Image-based 4-d Reconstruction Using 3-d Change Detection

Ali Osman Ulusoy and Joseph L. Mundy

School of Engineering, Brown University
Providence RI, USA

Abstract. This paper describes an approach to reconstruct the complete history of a 3-d scene over time from imagery. The proposed approach avoids rebuilding 3-d models of the scene at each time instant. Instead, the approach employs an initial 3-d model which is continuously updated with changes in the environment to form a full 4-d representation. This updating scheme is enabled by a novel algorithm that infers 3-d changes with respect to the model at one time step from images taken at a subsequent time step. This algorithm can effectively detect changes even when the illumination conditions between image collections are significantly different. The performance of the proposed framework is demonstrated on four challenging datasets in terms of 4-d modeling accuracy as well as quantitative evaluation of 3-d change detection.

1 Introduction

This paper proposes a new approach to reconstruct the evolution of a 3-d scene over time from imagery, *i.e.* image-based 4-d reconstruction. This work is targeted towards modeling of complex and ever-changing urban or natural environments. 4-d modeling of such scenes has various applications including urban growth analysis [1,2], construction site monitoring [3], natural resource management, surveillance and event analysis [4].

A naive approach to 4-d modeling is to reconstruct independent 3-d models for every time instant. This naive approach is motivated by the recent success of multi-view stereo (MVS) algorithms [5,6]. Nevertheless, 3-d reconstruction from images is an inherently ill-posed problem and reconstructing *dense* and *accurate* models of realistic urban or natural environments is still very challenging. Difficulties include areas of uniform appearance, transient objects, reflective surfaces, and self-occlusions. Model structure is highly dependent on illumination conditions, image resolution, and camera placement, which typically vary between data collections. Therefore, reconstructions of the same structures at different times typically contain many inconsistencies [4,7,8]. For instance, a building may be accurately reconstructed on an overcast day but not on a bright day with shadows and specular reflections that cause building surfaces to be fragmented in the model. These inconsistencies also complicate differentiating between actual changes in the scene from false alarms caused by inconsistent reconstructions, and therefore hinder 4-d analysis.

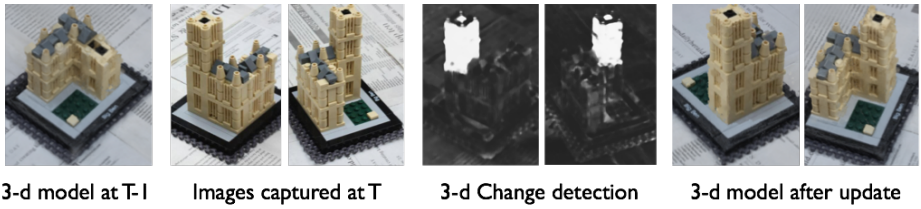


Fig. 1: Overview of the proposed 4-d modeling algorithm. The 3-d model at time T-1 is updated using the images captured at T. The algorithm first detects the 3-d changes in the scene by fusing change evidence from multiple views. Subsequently, surface geometry and appearance are estimated to update the evolving 4-d model.

Most importantly, this naive approach does not exploit the fact that in realistic scenes, many objects, such as buildings and roads, persist through long periods of time and need not be repeatedly estimated. Once a 3-d model of the scene is obtained, no further processing is necessary until a change occurs.

Motivated by these observations, this paper proposes a 4-d reconstruction algorithm based on change detection. The algorithm utilizes an initial accurate 3-d model of the scene. In subsequent frames, images from multiple (arbitrary) viewpoints are used to detect 3-d changes in the environment. This scheme circumvents reconstruction of the entire scene and can be achieved from only a sparse set of views. The image regions corresponding to unchanged parts of the scene can be safely discarded and model update is targeted only on the changed volumes. The steps of the proposed algorithm are depicted in Figure 1.

The algorithm continuously detects and incorporates changes in a scene over time to produce a complete 4-d representation. Such changes could include variations in both geometry and surface reflectance. However, if there exist vast illumination differences between data collections, the algorithm can effectively discard the effects of illumination and detect changes only in geometry.

The framework employs a probabilistic representation of 4-d surface geometry and appearance. Appearance is defined as the combination of surface reflectance and illumination giving rise to the pixel intensity or color seen in observations of the surface. This 4-d representation allows modeling the evolution of arbitrary and complex 3-d shapes, with no constraints on changing scene topology. The resulting 4-d models allow visualization of the full history of the scene from novel viewpoints as shown in Figure 2, as well as spatio-temporal analysis for applications such as tracking and event detection.

This paper makes three main contributions. The first contribution is a probabilistic formulation to detect 3-d changes at each time step without reconstructing a 3-d model for the entire scene. This inference process generates 3-d change volumes rather than pixel-level image change probabilities as in earlier applications of probabilistic volumetric representations [9]. The second contribution is an algorithm to discard the effects of varying illumination to detect structural 3-

d changes. The final contribution is the 4-d modeling pipeline which utilizes this change detection algorithm to continuously incorporate changes in a 4-d time-varying model. To the best of our knowledge, no other work has demonstrated such continuous operation in realistic and challenging urban scenes.

