Generalized Connectivity Constraints for Spatio-temporal 3D Reconstruction

Martin Ralf Oswald, Jan Stühmer, and Daniel Cremers

Department of Computer Science, Technische Universität München[*]
Boltzmannstr. 3, 85748 Garching, Germany

Abstract. This paper introduces connectivity preserving constraints into spatio-temporal multi-view reconstruction. We efficiently model connectivity constraints by precomputing a geodesic shortest path tree on the occupancy likelihood. Connectivity of the final occupancy labeling is ensured with a set of linear constraints on the labeling function. In order to generalize the connectivity constraints from objects with genus 0 to an arbitrary genus, we detect loops by analyzing the visual hull of the scene. A modification of the constraints ensures connectivity in the presence of loops. The proposed efficient implementation adds little runtime and memory overhead to the reconstruction method. Several experiments show significant improvement over state-of-the-art methods and validate the practical use of this approach in scenes with fine structured details.

Keywords: connectivity constraints, spatio-temporal 3D reconstruction.

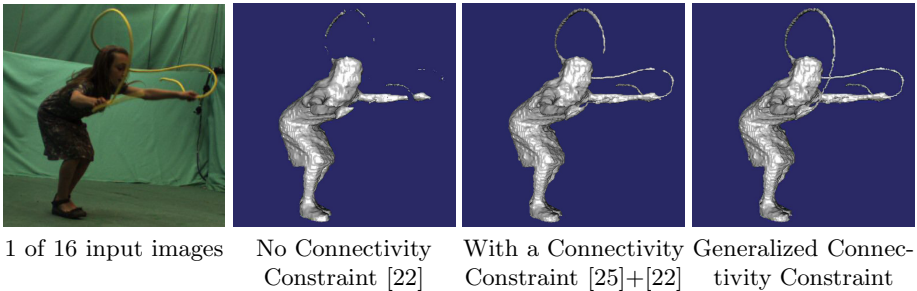| 1 of 16 input images | No Connectivity Constraint [22] | With a Connectivity Constraint [25]+[22] | Generalized Connectivity Constraint |

Fig. 1. Embedding connectivity constraints into multi-view reconstruction clearly helps to recover fine structures like the rope. The tree-shaped connectivity prior [25] only works for objects without holes (genus 0), resulting in disconnected parts when the rope touches the head. The proposed generalized connectivity constraint works for objects with arbitrary genus. Dataset: 'jumping rope' sequence from the INRIA 4D repository [16].

1   Introduction

Multi-view 3D reconstruction is a central research topic in computer vision that is driven in many different directions. Apart from a realistic physical modeling of the inverted imaging process, it is also of common interest to model learned and prior information (e.g. smoothness or shape priors), or imposing intuitive constraints on the solution, such as symmetry, connectedness or surface genus. In this work, we propose a method that is first: able to enforce connectedness of the computed solution, and second: able to preserve holes of the reconstructed scene within a multi-view reconstruction setup. We can guarantee that the solutions' surface genus is not smaller than the one of the visual hull.

Our approach is motivated by the spatio-temporal multi-view 3D reconstruction of scenes containing small object structures that we want to preserve in the reconstruction. Although fine object structures can also be preserved by incorporating exact silhouette information, such as in the work of Cremers and Kolev [6], this method is not applicable if the precomputed silhouettes are not accurate.

1.1   Contributions

- We embed the concept of connectivity constraints for image segmentation into a spatio-temporal multi-view reconstruction setup.
- Since the connectivity constraints proposed in [25] only work well for scenes and objects of genus zero, we propose a generalization of the connectivity constraints to an arbitrary genus.
- We suggest an efficient implementation of the generalized connectivity constraints with a small additional memory footprint and an almost unchanged computation runtime per optimization iteration. The necessary preprocessing only adds around one minute to the three minutes computation time per frame for the presented experiments.

1.2   Related Work

Spatio-temporal multi-view reconstruction on dense occupancy grids has been pioneered by Goldlücke et al. [12], [11] with a level set representation of the space-time surface. The drawback of this approach is its dependency on a good initialization due to the locally optimal optimization procedure. Aganj et al. [1] compute a spatio-temporal Delaunay mesh that automatically provides temporal correspondences of mesh vertices by using silhouettes. Starck and Hilton [24] proposed a spatio-temporal reconstruction pipeline which first estimates shapes from silhouettes and later refines the reconstruction with photometrically matched features and information about the reconstruction result from the previous time step. In [13], Guillemaut and Hilton propose a method that concurrently estimates a multi-layer segmentation and corresponding depth values of the scene based on confidence-weighted optical flow measures.

We use the spatio-temporal reconstruction method from Oswald and Cremers [22] as the basis of our work. This method is a generalization of the 3D reconstruction by Kolev et al. [19] to the temporal domain. Both approaches use a volumetric representation of the surface within an energy minimization framework which makes it easy to impose additional constraints on the solution.

