incorrect depth.

2) Too large matching range In the representative stereo matching problem, we need to find a right pixel in reference image corresponding to a certain pixel in local image. If there is no constraint, the searching range can be huge, leading to ambiguous matching and huge computational work.

3) Image degradation Image degradation contains image noise, motion blur, out-of-focus blur, lens distortion, etc. The noise of image is inevitable due to imaging mode and environment disturbances, but can be reduced by enhancing the illumination and adopting high quality sensors. Motion blur and out-of-focus blur both confuse the detail features in image and reduce the reconstruction accuracy. Lens distortion is a deviation from rectilinear projection which is common in short focal lens camera. A high level of lens distortion may lead to misalignment and even calibration failure.

To supplement the defects above, various priors are employed to improve consistency-based stereo. Here we give a definition of "prior" in image-based 3D modeling method: the supplementing information which reflects the characteristic of objects, images, or environments, and could be used to make the consistency-based 3D reconstruction efficacious, robust, or for special application.

Prior information has played an important role in image-based 3D modeling method to supplement the fragile consistency-based reconstruction. As the traditional matching methods have been studied sufficiently, the prior information determines the limit of best performance in image-based 3D modeling methods. In this survey, we concern the role of prior in 3D modeling problem, analyze and compare different kinds of priors in state-of-art method, and discuss the poten-

tial development of 3D modeling method. First, the priors presented in the past few decades are categorized into four types, and a systematic review of different prior applications is provided. Then we summarize the prevailing algorithm that adopts prior to generate better result. In the end, we introduce the trend and challenges of the image-based modeling problem.

3 Classical priors

Classical prior takes advantage of general attribute, material characteristics and illumination, which are assumptions that generally exist in various targets. In this section, we will discuss the classical priors in three categories: smooth/continous, silhouette and illumination.

3.1 Smooth and continuous prior

It is intuitive that most of the objects should be smooth and continuous, so these prior knowledge is studied at the earliest period. Generally speaking, smooth and continuous priors are applied to 3D modeling problem in two kinds: local methods and global methods. The former approaches smooth each part of the model according to local cues, which refers to small neighbor patch/window on disparity/depth image, or adjacent vertex on mesh. The local methods usually introduce hierarchy or iterative framework to ensure efficiency and robustness. In contrast, global algorithms rifely formulate energy function with extensive information and then minimize it. There are several mathematical translations of smooth hypothesis in this framework, and the smooth terms are various in different minimization algorithms, like Level set, Graph cuts, Variational, etc.

We first review the smooth prior in local means. Segment-based stereo [20] regards the scene structure as a set of planar surface patches because of continuous prior. Firstly, they adopt an over-segmentation on reference image and then generate a set of planar hypotheses for each segment. A series of algorithms [21–24] have been proposed to optimize the estimation of over the planar hypotheses, making the excellent results on the Middlebury stereo evaluation [25]. However, as these methods assume the scene is local planar, they cannot permit curved surface, and may lead to layering mesh when applied in 3D modeling problem. Furukawa et al. [26] provide a stereopsis pipeline that consists of match, expand, filter to generate a dense set of small rectangular patches covering the surface. Furthermore, an additional method could turn the resulting patch model into a mesh which can be further refined. The key idea of the expanding step is to spread the initial matches to nearby pixels due to the continuous prior, then the error matches are eliminated using visibility constraints in filtering step. These methods take advantage of smooth and continuous assumption to build a simple patch-like 3D model, which is convenient in computing and processing. The main drawback is that the patch-like model is open and sparse, as the global relation of different patches is limited. Also, patch based model neglects detailed shape.

Another local approach smooths the surface in disparity/depth image phase. As the value of each pixel represents the position of each point in 3D coordinate, traditional blur algorithm for intense image, like Gauss blur and Laplace blur, could be directly used in depth image. These methods effectively smooth the surface and filter out outliers in a limited window. Also, it is achievable to parallelize the computation, reaching the real-time level. Beeler et al. [27] add photometric and surface consistence in an iterative local framework to guide the smooth process. The sub-pixel disparity is updated in every iteration as a combination of surface-consistency term and photometric-consistency term: the surface consistency term alter the disparity according to normalized cross correlation (NCC) value of the contiguous region, and photometric consistency make depth in the textured areas of the image count more, so the surface is smoother in textureless area and accords with passive stereo result in texture area.

The defect of local means is its deficient effect in sparse and noisy data, since only small range of data are involved. Global methods, involving an energy function minimization problem, have attracted much concerns in the past decades. The smooth and continuous property is embodied as regularization, which penalises high frequency noise ingredient, forming smooth shape. The universal energy function to be minimized in stereo problem is the sum of data term $E_{data}(D)$ and smooth term $E_{smooth}(D)$, as follow:

$$E(D) = E_{data}(D) + E_{smooth}(D). \quad (1)$$

Here D is undetermined disparity map of local image and reference image, $E_{data}(D)$ is formulated with photo-consistency, $E_{smooth}(D)$ is formulated with smooth priors. We concentrate on the smooth term in following discussion.

Kolev et al. [28] enforce the smoothness implicitly by minimizing the corresponds to find the minimal surface with respect to Riemannian metric:

$$E_{smooth}(D) = \int_S (1 - \exp(-\tan(\frac{\pi}{4}(c(x) - 1))^2)/\sigma^2)\mathrm{d}S, \quad (2)$$

where S is estimated surface, $c(x)$ represents the photo-consistency in terms of normalized cross-correlations (NCC).

Liu et al. [29] formulate a variational energy function with the smooth term that represents the interaction between neighboring pixels:

$$E_{smooth}(D) = \int_a^b \psi_s(|\bigtriangledown D|^2)\,\mathrm{d}x, \tag{3}$$

where $\psi_s = \sqrt{s^2 + \epsilon^2}$ is a robust operator, and ϵ is a small positive constant. Yao et al. [30] further add the second-order gradient term to reinforce the smoothness of local areas. The variational methods have achieved a preferable result in Middlebury evaluation on both accuracy and integrity. The weakness of variational methods is that the result tends to get stuck into local optima, so it requires fair initial value to guarantee the right convergence. Li et al. [31] propose to encourage the second and third derivatives of depth to be zero as priors on slanted and curved surfaces. Woodford et al. [32] further stress the availability of second-order priors on the smoothness of 3D surfaces, the smooth term is formulated as:

$$E_{smooth}(D) = \sum_{\mathcal{N}\in\mathbb{N}}^{n} W((\mathcal{N})\rho_s(S(\mathcal{N}, D)). \tag{4}$$

Here, $S(\mathcal{N}, D)$ is the smoothness of a neighborhood $\mathcal{N}$, and $\rho_s(.)$ is smoothness cost function. $W(\mathcal{N})$ is an additional per-neighborhood conditional random field (CRF) weight.

Deformable model based algorithms like [33,34] generally combine smooth constraint with surface grid topology. Han et al. [33] employ the curvature force in level-set framework as a regularization force to counteract with the effect of image noise. In latter work, Esteban et al. [34] propose to formulate the regularization term as Laplacian [35] deformation to smooth the surface. In brief, the internal regularization tries to move a given mesh point v to the gravity center of its 1-ring neighborhood. The smooth force F_{smooth} is formulated as:

$$F_{smooth}(\mathbf{v}) = \frac{1}{m}\sum_{j\in\mathcal{N}_\infty(\mathbf{v})} \mathbf{v}_j - \mathbf{v}, \tag{5}$$

where $\mathbf{v}_j$ is the 1-ring neighbors of vertex $\mathbf{v}$, m is the total number of these neighbors. Zeng et al. [36] propose to integrate the local prior in a space carving framework, solved by graph-cut optimization. Instead of applying the smoothness term on the whole surface at once, they apply it on each patch separately, overcoming the parameterization limitation of the global approach. Their method reconstructs fairly smooth result and recovers more detail than level-set algorithm. Tasdizen et al. [37] generalize anisotropic techniques which have been useful in image processing to surface reconstruction by minimizing the second-order penalty functions. The experiment results show that the algorithm does well in preserving creases while denoising the input. The main shortcoming is the vast computation time.