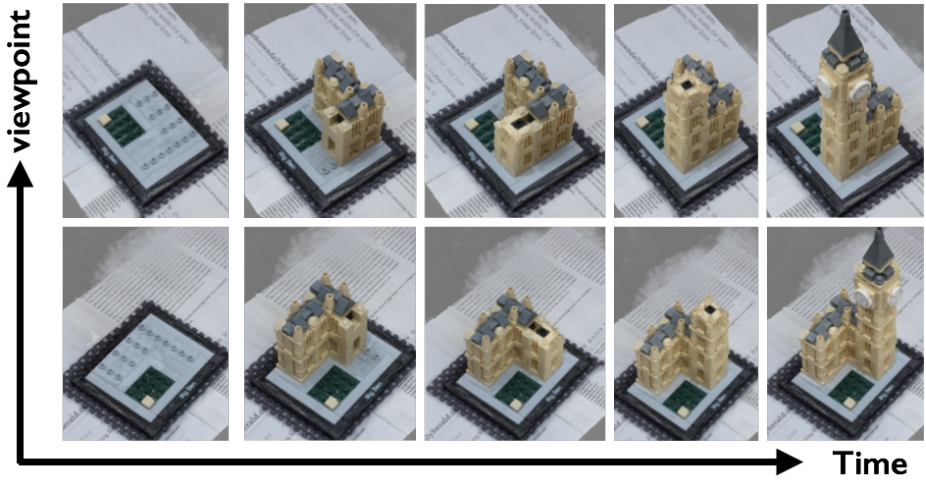


Fig. 2: Renderings of a 4-d model reconstructed using the proposed algorithm. The model captures the step-by-step construction of a toy Lego tower. The framework allows visualizing the 4-d model at any time instant from novel views.

2 Related Work

The proposed 4-d reconstruction pipeline performs 3-d change detection using the current set of images and the 3-d model from the previous time step. This 3-d change detection expands on the probabilistic and volumetric modeling (PVM) framework of Pollard and Mundy [9]. Their model encodes occupancy (surface) probability and an appearance distribution of each voxel in a volumetric grid, which is similar to other probabilistic voxel-based algorithms [10,11]. The surface and appearance distributions are jointly estimated using online Bayesian learning. The resulting 3-d model allows computing the probability of observing a given intensity or color, from an arbitrary (but known) viewpoint and without reference to training imagery. This probability can be used to compute the probability of change of each pixel in the image. Inspired by the PVM, the proposed framework encodes probabilistic surface and appearance information in 4-d.

The first contribution of this paper is a 3-d change detection algorithm based on combining pixel-wise change evidence from different viewpoints to compute the probability of change in each 3-d cell. Note that the algorithm of [9] detects changes *on the image domain*, whereas the proposed formulation estimates the volumetric structure of change *in the 3-d domain*.

The second contribution of this paper is an algorithm to detect changes only in 3-d geometry, *i.e.* structural change detection. Earlier algorithms that utilize the PVM for change detection assume either that illumination conditions vary slowly with respect to the model update rate or that illumination variations can be removed through global normalization [12]. However, in real outdoor settings, the variations in illumination and shadows are too complex to allow such normalization schemes. The proposed algorithm estimates a new appearance model for the PVM so as to match the appearance of the current image collection. Once the illumination conditions in the PVM and the corresponding appearance of surfaces as seen in the images are consistent, the 3-d change detection algorithm can be directly utilized to detect the structural differences.

The proposed approach is motivated by "appearance transfer"-based change detection methods [13] but at the same time, it extends them. Such methods exploit the prior 3-d scene geometry as a proxy to transfer pixels of one image onto the view of the other image and then compare the corresponding pixels. Matching pixels indicate where the geometry is still valid, *i.e.* background. Non-matching pixels indicate a foreground object. Kosecka exploit the appearance transfer concept for change detection in urban environments using StreetView images [14]. Deguchi *et al.* present a similar method where the uncertainty during depth estimation is also utilized [4]. Taneja *et al.* employs a surface mesh for appearance transfer between pairs of wide-baseline images [15,8]. Their approach also fuses the foreground maps in 3-d, using a similar approach to multi-view silhouette fusion [16]. The foreground estimate is then used as a coarse update to the initial mesh. Such methods compare pairs of images *on the image domain*. In contrast, the proposed method incorporates all image evidence simultaneously in 3-d using a probabilistic representation of surface geometry.