To the best of our knowledge the only previous work on connectivity in 3D reconstruction is the work of Bleyer et al. [3], in which the authors propose to use connectivity information for joint stereo matching and object segmentation. In contrast to our work, this method is rather a 2.5D than a 3D or even a 4D reconstruction method. While the authors in [3] correctly define connectivity as the existence of a connecting path, they instead propose to determine the connectivity of a pair of points by testing along a straight line that connects both points, thus only favoring convexity of objects.

In the field of image segmentation, topology preserving extensions have been proposed in different algorithmic frameworks. For the graph cut [4] algorithm, Zeng et al. [27] proposed a topology preserving refinement scheme. Chen et al. [5] propose to alternatingly estimate a graph cut segmentation and alter the respective unaries based on a level-set representation in order to fulfill predefined topological constraints. In contrast to our approach, this method does not compute minimal geodesic connections with respect to the input data and its runtime is much higher due to the iterative optimization. For the level set method a topology preserving extension was proposed by Han et al. [15]. Vicente et al. [26] use connectivity priors for a Markov random field segmentation. The authors propose an approximation scheme to enforce connectivity of the segmented object with respect to user given seed points. The drawback of all methods on connectivity mentioned so far is that they only converge to a local minimum and therefore depend on the initialization. Moreover, apart from Bleyer et al. [3] all approaches are made for a 2D domain.

Recently, three different globally optimal approaches were proposed. One is the work of Nowozin and Lampert [21], in which the constrained image segmentation problem is formulated as a linear programming relaxation. The drawback of this method is that the complexity does not scale well with the image size and therefore prevents its use for 3D or 4D reconstruction methods where the problem size easily reaches thousands or even millions of variables.

A closely related work is that of Gulshan et al. [14]. The foreground segment is restricted to the shape of a geodesic star with respect to a geodesic distance measure that depends on the image gradient. By placing several input seeds, this constraint allows several geodesic star shaped objects, their union is called a geodesic forest. However, the authors only present results on 2D image data and because the method is formulated in a graph-cut segmentation framework the boundary length regularizer is affected by the discretization.

Another globally optimal segmentation method with connectivity constraints is the work of Stühmer et al. [25]. The authors propose a geodesic tree-shaped connectivity prior for image segmentation in an efficient convex optimization framework that allows the segmentation of large scale problems as they arise

for example in 3D medical imaging data. In contrast to [14], this method is formulated using a continuous segmentation framework and does not suffer from discretization artifacts with respect to the boundary length regularizer. It is perfectly suited to accurately segment objects with a fine detailed tree-like structure, such as blood vessels in angiography, or the legs of insects in photographs. They first compute a single-source geodesic shortest path tree based on the image data. Then, the tree-connected segmentation is computed by imposing linear constraints on the solution, based on the precomputed shortest path tree. As such, these constraints only impose connectivity for objects without any holes or loops (genus 0).

   We follow this idea in the context of spatio-temporal multi-view reconstruction and generalize the connectivity constraint to objects with arbitrary genus.

2   3D Reconstruction with Connectivity Constraints

First, we give an introduction to the spatio-temporal 3D reconstruction approach of [22] which allows 3D reconstructions of moving scenes by using video data from several viewpoints. Then we show how the connectivity constraint of [25] can be incorporated into the reconstruction. The combination of both methods allows image based globally optimal 3D reconstruction while preserving connectivity of the object. As shown later in the experiments, this constraint also helps to reconstruct fine scale details of the scene.

2.1   Spatio-temporal Multi-view Reconstruction

This section briefly repeats the spatio-temporal 3D reconstruction approach in [22], which forms the basis of the proposed approach. The temporally changing scene is represented by an hypersurface $\Sigma \subset V \times T$ that is embedded in the spatio-temporal product space of the three-dimensional space $V \subset \mathbb{R}^3$ that changes over time $T \subset \mathbb{R}_{\geq 0}$. At every time instant t, the scene is observed by N static cameras with known projection matrices $\{\pi_i\}_{i=1}^N$ and approximate silhouettes $\{S_i(t)\}_{i=1}^N$. We do not need exact silhouettes, which is a desirable property in a 4D setup, because it is not easy to automatically generate exact silhouettes for all cameras and all time steps. The silhouettes contain valuable information about the number of holes in the scene, that is, the genus of the scene. Bringing this information into space-time, we will later use the visual hull $\mathcal{VH}(t) = \bigcap_{i=1}^N \pi_i^{-1}(S_i(t))$ to analyze the scene structure and to impose constraints on the connectedness of the reconstructed surface.