Smooth constraints of local and global method have merits and demerits respectively. Global methods produce more holistic smooth model, recovering more reasonable shape in featureless region, but require high computational cost. Experiments show that they could even achieve noise suppression and hole-filling on result models. Local methods are generally faster since the processing time is greatly reduced by parallelizing the algorithm. Therefore, some high efficiency demanding applications prefer local smooth methods.

3.2 Silhouette prior

Silhouette is the binary image that segments foreground object from the original image. Many 3D reconstruction studio's wall is designed to be single colored, making it easier to obtain silhouettes. In nature scene, there are still methods to achieve segmentation. Typical segmentation assumes that background pixel values are constant, whereas foreground pixel values vary. The well-known "Lazy Snapping" [38] achieves a coarse-to-fine segmentation approach with user interface, which leads to a fine silhouette using few clickings on the image. However, the automatic segmentation remains to be error-prone. Multi-view Segmentation methods [19,39] take advantage of multi-camera cues to make a robust automatic segmentation. The development of the segmentation algorithm makes it possible to utilize silhouette prior in automatic 3D modeling.

Silhouette priors are generally adopted in two ways:

In the first case, the maximal solid shape, namely visual hull, is firstly built using the silhouettes image of a series of viewpoints, then this rough shape is used as an initial model to constraint the subsequent process. Visual hull is firstly defined in [18], and several excellent algorithms [40–42] are proposed to solve this problem efficiently. Along with the camera viewing parameters, the silhouette defines a back-projected generalized cone that contains the actual object, then the visual hull is extracted as the intersection of every cones. When view points of images are dense and surround the object in various direction, the visual hull is complete and closed, which is suitable for initial estimation of one object. On the contrary, if the view points are few and adjacent, the visual hull will be open for extension and lack shape details.

The study of visual hull motivates the research on subsequent refinement. Esteban et al. [34] propose a deformable mesh model, which comes from the visual hull, to fuse texture and silhouette information together. The construction is

limited by the grid relation in visual hull. Later global optimizing reconstruction algorithms including [28,29,43–45] use the visual hull to initialize the photometric error function. The initial value is important to these methods because most of energy minimization algorithms are easy to fall into incorrect local minimal if the starting value is too deflecting. Also, the visual hull indicates the bounding box of the object, thus reduces the matching calculated amount. One of the limitations of using visual hull is that the initial model loses sight of sunken part, which means it is more suitable for convex object construction. Also, the method can not reconstruct some large scale scenes, like indoor scene and aerial photo reconstruction, because the shape cannot be carved from the inner visual cones.

In the second case, silhouettes are used to refine a pregenerated model in a gradual way. In point clouds edit, silhouettes are used to filter out the noisy points that fly away from truth shape, thus reduce the ambiguity and improve the accuracy [26,29,46]. In mesh edit, silhouettes help to move the vertices of the mesh towards the corresponding edge contours in multi-view images [47–49].

Sinha et al. [46] put forward to optimize photo-consistency and smoothness in a global graph-cut framework, then reconstruct a surface that exactly satisfies all the silhouettes. Vlasic et al. [48] propose to track human motion by deforming the template into the multi-view silhouettes in a non-rigid way. In their pipeline, first the template and skeleton are registered with the visual hull using linear blend skinning (LBS), then the Laplacian coordinates are used to preserve mesh detail while satisfying silhouette constraints. Gall et al. [47] adopt a Laplacian deformation framework to edit mesh model, constraining the projection of the vertices to lie on 2D positions on the image silhouette boundary. The refined surface is reconstructed by solving a least-square problem which minimize the distance between silhouette and mesh rim. The methods are widely used in motion track system [50], where detailed shape is not so important. Straka et al. [49] further use this technique to estimate hand or body shape from skeleton with multi-view silhouettes.

The strength of silhouette prior is its determinacy, which complements the ambiguity in the pixel-consistency match. However, the limitation of silhouette prior is distinct: first, the use of the silhouettes requires an accurate extraction of the object from the background, which is not easy to achieve in complex scene. Also, silhouette neglects concave details, and may produce a error-prone result when viewpoints are limited. As silhouettes based refinements focus on certain convex shape but not detail geometry, they are widely used in motion or skeleton track [48–50].

3.3 Illumination prior

Illumination has played a key role in image-based 3D reconstruction method. The basic function of illumination is cooperating with camera to acquire high quality images, which is bright and low-noise. In a further progress, illumination is used to generate structure information, the related research refers to shape from shading (SFS), photometric stereo (PS), and some active methods using specific illuminant. Shape from shading is first introduced by Horn [51] in 1970, which means recovering shape from a gradual variation of shading in the image. Photometric stereo is first described by Woodham [16] in 1980, and they estimate surface normal of a Lambertian object with the help of three distant light sources and corresponding images. Both problems have been studied for several decades, and some brilliant surveys have revisited the existing methods in SFS [15,52] and PS [53]. In this section, we will not repeat the work in aforementioned surveys, but concentrate on means that apply illumination information to achieve a synthetical reconstruction system, and reveal the role of illumination prior.

Shape from shading and multi-view stereo are naturally complementary to each other. The traditional multi-view stereo(MVS) methods are good at generating rough shape, but neglect the high-frequency details. In contrast, shape from shading methods concentrate on shading cues, which recover details in pixel level. Many algorithms have been proposed to integrate SFS and MVS to produce a more accurate result.

The early work [54,55] uses stereo or multi-view result as initialization, and then refines the model with SFS algorithm. Refs. [56–58] further combine MVS and SFS with a series of variational algorithms, achieving more detailed results. The aforementioned methods all make a simple hypothesis about the illumination, like a single distant light source, or unity of a distant point light source and uniform ambient illumination. They also assume that all the surface is Lambertian with no self-shadowing.

The latter studies try to break the constraint of specific illumination and Lambertian hypothesis, making the illumination prior easily applied into general scene. Yu et al. [59,60] consider reflectance model of non-lambertian object and take into account the effects of self-occlusions and self-shadows. The rough shape is iteratively optimized using multi-view images with Phong or Torrence-Sparrow model. Yoon et al. [61] formulate a global cost function with respect to both shape and reflectance in a variational framework, and use gradi-

ent decent to solve it. This method applies to a number of classical scenarios, and considers general dichromatic surfaces besides Lambertian surfaces. Wu et al. [62] propose a method that combines MVS and SFS for general, unknown illumination. They use low-frequency spherical harmonics and wavelet to model the illumination, assuming a single lighting condition and explicitly handling self-shadowing. Han et al. [63] further study the lighting model in natural scene, and introduce a general lighting model consisting of global and local lights. The result shows the accurate performance in uncontrolled natural illumination conditions, while their method considers only uncontrolled natural illumination conditions accurately.