The third contribution of this paper is the 4-d modeling pipeline suitable for realistic urban or natural scenes. Another approach which utilizes the PVM in modeling 4-d scenes is by Ulusoy *et al.* [17], which generates 4-d models by reconstructing full 3-d models (PVMs) at each time step. Their results are limited to indoor studio data, which has static cameras as well as foreground segregation, that allows producing high-quality independent reconstructions. Such independent reconstruction is highly undesirable in realistic urban scenes for two reasons. First, such scenes contain many static or slowly varying structures which need not be repeatedly estimated at every frame. Second, 3-d reconstruction from images is an ill-posed problem and variations in imaging conditions typically result in significantly different reconstructions of the same structures. In contrast, the proposed pipeline estimates updates to the previous time step's PVM from subsequent imagery in two phases. In the first phase, the 3-d change detection algorithm is used to compute the probability of change in each 3-d cell. Then, reconstruction is targeted at cells with significant change probability so as to estimate only their changed surface and appearance distributions.

Ulusoy *et al.* define an efficient representation for 4-d volumetric scenes, which is also used in the algorithms proposed here. The representation,based on shallow octrees with binary time trees at each leaf is well-matched to highly

parallel GPU processing. It is emphasized that the change-driven 4-d modeling algorithms proposed here represent a significant departure from their approach of reconstructing a full 3-d model at each time step.

3 4-d Reconstruction Pipeline Overview

The proposed framework encodes a occupancy (surface) probability and an appearance distribution of each cell in 4-d. These quantities are stored in the 4-d data structure proposed by Ulusoy *et al.* [17]. In brief, a volume is decomposed as an octree and the temporal variation in each cell is modeled by a binary tree. This 4-d representation is initialized with a PVM (3-d model) that represents the initial state of the scene. At subsequent times, imagery taken roughly simultaneously from arbitrary viewpoints are input to the 3-d change detection algorithm. The images are registered to the 3-d model using standard tools. The details of registration can be found in the experiments section. The proposed change detection algorithm and its extension to varying illumination conditions are explained in Sections 4 and 5 respectively. These algorithms compute the probability of change in each 3-d cell. Subsequently, the new surface and appearance distributions of cells with significant change probability estimated from the current set of images to update the current 3-d model. Namely, the binary time trees of the changed cells in the 4-d representation are sub-divided to allocate new memory for the update. The cells to be updated are initialized as empty (low surface probability) and the online learning algorithm of [9] is used to estimate the surface probabilities and appearance distributions. It should be noted that this process does not estimate the changed parts of the model in isolation, but uses the entire PVM for visibility and occlusion reasoning. This 3-d context supports change analysis in cases where objects are heavily occluded by background surfaces. These steps of 3-d change detection and model update are repeated for each time instant to achieve a complete 4-d model.

4 3-d Change Detection

This section describes 3-d change detection in the PVM from subsequent images captured from multiple viewpoints. The simplest situation is where illumination conditions are assumed to vary slowly such that the appearance encoded in the PVM at frame $T-1$ is similar to that of the images captured at T. Note that this assumption does not require the appearance of the initial PVM and all subsequent frames to be similar; it only requires the appearance of neighboring times to be similar. The assumption of slowly varying illumination will later be relaxed but enables a straightforward description of the algorithm. Based on this assumption, events that vary the appearance of the scene qualify as change. Such events include structural changes, *e.g.* a new building, as well as changes in surface colors, *e.g.* new paint on a wall.

The PVM is represented in a grid of cells, where each cell X contains a surface probability, $p(X_S)$, and an appearance distribution, $p(X_I)$. Appearance

is defined as the intensity or color of surface voxels as seen in the image pixels. The problem is to compute the presence or absence of change in each cell given subsequent images $\{I_i\}_{i=1}^{N}$. A binary random variable, X_C is defined for each cell, where $X_C = 1$ denotes the presence of change and $X_C = 0$ denotes no change. The problem can be posed as maximum a-posteriori estimation,

$$\{X_C\}^* = \underset{\{X_C\}}{\arg\max}\, p(\{X_C\}|\{I\}) \propto p(\{I\}|\{X_C\})\, p(\{X_C\}) \tag{1}$$

The likelihood term, $p(\{I\}|\{X_C\})$, models evidence from the images. The prior term, $p(\{X_C\})$, represents all previous knowledge regarding which parts of scene may contain changes. Such information can be obtained from various sources including semantic cues, *e.g.* a construction site is likely to contain many changes, whereas the Eiffel tower is almost surely static. However in the work here, a uniform prior is assumed, *i.e.* each voxel is equally likely to change.

For tractability, the change probability of each cell is considered conditionally independent given the images. This assumption leads to a simple solution to eq. (1) where a cell is labeled as containing change if,

$$\frac{p(\{I\}|X_C = 1)}{p(\{I\}|X_C = 0)} > \frac{p(X_C = 0)}{p(X_C = 1)} \tag{2}$$

Note that this solution is the likelihood ratio test arising from Bayesian decision theory. The terms $p(\{I\}|X_C = 1)$ and $p(\{I\}|X_C = 0)$ model image evidence in the presence and absence of change respectively. Assuming conditional independence, the terms are simplified as follows,

$$p(\{I\}|X_C) = \prod_{i=1}^{N} p(I_i|X_C) \tag{3}$$

The image intensities (or color) observed at a pixel are either due to background (PVM) or foreground (change). In general, no a-priori information is available about the appearance of a "change" and therefore, the intensities are assumed to be uniformly distributed. On the other hand, if the pixel is generated by the background, the PVM allows computing the probability of observing intensity I along ray R as follows:

$$p(I) = \sum_{X \in R} p(I|V = X)\, p(V = X) \tag{4}$$

$$p(V = X) = p(X_S)\, p(X \text{ is visible}) \tag{5}$$

$$p(X \text{ is visible}) = \prod_{X' < X} (1 - p(X'_S)) \tag{6}$$

where V denotes the cell along the ray that is responsible for producing the pixel intensity and $p(I|V = X)$ is the appearance distribution of the cell.

Assuming the cell X back-projects into a single pixel in image I_i, predicates B_X^i and F_X^i can be defined for each pixel, indicating whether the pixel was

generated by background or change respectively. The term $p(I_i|X_C = 1)$ can be expressed using these predicates as follows,

$$p(I_i|X_C = 1) = p(I_i|F_X^i)\, p(F_X^i|X_C = 1) + p(I_i|B_X^i)\, p(B_X^i|X_C = 1) \quad (7)$$

where $p(I_i|F_X^i)$ is the uniform distribution and $p(I_i|B_X^i)$ is computed as in eq. (4). The term $p(B_X^i|X_C = 1)$ is the probability of observing background in the pixel when cell X along the ray contains change. It is expanded by applying the Bayes rule,

$$p(B_X^i|X_C) = \frac{p(X_C|B_X^i)\, p(B_X^i)}{p(X_C|B_X^i)\, p(B_X^i) + p(X_C|F_X^i)\, p(F_X^i)} \quad (8)$$

where $p(F_X^i)$ and $p(B_X^i)$ can be set to 0.5 in the absence of prior knowledge. The likelihood term $p(X_C = 1|B_X^i)$ can be expressed based on the following intuition. Given the pixel is background, the voxels between the camera and the surface voxel, *i.e.* visible voxels, can be labeled as containing no change. However, nothing can be said about the voxels after the surface voxel along the ray, *i.e.* occluded voxels. Thus, the term can be expanded as follows,

$$p(X_C|B_X^i) = p(X_C|vis, B_X^i)\, p(vis|B_X^i) + p(X_C|\neg vis, B_X^i)\, p(\neg vis|B_X^i) \quad (9)$$

where $vis|B_X^i$ is shorthand for "X is visible" given the pixel is generated by background. Note that $p(X_C = 1|vis, B_X^i) = 0$, since a visible voxel that back-projects into a background pixel cannot contain change. However, when the voxel is not visible, the term $p(X_C = 1|\neg vis, B_X^i)$ is set to 0.5 to denote the uncertainty. The probability of visibility, $p(vis|B_X^i)$, can be computed exactly as in eq. (6). However, note that this computation is only possible because the pixel is known to be background, *i.e.* the PVM geometry along the ray is still accurate. In general, visibility or other measurements in the PVM may no longer be correct due to the changes in the scene. The likelihood is simplified as,

$$p(X_C|B_X^i) = 0.5\, p(X \text{ is not visible}) \quad (10)$$

Finally, $p(X_C = 1|F_X^i)$ is set to 1. Since the pixel is foreground, the geometry in the PVM is no longer accurate and therefore, nothing can be said regarding visibility of change. In the absence of such information, it is best to assume change may be present in any cell along the ray.

Overall, this formulation incorporates change evidence from all viewpoints while reasoning about visibility and occlusions. The solution, eq. (2), allows processing each image separately and then making a decision locally at each 3-d cell. Note that this framework can be augmented with spatial priors. However, such priors typically require large scale global inference algorithms. Moreover, traditional priors such as smoothness or planarity are often not valid in complex urban settings with changes due to construction or transportation.

5 Structural 3-d Change Detection

The change detection algorithm described in the previous section utilizes 3-d appearance in the PVM to detect 3-d changes in the model. When the illumination conditions of a new image collection are significantly different from that of an existing PVM, this algorithm can no longer be used. In such cases, the entire scene might have changed in appearance and therefore, changes only in geometry are considered. This section describes a method to effectively discard the effects of illumination change between collections to detect changes only in geometry.

In such cases, the appearance distributions stored in the PVM do not directly relate to the new appearance of the scene. Therefore, the proposed algorithm discards these distributions and retains only the probabilistic 3-d geometry. Then, the algorithm estimates the current appearance probability distributions in the PVM from the input images. Once the illumination conditions in the PVM match to that of the images, the change detection algorithm of Section 4 can be applied to detect the changes in geometry.