Unlike shape from shading, photometric stereo (PS) utilizes images taken under different lighting conditions to fully constrain the surface normal. Many early researches work on extending PS to non-Lambertian surfaces. Zhang et al. [64] take all structures from motion, photometric stereo and multi-view stereo into account, making a dense reconstruction of textured and texture-less surfaces from a monocular image sequence. Their main contribution includes stereo matching with changes in lighting and photometric stereo for moving scenes, but the pipeline is not robust to complex scenes with occlusion and mixed lighting. Basri et al. [65] present a photometric stereo framework using a first order harmonic approximation, or a second order harmonic approximation. Both methods produce favorable result but require at least four or nine images with various illumination as input. Their work is still based on Lambertian assumption, but is robust under general unknown lighting conditions. Beeler et al. [27] propose to recover mesoscopic geometry that can not be reconstructed by stereo algorithm with a simple illumination model. The method is based on prior that the incoming light in small concavities (like pores on face) is blocked and the point thus appears darker. Therefore, they extract mesoscopic gradient information in images to reconstruct local high-frequency gradient geometry. The rendering mesh is vivid like real human skin texture, but not metrically correct. Hernández and Brostow [66,67] extend photometric stereo to multi-spectral method. Their setup consists of an ordinary video camera and three colored light sources, which emit red, green, and blue light from different directions. As the reflection of three kinds of lights will not be mixed, they obtain more information to produce better normal maps. They have achieved fair reconstruction on human faces and cloth.

Roth et al. [68] combine photometric stereo-based methods with face alignment techniques, recovering face model from the unconstrained face images. They prepare an enhanced 3D template with surface normal from a generic 3D face template, of which 3D landmarks are consistent with the estimated 2D landmarks on all images. In reconstructing phase, 2D face images at all poses are back projected onto the 3D surface, where the collection of projections will form a data matrix spanning all vertices of the template. The method inherits the advantages of PS, meanwhile maintains the consistency of the overall shape with 2D landmarks, making it feasible to unconstrained images.

In brief, illumination priors recover an even detailed geometry, which is an appropriate supplement to consistency based stereo. The prior strongly relies on assumption of albedo and illumination. Although more adaptable and efficient reflectance models are presented, the results are still unstable especially when the object is consist of various materials. For example, texture-copying, which means producing false shape fluctuate in textured region, is one of the unavoidable problem in SFS issue. As for illumination, nature lighting is more universal, but increases the reconstruction difficulty. Therefore, complex studio with elaborate lighting system, like Refs. [66,67,69–72], is preferred when modeling exquisite geometry.

The aforementioned classical priors are summarized in Table 1. The application and comparison of various classical priors are demonstrated. As shown in Fig. 2(b), with no constraint, mesh cannot be reconstructed as the point cloud is scattered. Figure 2(c) is the rough model generated using EPVH [40], and Fig. 2(d) is reconstructed using method [29] with global smooth prior. Figure 2(e) is refined by illumination prior [62], which strengthen detailed shape. In Fig. 3, the ground truth (Fig. 3(a)) is scanned using Kinect and reconstructed using KinectFusion [5]. Models (Figs. 3(b)–3(e)) are rendered according to error distance. Specifically, Fig. 3(b) is generated using EPVH [40], the model is rough and miss sunk geometry. Figures 3(c) and 3(d) are reconstructed using local smooth prior [27] and global smooth prior [29]. These priors smooth the surface but cannot tackle the shape decay in textureless region (e.g., hair, trousers and shoes). Figure 3(e) is the refined model of Fig. 3(d) using silhouette adaption method [47], which restores some of the decay regions. Figure 4 mainly compares the effect of smooth prior and illumination prior.

4 Specific priors

In recent years, increasing image-based 3D modeling researches put forward specific priors to break the limitation

Table 1 Comparison of different classical priors

Prior	Strength	Limitation	Time
Smooth and continuous (local)	• Suitable for most object and scene • Low computational-cost	• Tend to fail when texture is too sparse • Weaken high-frequency details • Cannot tackle discontinuity case	Short
Smooth and continuous (global)	• Make reasonable continuous model • Take global depth into account	• High computational-cost • Unmanageable for the trade-off between smooth and high-frequency details	Long
Silhouette (visual hull)	• Low ambiguity • Constrain the bounding of object	• No sunk details • Need adequate images in multi-view to make a whole model	Short
Silhouette adaption	• Effectively refine convex shape • Constrain the bounding of object	• The relation of silhouette and surface is ambiguous • Easy to fail in occlusion part	Relatively long
Fixed illumination	• Emphasis on high frequency detail • Do not need any Additional Settings	• Ill-posed problem, usually need preliminary shape • Texture-copying problem • Subject to specific albedo and lighting assumption	Relatively short
Varying illumination	• Emphasis on high frequency detail • More robust than unknown lighting method	• Require complex lighting setup • Cannot reconstruct object that is too large to be placed in a studio	Medium
Multispectral light	• More robust for matching • Support other illumination priors	• Require complex lighting setup • Cannot reconstruct object that is too large to be placed in a studio	Relatively short

Fig. 2 Reconstruction result of a model with different priors. (a) One frame of image sequence; (b) point cloud with no constraint; (c) shape from silhouette (visual hull); (d) consistency matching with smooth prior; (e) refined using ilumination prior; (f) rendered in color

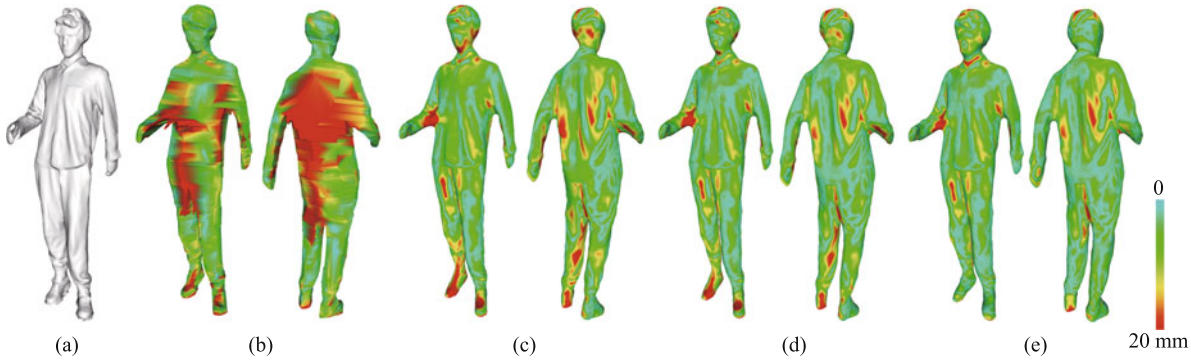


Fig. 3 Reconstruction result of the model with different priors and quantitive evaluations. (a) Ground truth; (b) visual hull (silhouette prior); (c) MVS with local smooth prior; (d) MVS with global smooth prior; (e) MVS (global smooth) + silhouette adapt

of universal multi-view reconstruction systems. Comparing to classical priors, specific priors provide more forceful constraint, which reflects deep characteristic and generates fine result in some challenging case. In exchange, these approaches are certain to decay when object does not meet the supposed case. Currently, the specific priors become more

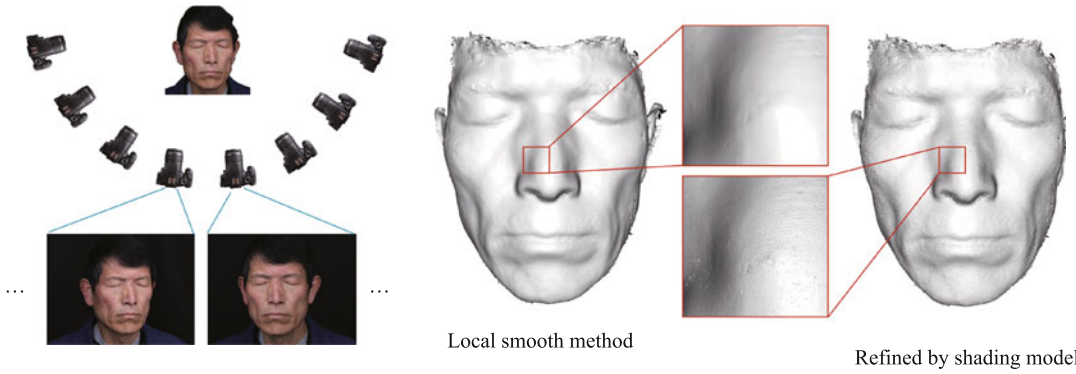


Fig. 4 Face reconstruction using Beeler's algorithm [27], the left model takes advantage of local smooth method, while the right one is refined with shading model, which recovers micro structure

and more intelligent, and improve the 3D reconstruction quality in many major applications.

We divide specific priors into following categories:

• **Manhattan prior** To strengthen the geometric constraint in artificial buildings or indoor scenes, Manhattan priors [73] are introduced in Refs. [74,75]. Manhattan priors assume that all surfaces in the world are aligned with three dominant directions, typically corresponding to the x, y, and z axes. Furukawa et al. [74] detect dominant orientations that most of the geometries lie, and then assign one of the candidate planes to each pixel in the image using Markov random field. In this way, the conventional smoothness prior is replaced with a structured model of axis-aligned planes, forming depth image with complex surface. Vanegas et al. [75] extend previous work to model entire buildings from oblique-angle aerial images by using grammar-based techniques, which describes a compact set of transitions between consecutive floors. The method employs an extruded bounding box of the building footprint extracted from GIS data as the initial models, then rewriting rules are used to perform transitions to the floors of the initial 3D model, producing a refined building model. Zeisl et al. [76] focus on building interiors reconstruction, and replace the Manhattan-world assumption by a modified prior: the space of building interior is bounded by parallel ground and ceiling planes, and purely vertical structures, like walls and doors. The Manhattan prior based algorithm produces clean and simple models in indoor scenes and outdoor buildings, even if the area is textureless.

• **Piecewise planar prior** Piecewise planar prior assumes that the object consists of many piecewise planes, so majority of these methods take advantage of over-segmentation result to guide the stereo matching model (e.g., [23,77]), or initialize the disparity maps using segmentation prior (e.g., [78]).

The segmentation based method reduces the ambiguities caused by textureless to some extent. On the base of color segmentation, plane-fitting algorithm [79,80] is put forward to recover dominant scene planes, which is not limited to Manhattan scene, like [74].

Gallup et al. [81] further distinguish piecewise planar and non-planar region in multi-view stereo. They first segment an image into piecewise planar and non-planar region using a classifier, which is pre-learned from hand-labeled planar and non-planar image regions. Then the piecewise planar region is recovered with planar while the non-planar regions are modeled with standard multi-view stereo algorithm. Kim et al. [82] propose the two-stage method for piece-wise planar scene reconstruction. First initial planes are allocated to each segment, then the planes are refined with non-linear optimization, generating a filtered and more accurate result. Mathias et al. [83] introduce shape grammars to SFM and image-based analysis, composing a system that improves modeling by automatically specializing the applied detectors.

• **Multiple geometry prior** Obviously, it is insufficient to use merely plane to describe complex scene, therefore, multiple geometry prior incorporates various shape templates in the reconstruction framework.

Zebedin et al. [84] take advantage of both planes and surfaces of revolution template prior, generating a much broader family of roof shapes. Their approach is mainly used for automatic urban modeling from aerial imagery, so the prior is specially designed for roof reconstruction. Wu et al. [85] propose swept surface prior to make a schematic surface representation, which is preferred by architects. They assume that the architectural scene is consist of transport curves which lie in planes parallel to the ground, and profile curves which lie in planes that are orthogonal to the ground. Such prior is extraordinarily suitable to outdoor architecture, and the method could even tackle point clouds generated from SFM, which is comparatively sparse and incomplete. Lafarge et al. [86,87] regard urban scenes as a combination of meshes and geo-

metric primitives, including planes, spheres, cylinders, cones, and tori. They make an efficient iterative mesh editing to produce a compact model, where detail elements are described by meshes and regular structures are described by primitives. Mahabadi et al. [88] present a shape prior which splits the object into multiple convex parts. The idea of this shape prior comes from Wulff shape, the equilibrium shape of a crystal, for which it is natural that the input object shapes are split into convex or almost convex segments. The first step is achieving a volumetric multi-label segmentation, then each of the transitions between labels is penalized with its individual anisotropic smoothness term. The means promote the robustness and accuracy of textureless object modeling, while the experiment results are limited to the object in relatively simple geometry.

• **Example based prior** Example based priors are designed for some certain objects, like face, hair, buildings, plants, etc. These priors have the most specific templates and achieve extremely difficult reconstruction.

Face modeling using monocular image strongly relies on template methods due to its lacking depth cues. Early work has been well reviewed in [89]. Existing algorithms recover shape from silhouette, shape form shading, but the most successful approach of that time is analysis by synthesis [90,91], where the parameters of the 3D statistical model are adjusted to match 2D face image to the reconstructed face, namely, a mean template. As depth cues from one single image are fairly limited, many studies adapt multi-view method, cooperating with template prior, to make more accurate and robust face reconstruction. Cheng et al. [92] propose to extract face features using structure from motion/silhouette, and then adapt these features to a generic face model using radial basis function interpolation in 3D space. Fidaleo et al. [93] introduce a deformable generic face model at the pose estimation, face segmentation, and preprocessing stages, improving the robustness and flexibility. Tytgat et al. [94] further employ 2D morphing techniques for generating an animated, personalized 3D model in real time. Baumberger et al. [95] present a robust reconstruction system using a statistical shape model and facial landmarks, which are defined in the first frame. The pipeline is proved to be robust and efficient, and the rendering model looks rather fine. Roth et al. [68] and Kemelmacher-Shlizerman et al. [96] propose to recover 3D human face model from vast internet images. Their methods integrate facial landmarks driven and unconstrained image aligning technique, achieving an unconstrained face modeling system.

Besides human face, example-based methods have been developed in 3D reconstruction of eyes [97,98], hair [99,100], structures [101,102], plants [103–105], etc. The main idea of these method is combining traditional image based stereo with specific template priors to simulate the ideal model.

• **Trained category prior** Comparing to example based priors, trained category priors do not employ fixed pipeline, but try to explore deep constraint from big data or statistics. Therefore, these methods have the potential to be expanded to different groups of objects.

Blanz et al. [106] recover 3D surface using a statistical method. Their system relies on a dataset of 3D scans, which are converted into a vector space representation (Morphable Model). Then the missing vertex coordinates are inferred by estimating the probability density of 3D faces. The regularization trades off between fitting the surface to the feature points and producing a plausible solution in terms of prior probability. Therefore, the method requires sparse features and is more robust than traditional consistency based reconstruction.

Bao et al. [107] propose a semantic structure from motion (SSFM), which takes advantage of semantic and geometrical properties associated with objects besides geometry constraints to recover structure of the scene. In follow-up work [108], he introduces semantic information as prior to promote performance of multi-view stereo. Their method includes two phase: learning and reconstruction. In learning stage, 3D models of a set of samples are scanned in advance, then they model semantic similarity as a shape prior which consists of a set of automatically learned anchor points and a learned mean shape. In reconstruction stage, the shape variation across instances and capturing semantic similarities are combined to generate a fine model. The method has shown its superiority on some challenge cases like textureless spherical fruits or car models.