First, assume that the 3-d geometry represented by the PVM is unchanged. Under this assumption, the appearance of a surface voxel can be estimated from the images it is visible in. Each image provides an intensity or color observation, as well as a weight (see eq. 5) for each voxel in its field of view. An appearance distribution is fit to the set of weighted samples. The weight prevents occluded voxels from being trained with occluding surfaces' color. Under the Lambertian reflectance assumption, a Gaussian distribution is sufficient to explain the intensities observed from different viewpoints.

In general, the scene will contain geometric changes (additions or removals) that are currently not modeled in the PVM. The changes will cause the appearance models of some voxels to be trained with erroneous observations. This situation can be analyzed in Figure 3, where the addition and removal of a blue object are shown in Figures 3a and 3b respectively. In the case of addition, the appearance model of the occluded voxel X_1 is incorrectly trained with the object color. When the object is removed voxel X_2, which no longer contains a surface, is trained with colors from different parts of the surface behind. Also, X_1 is no longer occluded in the middle camera but the pixel observation is still weighted with an incorrect low visibility probability.

The challenge of appearance estimation is then to segregate in each voxel, the actual color of the surface from the pixel observations caused by changes. Note that this problem is very similar to background modeling in the image domain. Inspired by Stauffer and Grimson's approach [18], the appearance of each voxel is modeled using a mixture of Gaussians distribution. The training follows the online EM based algorithm of [18]. The "background" mixture modes, which correspond to the actual surface color, are chosen to be those with high mixture weight and low variance. The rest of the modes, caused by changes in the scene, can be discarded for the purposes of this algorithm.

This scheme relies on being able to classify the actual color modes of surfaces, which depends on the scene and camera poses. In general, multiple observations

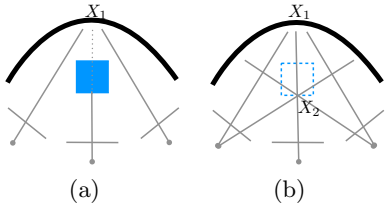


Fig. 3: A simple scene to study the effects of appearance estimation when new object (blue square) is added (a) or removed (b) from the scene.

of the surface are needed to form "background" modes such that any other mode created by the changes can be classified as otherwise. Consider voxel X_1 in Figure 3a, which receives one incorrect and two correct observations, resulting in a low variance mode for the actual color and a higher variance mode for the occluding object color. Hence, the appearance of X_1 can be correctly computed.

A demonstration of appearance estimation is provided in Figure 4 on a street scene. A rendering of the initial PVM is presented in Figure 4a. Subsequent images (see Figure 4b) were captured under significantly different illumination and reveal the addition of a speed camera next to the tree. Note that this object is not present in the initial PVM. Figure 4c displays the PVM after having discarded the old appearance models and using a single image to estimate the new appearance. The image pixels corresponding to the speed camera get projected onto the wall since the speed camera's 3-d geometry is not yet modeled in the PVM. Figures 4d and 4e display renderings of the PVM after using two and eight images respectively. As more images are used, the ghosting effects of the speed camera fade away and the correct (current) color of the wall can be estimated. Once the new appearance distributions are estimated, the change detection algorithm of Section 4 can applied directly to detect changes in geometry, such as the shape of the speed camera.

6 Experiments

The framework is evaluated on four different datasets. Each dataset consists of imagery captured from multiple arbitrary viewpoints at each time instant. All datasets except the last one contain multiple time steps to demonstrate continuous operation of the 4-d reconstruction framework.

The images for the first three datasets were captured using a handheld camera with 1 megapixel resolution. The images were calibrated with incremental Structure-from-Motion using the VisualSFM software [19]. The initial set of images were calibrated using the standard SfM pipeline. The resulting image matches and point cloud are used to register subsequent images. In all three datasets, this procedure achieved sub-pixel reprojection error. These datasets are available at `http://www.lems.brown.edu/~au/4d_datasets`.

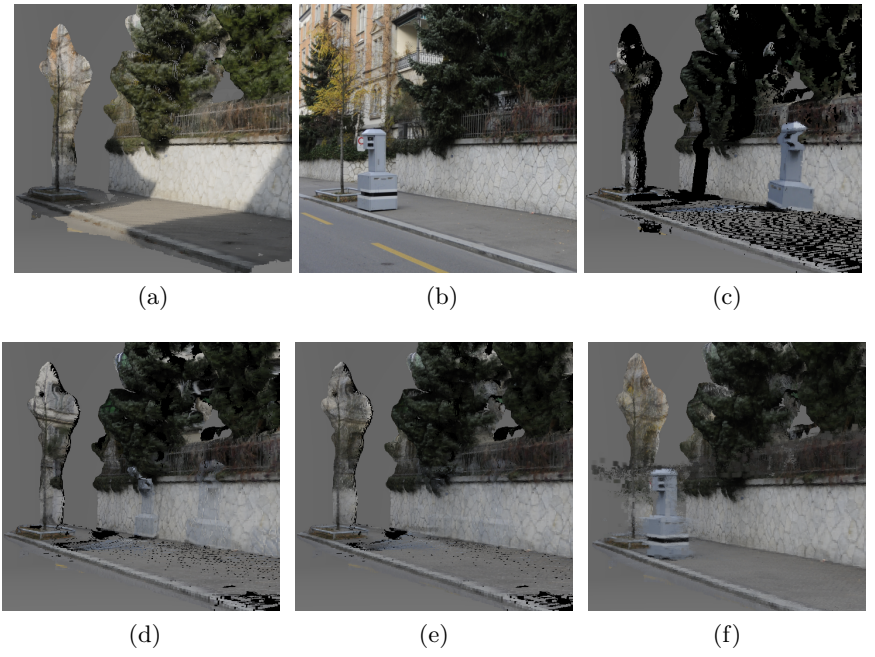
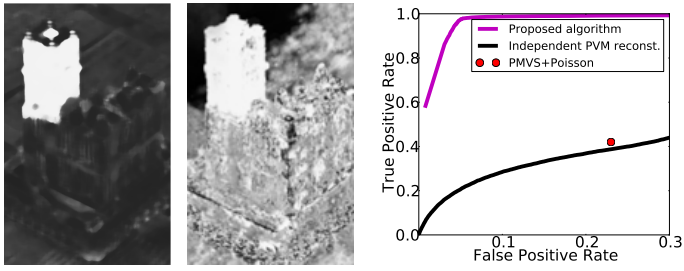


Fig. 4: Demonstration of estimating the new appearance of the PVM from subsequent images. (a) The rendering of the initial PVM. (b) One of eight subsequent images captured under different illumination conditions than that of the PVM. Note the additional object (speed camera) next to the tree. (c-e) The rendering of the PVM after having discarded the old appearance model and estimated the new appearance with a single image (c), two images (d) and eight images (e). (f) Rendering of the PVM after 3-d model update.

For the first two datasets, the illumination conditions vary slowly with respect to the time step such that the appearance of the scene does not change much between collections. Both datasets contain small-scale objects so that changes can be easily produced, but were captured *outdoors under natural illumination*. The 3-d change detection algorithm of Section 4 is employed here to reconstruct the 4-d model. The final two datasets contain image sets captured under significantly different illumination in real ground-level urban scenes and are used to evaluate the framework with the structural change detection algorithm. The choice of which change detection algorithm to use is decided manually for now. The last dataset, which consists of two time steps, was taken from Taneja *et al.* [8] where they used it to demonstrate their structural change detection approach. This dataset is used to compare the proposed algorithm to this previous work.

The first dataset captures the step by step construction of a Lego Big Ben tower. The scene initially contains the base plate and the tower is erected in a total of six time steps. Each time step is modeled from roughly forty images

taken around the object in two circles. Renderings from the resulting 4-d model is displayed in Figure 2. The proposed algorithm effectively captures the evolution of the scene with high resolution.



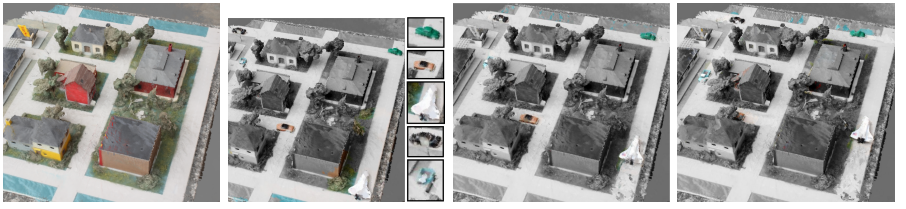
(a) (b) (c)

Fig. 5: Comparison of 3-d change detection results for the first dataset. Volume rendering of changes predicted by the proposed approach (a) and the approach of independently reconstructing PVMs (b). The ROC curve is displayed in (c).

Example 3-d change detection and model update results are displayed in Figure 1. It can be observed that the detections are correctly localized on the two new blocks added to the tower. The change detection results are compared to the approach of reconstructing 3-d models for each time step in Figure 5. Separate PVM models are reconstructed for $T-1$ and T and the models are compared in 3-d, where the change probability of a cell is computed as the difference between the corresponding surface probabilities at $T-1$ and T. The resulting change probabilities are displayed using volume rendering in Figure 5b, where the inconsistencies between the models can be observed. The comparison is quantified using an ROC curve shown in Figure 5c. The ground truth 3-d changes are delineated manually and the detection threshold in eq. (2) is varied to construct the curves. The significant inconsistency between individually reconstructed PVM models results in a high number of false positives. This comparison was also repeated using PMVS[20]+Poisson[21] reconstruction instead of the PVM approach. The resulting meshes were voxelized for 3-d binary change detection, which yields a single point on the ROC curve. The PMVS+Poisson result is similar to that of the PVM. These results demonstrate the inherent ill-posed nature of inferring 3-d geometry from images. Small imaging differences (slight variations in illumination and camera poses) produce larger 3-d model surface variations which lead to dissimilarities in the reconstructions of unchanged regions. These dissimilarities complicate distinguishing between actual changes in the scene from false alarms. In contrast, the proposed algorithm is robust to these imaging differences due to its probabilistic multi-view change reasoning.