Dame et al. [109] integrate dense SLAM with 3D shape and pose recovery. Initially, a dense representation of the scene is reconstructed using photo-consistency. After that, an object-class detector is used to identify the object, recovering the 6D pose and geometry. The system reconstructs unseen part of the object, and reduces the erroneous possibility in SLAM. Hane et al. [110] propose to incorporate shape priors based on surface normal distributions into convex multi-label optimization. The object class specific shape prior is formulated in the form of spatially varying anisotropic smoothness terms, of which parameters are extracted from the training data. This kind of prior can be generalized to various classified shapes and improves the robustness considerably.

Specific priors explore deep-level rules of the objects in the same category. The aforementioned priors are summarized

in Table 2. Figure 5 [26,74,81,86,99,103,108,111] illustrates the improvement of specific priors in various cases. Although these priors are limited in some certain occasions or require pre-training process, they are more efficient in targeted assignment than classical priors, opening up a new prospect to solve challenging problem in image-based 3D modeling field.

5 Future trend

Image-based 3D modeling has been studied for several decades, and the study of consistency based methods reaches bottleneck constraint. The introduction of prior knowledge promotes the 3D reconstruction in both accuracy and efficiency. In the future study, prior will still be the hot topic in 3D modeling field. We predict that the future trend will focus on two aspects:

1) More intelligent prior The emerging researches of machine learning are gradually revealing the way of human intelligence, and are successfully applied in many computer vision fields. In 3D modeling field, intelligent prior means having the ability to learn from the big data and enhance the reconstruction quality. Several recent studies have explored the feasibility to take advantage of learning algorithm to enhance 3D modeling quality. Zhang et al. [112] explore the method to learn from huge visual data, then reconstruct the 3D model from a single image using category detector. Bao et al. [108] take advantage of learned category-level shape priors and object detection to enhance multi-view stereo. Mehrdad et al. [113] propose a content based descriptor which employs histogram of local orientation (HLO) as a geometric property of the shape to retrieve 3D models. In addition to this, data-learned methods are used to optimize normals [114], enhance matching confidence [115] and perform large-distances regularization [116]. We have seen that the trend to combine track, categorization and recognition with 3D reconstruction, forming more intelligent ways to utilize big visual data.

2) Practical and challenging application Although 3D reconstruction problem is studied for a long period of time, this technology enters daily life merely in recent five years. Image-based 3D reconstruction softwares, like PhotoScan[1)] and 123D Catch[2)], achieve classical 3D modeling pipeline, however, their efficiency and robustness are still limited. Refs. [117,118] integrate the developed monocular 3D modeling method on a mobile phone, building a substantially 3D scanning system. The works [68,96] recover 3D human face model from a great many internet images, which are captured in unconstrained case. These study shows the potential to achieve low-cost, convenient and unconstrained 3D reconstruction in the future, and we believe there still exists large room for improvement.

6 Conclusion

Inchoate study on prior focuses on intuitionistic information like smooth, continuous and silhouette, which usually applies to common objects. The introduction of reflectance model takes illumination and surface property into account, developing into structure from shading (SfS) and photometric stereo (PS). All these classical priors are utilized in the most advanced modeling systems, which reconstruct steady and fair model of still things and human body.

Recently, increasing researches attempt to expand image-based modeling method to out-of-lab environment, where images contain much more noise and uncertain influence. Con-

Table 2 Comparison of different specific priors

Prior	Strength	Limitation	Time
Manhattan	• Produce clear model of Manhattan scenes • Resistance to noise • Suitable for texture-sparse region	• Specific to certain scene • Lack of details	Short
Planar piece	• Make preferable result for building or indoor scenarios • Resistance to noise	• Specific to certain scene • Lack of details	Relatively short
Geometric template	• Identify existing model • Greatly improve the reconstruction result in some challenge cases	• Require strong manual inference • Need pre-category	Medium
Example based	• Focus on decent reconstruction of some specified objects	• The range of application is limited in extremely small category • Do not ensure high accuracy	Medium
Category learning	• High roust result • Relatively wider application range than example based methods	• Require pre-training process • May be prone to mean template	Relatively long

[1)] http://www.agisoft.com
[2)] http://www.123dapp.com/catch

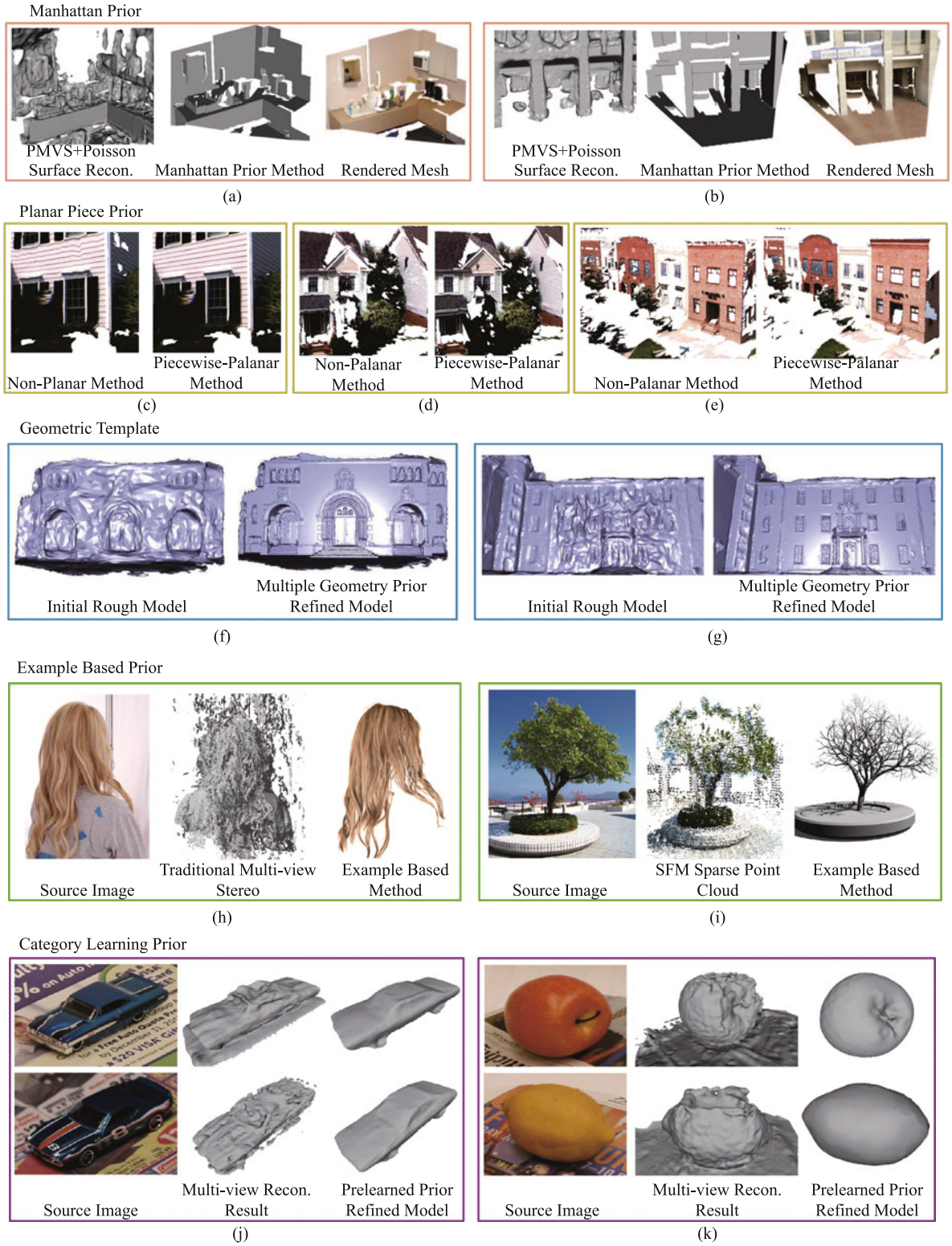


Fig. 5 The improvement of specific priors in various cases. (a) and (b) [74] demonstrate that the manhattan prior greatly improves accuracy in office and hall reconstruction; (c)–(e) [81] compare the result of non-planar method [26,111] and piecewise planar method in street building scenes; (f) and (g) [86] show the effect of geometric template prior which refines the building model; (h) [99] and (i) [103] take advantage of example based prior to reconstruct hair and trees; (j) and (k) [108] show one of the category learning prior method which reconstructs some extremely challenging objects, including toy cars and fruits (In (j) and (k), the multi-view model is reconstructed using [26,111]. The specific prior has shown its superiority on some extremely challenging modeling problem)