This dataset was also used to assess the computation time of the proposed algorithm with respect to reconstructing 3-d models from scratch. Both methods

were implemented based on the code made available in [17], which provides an implementation of the 4-d representation optimized for computation in the GPU. The change detection algorithms and the model update algorithm were implemented to exploit this parallelism. All algorithms process images in an online fashion. On average, reconstructing 3-d models from scratch takes roughly 7 minutes while the proposed pipeline takes 5 minutes. The computation time is improved because during model update, only the changed pixels are used to cast rays to update the volume. Moreover, since many structures in the scene are already reconstructed, the update algorithm converges much faster.



(a) (b) (c) (d)

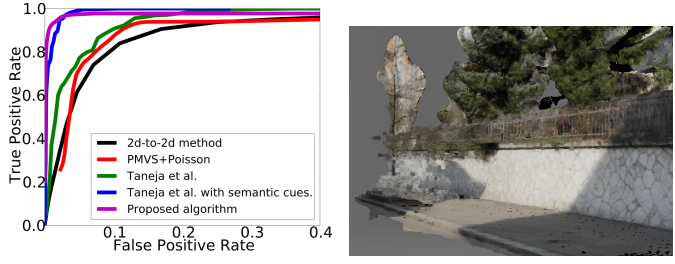
Fig. 6: Results from the model town dataset. (a) The initial time step. (b-d) Renderings from frames 1, 3 and 5 respectively. The colors of pixels that do not observe a change are converted to grey to highlight the dynamic objects. The column to the right of (b) shows a close up view of the changes for that time step. Please zoom in the images to appreciate the reconstruction quality.

The second dataset contains toy cars moving on a lifelike miniature model town. The initial time step consists of only the model town and is modeled from roughly fifty images. The initial PVM is presented in Figure 6a. For each of the rest of the time steps, only ten images are available. In the second time step, the cars are placed on the model town. Then, in each subsequent frame, the cars are moved gradually along the road. This dataset poses additional challenges due to the limited number of views at each time step. Moreover, the scene contains many complex structures, such as the buildings and trees, that frequently occlude the cars. The resulting 4-d model is displayed from a novel viewpoint for three time steps in Figures 6b-6d. The detected 3-d changes are used to grey out the pixels that do not observe a change. It can be seen that the moving cars are accurately detected and reconstructed.

Static structures in the model town, that were reconstructed accurately in the initial time step, are detected as unchanging, which allows their reconstructions to persist from the initial time step to future time steps. In contrast, the method of reconstructing individual models at each time step attempts to reconstruct the *entire scene structure* from the sparse set of images. Experiments show both PVM and PMVS+Poisson fail to produce acceptable 3-d models in this setting. Please see the supplementary document for the comparisons.

Fig. 7: Renderings of the 4-d model of the third dataset.

The third dataset was captured in an urban environment. The initial five time steps involve the process of taking two trash cans outside and then removing them, one by one. The first time step was modeled from fifteen images and the next four time steps from six images on average. The resulting 4-d model is rendered in Figure 7, where the appearance and disappearance of the trash cans are modeled accurately ($T2 - T4$). Next, the scene was imaged during a construction several months later ($T = 5$), where no major structural changes are present. The structural change detection algorithm correctly detects no significant changes in the scene even though the illumination has changed drastically. The rendering of the scene after the new appearance can be seen in the image labeled $T = 5$. Note the presence of direct sunlight and shadows. Finally, the next day ($T = 6$), images were taken under new illumination conditions, when a construction cone was placed on the sidewalk. The geometry of the cone is modeled accurately as seen in the figure. Similarly to the previous dataset, the small number of images in some time steps does not allow independent 3-d reconstruction of each time step.



(a) (b)

Fig. 8: (a) Change detection ROC curve for the fourth dataset. (b) The rendering of the PVM after the reverse experiment of removing the object.

The final dataset consist of two sets of eight images of an urban environment with (see Figure 4b) and without a speed camera, as well as a mesh model of the scene without it. Taneja *et al*. exploit this mesh model as a proxy to compare pairs of images and integrate image inconsistencies in the volume [8]. Their approach assumes spatial smoothness of changed regions and exploits the output of object (vegetation, human, and car) detectors to suppress false positives. In contrast, the proposed approach employs all image evidence (not just image pairs) simultaneously in an entirely probabilistic representation of surface geometry, appearance and change. Moreover, the proposed approach does not assume a smoothness prior or assume such semantic cues are available. The comparison is carried out using the ground truth change masks provided in the dataset. The 3-d changes are projected onto the ground truth masks to compute true and false positive rates. The resulting ROC curve is displayed in Figure 8a, where rest of the curves were taken from [8]. The proposed approach performs better than Taneja *et al*.'s approach (green curve), which uses a spatial smoothness prior to regularize the detections, as well as when their approach additionally exploits semantic cues (blue curve). However, it can be observed that the proposed approach does not reach a full detection rate. The missed detections occur at the bottom of the speed camera, where the camera (changed region) borders the pavement (original geometry). Taneja *et al*.'s approach is able to detect this region due to the spatial smoothness term in the formulation.

Figure 4f displays the updated model which includes the speed camera. The reverse experiment was also performed, where this updated model was used together with the images excluding the object to detect the disappearance of the object. The resulting model is presented in Figure 8b.

7 Conclusion

This paper presented a novel image-based 4-d reconstruction algorithm suited for capturing the evolution of ever-changing urban or natural environments. The proposed algorithm avoids rebuilding 3-d models of the scene at each time instant through detection of changes and reconstructing only the changed surfaces. Through accurate change detection and model update, the algorithm is able to achieve continuous operation in challenging datasets. Experiments confirm that this approach yields higher quality 4-d models, better change detection accuracy and running time compared to the baseline approach of independent 3-d reconstruction of each time step. The reconstruction algorithm is based on a novel 3-d change detection algorithm which can detect changes in geometry and appearance, or only in geometry in case of significant illumination difference between collections. The algorithm performs favorably compared to previous works.

Future work includes city-wide large scale 4-d reconstruction, where various sources of information can be exploited, *e.g*. aerial, ground level and satellite imagery, cadastral 3-d models [1], as well as LIDAR. Community photo collections also provide a great source for 4-d modeling but bring many additional challenges yet to be addressed [22].

References

1. Taneja, A., Ballan, L., Pollefeys, M.: City-Scale Change Detection in Cadastral 3D Models using Images. CVPR (2013)
2. Schindler, G., Dellaert, F.: Probabilistic temporal inference on reconstructed 3D scenes. CVPR (June 2010)
3. Golparvar-Fard, M. and Pena-Mora, F. and Savarese, S.: Monitoring changes of 3D building elements from unordered photo collections. Proc., IEEE workshop on Computer Vision for Remote Sensing of the Environment (in conjunction with ICCV-11) (2011)
4. Sakurada, K., Takayuki, O., Deguchi, K.: Detecting Changes in 3D Structure of a Scene from Multi-view Images Captured by a Vehicle-mounted Camera. CVPR (2013)
5. Pons, J.P., Labatut, P., Vu, H.H., Keriven, R.: High Accuracy and Visibility-Consistent Dense Multiview Stereo. PAMI (2012)
6. Tola, E., Strecha, C., Fua, P.: Efficient large-scale multi-view stereo for ultra high-resolution image sets. Machine Vision and Applications **27**(5) (2011) 1–18
7. Restrepo, I.M.: Characterization of Probabilistic Volumetric Models for 3-d Computer Vision. PhD thesis, Brown University (2013)
8. Taneja, A., Ballan, L., Pollefeys, M.: Image based detection of geometric changes in urban environments. ICCV (November 2011)
9. Pollard, T., Mundy, J.L.: Change Detection in a 3-d World. In: CVPR. (June 2007)
10. Bonet, J.D., Viola, P.: Poxels: Probabilistic voxelized volume reconstruction. ICCV (1999)
11. Broadhurst, A., Drummond, T.W., Cipolla, R.: A Probabilistic Framework for Space Carving. ICCV (2001)
12. Pollard, T.B.: Comprehensive 3-d Change Detection Using Volumetric Appearance Modeling. PhD thesis, Brown University (2009)
13. Ivanov, Y., Bobick, A., Liu, J.: Fast lighting independent background subtraction. IJCV (2000) 1–14
14. Košecka, J.: Detecting changes in images of street scenes. ACCV (2012)
15. Taneja, A., Ballan, L., Pollefeys, M.: Modeling dynamic scenes recorded with freely moving cameras. ACCV (2011)
16. Franco, J.S., Boyer, E.: Fusion of multiview silhouette cues using a space occupancy grid. ICCV (2005)
17. Ulusoy, A.O., Biris, O., Mundy, J.L.: Dynamic Probabilistic Volumetric Models. In: ICCV. (2013)
18. Stauffer, C., Grimson, W.: Adaptive background mixture models for real-time tracking. In: CVPR. (1998)
19. Wu, C.: Towards Linear-time Incremental Structure from Motion. 3DV (June 2013)
20. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. PAMI (August 2010)
21. Kazhdan, M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. Eurographics (2006)
22. Snavely, N., Seitz, S.M., Szeliski, R.: Modeling the World from Internet Photo Collections. IJCV (December 2007)