sidering multifarious application, targeted priors have greater potential to popularize 3D modeling technique into more practical and simple use. Geometry priors like piecewise planar and Manhattan assumption greatly improve the feasibility of urban reconstruction, meanwhile, example based prior opens the door to some extreme challenging object, including hair, face, trees, so on and so forth. Training and learning method makes it possible to reconstruct a class of object robustly and efficiently.

The priors to supplement ambiguous shape estimation have been studied for a long time but still cannot be completely solved. The enhancement of classical cues is limited when images quality is low and lighting is unstable. By involving geometry and template prior, reconstruction is more efficient, but current assumptions such as planar piece do not work for more complex object. In the future research, we believe more intelligent prior and practical application are two hot topics, and there still exists large room for accuracy and practicability improvement.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 61371166, and 61422107) and the Natural Science Foundation of Jiangsu Province, China (BK20130583).

References

1. Dyer C R. Volumetric scene reconstruction from multiple views. Foundations of Image Understanding, 2001, 628: 469–489
2. Slabaugh G, Schafer R, Malzbender T, Culbertson B. A survey of methods for volumetric scene reconstruction from photographs. In: Proceedings of the Joint IEEE TCVG and Eurographics Workshop in Stony Brook. 2010, 81–100
3. Seitz S M, Curless B, Diebel J, Scharstein D, Szeliski R. A comparison and evaluation of multi-view stereo reconstruction algorithms. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2006, 519–528
4. Brenner C. Building reconstruction from images and laser scanning. International Journal of Applied Earth Observation and Geoinformation, 2005, 6(3): 187–198
5. Newcombe R A, Izadi S, Hilliges O, Molyneaux D, Kim D, Davison A J, Kohi P, Shotton J, Hodges S, Fitzgibbon A. KinectFusion: real-time dense surface mapping and tracking. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. 2011, 127–136
6. Izadi S, Kim D, Hilliges O, Molyneaux D, Newcombe R, Kohli P, Shotton J, Hodges S, Freeman D, Davison A, Fitzgibbon A. KinectFusion: real-time 3D reconstruction and interaction using a moving depth camera. In: Proceedings of the ACM Symposium on User Interface Software and Technology. 2011, 559–568
7. Scharstein D, Szeliski R, Hirschmüller H. Stereo. http://vision.middlebury.edu/stereo/, 2015
8. Huang T S, Netravali A N. Motion and structure from feature correspondences: a review. Proceedings of the IEEE, 1994, 82(2): 252–268
9. Oliensis J. A critique of structure-from-motion algorithms. Computer Vision and Image Understanding, 2000, 80(2): 172–214
10. Lowe D G. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2004, 60(2): 91–110
11. Durrant-Whyte H, Bailey T. Simultaneous localization and mapping. IEEE Robotics & Automation Magazine, 2006, 13(2): 99–110
12. Williams B, Klein G, Reid I. Real-time SLAM relocalisation. In: Proceedings of the 11th IEEE International Conference on Computer Vision. 2007, 1–8
13. Newcombe R A, Lovegrove S J, Davison A J. DTAM: dense tracking and mapping in real-time. In: Proceedings of the IEEE International Conference on Computer Vision. 2011, 2320–2327
14. Tan W, Liu H M, Dong Z L, Zhang G F, Bao H J. Robust monocular SLAM in dynamic environments. In: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality. 2013, 209–218
15. Zhang R, Tsai P S, Cryer J E, Shah M. Shape-from-shading: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1999, 21(8): 690–706
16. Woodham R J. Photometric method for determining surface orientation from multiple images. Optical Engineering, 1980, 19(1): 1–22
17. Bartoli A, Gerard Y, Chadebecq F, Collins T, Pizarro D. The visual hull concept for silhouette-based image understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(10): 2099–2118
18. Laurentini A. The visual hull concept for silhouette-based image understanding. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1994, 16(2): 150–162
19. Lee W, Woo W, Boyer E. Silhouette segmentation in multiple views. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(7): 1429–1441
20. Tao H, Sawhney H S, Kumar R. A global matching framework for stereo computation. In: Proceedings of the 8th IEEE International Conference on Computer Vision. 2001, 532–539
21. Bleyer M, Gelautz M. A layered stereo algorithm using image segmentation and global visibility constraints. In: Proceedings of the IEEE International Conference on Image Processing. 2004, 2997–3000
22. Hong L, Chen G. Segment-based stereo matching using graph cuts. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004, 74–81
23. Klaus A, Sormann M, Karner K. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In: Proceedings of the 18th IEEE International Conference on Pattern Recognition. 2006, 15–18
24. Yang Q X, Wang L, Yang R G, Stewénius H, Nistér D. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3): 492–504
25. Scharstein D, Szeliski R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision, 2002, 47(1–3): 7–42
26. Furukawa Y, Ponce J. Accurate, dense, and robust multiview stereop-

sis. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2010, 32(8): 1362–1376

27. Beeler T, Bickel B, Beardsley P, Sumner B, Gross M. High-quality single-shot capture of facial geometry. ACM Transactions on Graphics, 2010, 29(4): 40
28. Kolev K, Klodt M, Brox T, Esedoglu S, Cremers D. Continuous global optimization in multiview 3D reconstruction. International Journal of Computer Vision, 2009, 84(1): 80–96
29. Liu Y B, Cao X, Dai Q H, Xu W L. Continuous depth estimation for multi-view stereo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2009, 2121–2128
30. Yao Y, Zhu H, Nie Y M, Ji X L, Cao X. Revised depth map estimation for multi-view stereo. In: Proceedings of the IEEE International Conference on 3D Imaging. 2014, 1–7
31. Li G, Zucker S W. Surface geometric constraints for stereo in belief propagation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2006, 2355–2362
32. Woodford O, Torr P, Reid I, Reid I, Fitzgibbon A. Global stereo reconstruction under second-order smoothness priors. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(12): 2115–2128
33. Han X, Xu C Y, Prince J L. A topology preserving level set method for geometric deformable models. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2003, 25(6): 755–768
34. Esteban C H, Schmitt F. Silhouette and stereo fusion for 3D object modeling. Computer Vision and Image Understanding, 2004, 96(3): 367–392
35. Sorkine O, Cohen-Or D, Lipman Y, Alexa M, Rössl C, Seidel H P. Laplacian surface editing. In: Proceedings of the ACM SIGGRAPH symposium on Geometry processing. 2004, 175–184
36. Zeng G, Paris S, Quan L, Sillion F. Progressive surface reconstruction from images using a local prior. In: Proceedings of the 10th IEEE International Conference on Computer Vision. 2005, 1230–1237
37. Tasdizen T, Whitaker R. Higher-order nonlinear priors for surface reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2004, 26(7): 878–891
38. Li Y, Sun J, Tang C K, Shum H Y. Lazy snapping. ACM Transactions on Graphics, 2004, 23(3): 303–308
39. Kolmogorov V, Criminisi A, Blake A, Cross G. Probabilistic fusion of stereo with color and contrast for bilayer segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(9): 1480–1492
40. Franco J S, Boyer E. Efficient polyhedral modeling from silhouettes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3): 414–427
41. Matusik W, Buehler C, Raskar R, Gortler S, McMillan L. Image-based visual hulls. In: Proceedings of the 27th ACM Annual Conference on Computer Graphics and Interactive Techniques. 2000, 369–374
42. Miller G, Hilton A. Exact view-dependent visual hulls. In: Proceedings of the IEEE International Conference on Pattern Recognition. 2006, 107–111
43. Vogiatzis G, Torr P H S, Cipolla R. Multi-view stereo via volumetric graph-cuts. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005, 391–398
44. Furukawa Y, Ponce J. Carved visual hulls for image-based modeling. International Journal of Computer Vision, 2009, 81(1): 53–67
45. Zheng Z Y, Ma L Z, Li Z, Chen Z H. Reconstruction of shape and reflectance properties based on visual hull. In: Proceedings of the ACM Computer Graphics International Conference. 2009, 29–38
46. Sinha S N, Pollefeys M. Multi-view reconstruction using photo-consistency and exact silhouette constraints: a maximum-flow formulation. In: Proceedings of the 10th IEEE International Conference on Computer Vision. 2005, 349–356
47. Gall J, Stoll C, De Aguiar E, Theobalt C, Rosenhahn B, Seidel H. Motion capture using joint skeleton tracking and surface estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2009, 1746–1753
48. Vlasic D, Baran I, Matusik W, Popović J. Articulated mesh animation from multi-view silhouettes. ACM Transactions on Graphics, 2008, 27(3): 97
49. Straka M, Hauswiesner S, Rüther M, Bischof H. Simultaneous shape and pose adaption of articulated models using linear optimization. In: Proceedings of European Conference on Computer Vision. 2012, 724–737
50. Liu Y, Gall J, Stoll C, Dai Q, Seidel H, Theobalt C. Markerless motion capture of multiple characters using multiview image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(11): 2720–2735
51. Horn B K P. Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical Report. 1970
52. Durou J D, Falcone M, Sagona M. Numerical methods for shape-from-shading: a new survey with benchmarks. Computer Vision and Image Understanding, 2008, 109(1): 22–43
53. Herbort S, Wohler C. An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. 3D Research, 2011, 2(3): 1–17
54. Leclerc Y G, Bobick A F. The direct computation of height from shading. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 1991, 552–558
55. Fua P, Leclerc Y G. Object-centered surface reconstruction: combining multi-image stereo and shading. International Journal of Computer Vision, 1995, 16(1): 35–56
56. Jin H, Cremers D, Yezzi A J, Soatto S. Shedding light on stereoscopic segmentation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004, 36–42
57. Jin H, Yezz A, Soatto S. Stereoscopic shading: integrating multi-frame shape cues in a variational framework. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2000, 169–176
58. Jin H, Yezzi A J, Soatto S. Region-based segmentation on evolving surfaces with application to 3D reconstruction of shape and piecewise constant radiance. In: Proceedings of European Conference on Computer Vision. 2004, 114–125
59. Yu T L, Xu N, Ahuja N. Recovering shape and reflectance model of non-lambertian objects from multiple views. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2004, 226–233
60. Yu T L, Xu N, Ahuja N. Shape and view independent reflectance map from multiple views. International Journal of Computer Vision, 2007,

73(2): 123–138

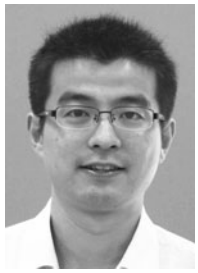
61. Yoon K J, Prados E, Sturm P. Joint estimation of shape and reflectance using multiple images with known illumination conditions. International Journal of Computer Vision, 2010, 86(2–3): 192–210
62. Wu C L, Wilburn B, Matsushita Y, Theobalt C. High-quality shape from multi-view stereo and shading under general illumination. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2011, 969–976
63. Han Y, Lee J Y, Kweon I S. High quality shape from a single RGB-D image under uncalibrated natural illumination. In: Proceedings of the IEEE International Conference on Computer Vision. 2013, 1617–1624
64. Zhang L, Curless B, Hertzmann A, Seitz S M. Shape and motion under varying illumination: unifying structure from motion, photometric stereo, and multiview stereo. In: Proceedings of the IEEE International Conference on Computer Vision. 2003, 618–625
65. Basri R, Jacobs D, Kemelmacher I. Photometric stereo with general, unknown lighting. International Journal of Computer Vision, 2007, 72(3): 239–257
66. Hernndez C, Vogiatzis G, Brostow G J, Stenger B, Cipolla R. Non-rigid photometric stereo with colored lights. In: Proceedings of the 11th IEEE International Conference on Computer Vision. 2007, 1–8
67. Brostow G J, Hernández C, Vogiatzis G, Stenger B, Cipolla R. Video normals from colored lights. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2011, 33(10): 2104–2114
68. Roth J, Tong Y, Liu X. Unconstrained 3D face reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 2606–2615
69. Debevec P. The light stages and their applications to photoreal digital actors. SIGGRAPH Asia Technical Briefs, 2012, 2
70. Ghosh A, Fyffe G, Tunwattanapong B, Busch J, Yu X M, Debevec P. Multiview face capture using polarized spherical gradient illumination. ACM Transactions on Graphics, 2011, 30(6): 129
71. Liu Y B, Dai Q H, Xu W L. A point-cloud-based multiview stereo algorithm for free-viewpoint video. IEEE Transactions on Visualization and Computer Graphics, 2010, 16(3): 407–418
72. Wu C L, Liu Y B, Dai Q H, Bennett W. Fusing multiview and photometric stereo for 3D reconstruction under uncalibrated illumination. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(8): 1082–1095
73. Coughlan J M, Yuille A L. Manhattan world: compass direction from a single image by Bayesian inference. In: Proceedings of the 7th IEEE International Conference on Computer Vision. 1999, 941–947
74. Furukawa Y, Curless B, Seitz S M, Szeliski R. Manhattan-world stereo. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2009, 1422–1429
75. Vanegas C A, Aliaga D G, Beneš B. Building reconstruction using manhattan-world grammars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010, 358–365
76. Zeisl B, Zach C, Pollefeys M. Stereo reconstruction of building interiors with a vertical structure prior. In: Proceedings of the IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission. 2011, 366–373
77. Sun J, Li Y, Kang S B, Shum H Y. Symmetric stereo matching for occlusion handling. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005, 399–406
78. Zhang G, Jia J, Wong T T, Bao H. Consistent depth maps recovery from a video sequence. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(6): 974–988
79. Yang Q X, Wang L, Yang R G, Stewénius H, Nistér D. Stereo matching with color-weighted correlation, hierarchical belief propagation, and occlusion handling. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(3): 492–504
80. Sinha S N, Steedly D, Szeliski R. Piecewise planar stereo for image-based rendering. In: Proceedings of the IEEE International Conference on Computer Vision. 2009, 1881–1888
81. Gallup D, Frahm J M, Pollefeys M. Piecewise planar and non-planar stereo for urban scene reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010, 1418–1425
82. Kim H, Xiao H, Max N. Piecewise planar scene reconstruction and optimization for multi-view stereo. In: Proceedings of Asian Conference on Computer Vision. 2013, 191–204
83. Mathias M, Martinovic A, Weissenberg J, Van Gool L. Procedural 3D building reconstruction using shape grammars and detectors. In: Proceedings of the IEEE International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission. 2011, 304–311
84. Zebedin L, Bauer J, Karner K, Bischof H. Fusion of feature-and area-based information for urban buildings modeling from aerial imagery. In: Proceedings of European Conference on Computer Vision. 2008, 873–886
85. Wu C C, Agarwal S, Curless B, Seitz S M. Schematic surface reconstruction. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012, 1498–1505
86. Lafarge F, Keriven R, Brédif M, Hoang-Hiep V. A hybrid multiview stereo algorithm for modeling urban scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 5–17
87. Lafarge F, Keriven R, Brédif M, Hoang-Hiep V. Hybrid multi-view reconstruction by jump-diffusion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2010, 350–357
88. Mahabadi R K, Hane C, Pollefeys M. Segment based 3D object shape priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 2838–2846
89. Widanagamaachchi W N, Dharmaratne A T. 3D face reconstruction from 2D images. In: Proceedings of the IEEE Digital Image Computing: Techniques and Applications. 2008, 365–371
90. Amin S H, Gillies D. Analysis of 3D face reconstruction. In: Proceedings of the 14th IEEE International Conference on Image Analysis and Processing. 2007, 413–418
91. Hassner T, Basri R. Example based 3D reconstruction from single 2D images. In: Proceedings of the IEEE Computer Vision and Pattern Recognition Workshop. 2006, 15
92. Cheng C M, Lai S H. An integrated approach to 3D face model reconstruction from video. In: Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems. 2001, 16–22
93. Fidaleo D, Medioni G. Model-assisted 3D face reconstruction from video. In: Proceedings of the International Workshop on Analysis and Modeling of Faces and Gestures. 2007, 124–138

94. Tytgat D, Lievens S, Six E. A prior-based approach to 3D face reconstruction using depth images. Advances in Depth Image Analysis and Applications, 2013, 32–41
95. Baumberger C, Reyes M, Constantinescu M, Olariu R, Aguiar E, Oliveira-Santos T. 3D face reconstruction from video using 3D morphable model and silhouette. In: Proceedings of the 27th IEEE SIBGRAPI Conference on Graphics, Patterns and Images. 2014, 1–8
96. Kemelmacher-Shlizerman I, Seitz S M. Face reconstruction in the wild. In: Proceedings of the IEEE International Conference on Computer Vision. 2011, 1746–1753
97. Bérard P, Bradley D, Nitti M, Beeler T, Gross M. High-quality capture of eyes. ACM Transactions on Graphics, 2014, 33(6): 1–12
98. Bermano A, Beeler T, Kozlov Y, Bradley D, Bickel B, Gross M. Detailed spatio-temporal reconstruction of eyelids. ACM Transactions on Graphics, 2015, 34(4): 44
99. Hu L W, Ma C Y, Luo L J, Li H. Robust hair capture using simulated examples. ACM Transactions on Graphics, 2014, 33(4): 126
100. Luo L, Li H, Rusinkiewicz S. Structure-aware hair capture. ACM Transactions on Graphics, 2013, 32(4): 76
101. Xiao J, Fang T, Zhao P, Lhuillier M, Quan L. Image-based street-side city modeling. ACM Transactions on Graphics, 2009, 28(5): 89–97
102. Nan L, Sharf A, Zhang H, Cohen-Or D, Chen B. SmartBoxes for interactive urban reconstruction. ACM Transactions on Graphics, 2010, 29(4): 157–166
103. Tan P, Zeng G, Wang J D, Kang S B, Quan L. Image-based tree modeling. ACM Transactions on Graphics, 2007, 26(3): 87
104. Quan L, Tan P, Zeng G, Yuan L, Wang J D, Kang S B. Image-based plant modeling. ACM Transactions on Graphics, 2006, 25(3): 599–604
105. Livny Y, Yan F, Olson M, Chen B, Zhang H, El-Sana J. Automatic reconstruction of tree skeletal structures from point clouds. ACM Transactions on Graphics, 2010, 29(6): 151
106. Blanz V, Mehl A, Vetter T, Seidel H. A statistical method for robust 3D surface reconstruction from sparse data. In: Proceedings of the IEEE International Symposium on 3D Data Processing, Visualization and Transmission. 2004, 293–300
107. Bao S Y, Savarese S. Semantic structure from motion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2011, 2025–2032
108. Bao S Y, Chandraker M, Lin Y, Savarese S. Dense object reconstruction with semantic priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013, 1264–1271
109. Dame A, Prisacariu V A, Ren C Y, Reid I. Dense reconstruction using 3D object shape priors. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013, 1288–1295
110. Hane C, Savinov N, Pollefeys M. Class specific 3D object shape priors using surface normals. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, 652–659
111. Kazhdan M, Bolitho M, Hoppe H. Poisson surface reconstruction. In: Proceedings of Eurographics Symposium on Geometry Processing. 2006, 61–70
112. Zhang Q S, Song X, Shao X W, Zhao H J, Shibasaki R. When 3D reconstruction meets ubiquitous RGB-D images. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, 700–707
113. Mehrdad V, Ebrahimnezhad H. 3D object retrieval based on histogram of local orientation using one-shot score support vector machine. Frontiers of Computer Science, 2015, 9(6): 990–1005
114. Hane C, Ladicky L, Pollefeys M. Direction matters: depth estimation with a surface normal classifier. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 381–389
115. Park M G, Yoon K J. Leveraging stereo matching with learning-based confidence measures. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 101–109
116. Guney F, Geiger A. Displets: resolving stereo ambiguities using object knowledge. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015, 4165–4175
117. Tanskanen P, Kolev K, Meier L, Camposeco F, Saurer O, Pollefeys M. Live metric 3D reconstruction on mobile phones. In: Proceedings of the IEEE International Conference on Computer Vision. 2013, 65–72
118. Kolev K, Tanskanen P, Speciale P, Pollefeyset M. Turning mobile phones into 3D scanners. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014, 3946–3953

Hao Zhu received the BS degree from Department of Electronic Science and Technology, Nanjing University (NJU), China in 2013. He is currently working toward the PhD degree of electronic science and technology at NJU. His interests include computer vision and computational imaging.

Yongming Nie received the BS degree from Department of Mathematics, Tianshui Normal University, China in 2009, and MS degree from the Department of Mathematics, Nanchang University, China in 2012. He is currently working toward the PhD degree of electronic science and technology in Nanjing University. His research interests mainly include computer vision and machine learning.

Tao Yue received the BS degree in automation from Northwestern Polytechnical University, China in 2009, and the PhD degree from Tsinghua University, China in 2015. He is currently an associate researcher with the School of Electronic Science and Engineering, Nanjing University, China. His research interests mainly include image processing and computational photography.



Xun Cao received his BS degree from Nanjing University (NJU), China in 2006, and PhD degree from the Department of Automation, Tsinghua University, China in 2012. He is currently a professor at the School of Electronic Science and Engineering, NJU. Dr. Cao was visiting Philips Research, Aachen, Germany during 2008, and Microsoft Research Asia, China during 2009 and 2010. He was a visiting scholar at the University of Texas at Austin, USA from 2010 to 2011. His research interests include computational photography, image based modeling and rendering, and 3D TV systems. He is the awardee of the NSFC Excellent Young Scholars Program in 2014.