   For mathematical convenience with respect to the final optimization procedure, we represent the hypersurface Σ as the boundary between the interior and the exterior part of the scene. Thus, hypersurface Σ is expressed by the binary labeling function $u : V \times T \mapsto \{0, 1\}$, indicating either interior or exterior for every point in space-time. This automatically ensures a closed manifold without boundaries and easily deals with arbitrary topologies. Stated as an energy minimization, the 3D reconstruction problem is described as finding a surface with

minimal area that best fits the input data, represented by a photoconsistency measure $\rho : V{\times}T \mapsto \mathbb{R}_{\geq 0}$ and a data term $f : V{\times}T \mapsto \mathbb{R}$:

$$E(u) = \int\limits_{V{\times}T} \left(\rho|\nabla_{\boldsymbol{x}}u| + g_t|\nabla_t u| \right) \ d\boldsymbol{x}dt + \lambda \int\limits_{V{\times}T} fu \ d\boldsymbol{x}dt \tag{1}$$

with $\lambda > 0$ steering the smoothness of reconstructed hypersurface. The data term $f : V{\times}T \mapsto \mathbb{R}$ locally expresses an affinity to an interior ($f < 0$) or an exterior ($f > 0$) labeling. Similar to [22], we restrict the solution space of the energy minimization in (1) to the visual hull. As a consequence, the approximate silhouettes can be 'larger', but not 'smaller' in order to ensure that the corresponding visual hull fully contains the true scene. The function $g_t(\boldsymbol{x}, t) = \exp\left(-|\nabla f(\boldsymbol{x}, t)| \right)$ weights the temporal smoothing based on f to account for fast motions.

The photoconsistency measure $\rho(\boldsymbol{x})$ resembles truncated normalized cross-correlation matching scores C_i between neighboring camera pairs and is defined as

$$\rho(\boldsymbol{x}) = \exp\left[-\mu \sum_{i \in \mathcal{C}} \underbrace{\delta\left(d_i^{\max} {=} \mathrm{depth}_i(\boldsymbol{x}) \right) \cdot C_i(\boldsymbol{x}, d_i^{\max})}_{\mathrm{VOTE}_i(\boldsymbol{x})} \right] \ . \tag{2}$$

The delta function δ in combination with $d_i^{\max} = \arg\max_d C_i(\boldsymbol{x}, d)$ performs a ray-based denoising of these measures and represents the voting scheme proposed by Hernández and Schmitt [9] and μ is a scaling parameter. The data term f avoids trivial solutions of energy (1) by propagating the photometric information from Eq. (2) in a probabilistic manner into the volume.

$$f(\boldsymbol{x}, t) = -\ln\left(\frac{1 - P(\boldsymbol{x} \in int(\Sigma))}{P(\boldsymbol{x} \in int(\Sigma))} \right) \ . \tag{3}$$

The probability $P(\boldsymbol{x} \in int(\Sigma))$ that point $\boldsymbol{x}$ belongs to the interior of surface Σ is defined based on the voting locations and qualities of corresponding camera rays $r_i(\boldsymbol{x}, \cdot)$ through point $\boldsymbol{x}$

$$P(\boldsymbol{x} \in int(\Sigma)) = \prod_{i=1}^{N}\prod_{j=1}^{N} \prod_{\substack{\mathrm{depth}_i(\boldsymbol{x}) < d \leq d_i^{\max}}} \frac{1}{Z_j} \exp\left[-\eta \cdot \mathrm{VOTE}_j\left(r_i(\boldsymbol{x}, d) \right) \right] \tag{4}$$

As suggested in [22] we limit the memory consumption of the method by setting $|T| = 3$ and taking the center frame as a smooth solution. For each frame a mesh is extracted with the Marching Cubes algorithm [20] at an iso-level of 0.5.

2.2   Connectivity Constraints via Directed Graphs

Without loss of generality we assume that the visual hull is connected. For the case that is not connected, the same approach can be applied component-wise after identifying independent connected components of the visual hull. We define connectivity constraints independently for each time step to allow for topology changes between time steps. For better readability we drop the temporal dependency in the following notation.

Graph Structure. For every time step we define a geodesic shortest path tree $\mathcal{G}_s$ on the visual hull $\mathcal{VH}$ with respect to a given source node s that contains for each point $x \in \mathcal{VH}$ inside the visual hull the shortest geodesic path C_s^x from s to x that minimizes the cost function

$$\mathcal{D}_s(x) = \ell(C_s^x) = \int_0^1 e^{f(C_s^x(r))} dr \ , \tag{5}$$

which is a positive geodesic measure that depends on the data term. Variable r parametrizes the path from s to x. $\mathcal{D}_s(x)$ is a shorthand for the distance map of the shortest geodesic path from the source node s to any point $x \in \mathcal{VH}$. The edges of the shortest paths form the edge set $\mathcal{E}$ of the shortest path tree $\mathcal{G}_s$.

Source Node Computation. It is desirable to center the source node for the geodesic shortest path computation within the data term. To this end, we compute the source node $s(t)$ as the point which minimizes a spatio-temporal convolution of the data term f with a sufficiently large Gaussian kernel $\mathcal{G}$.

$$s(t) = \arg\min_x \int_{t-1}^{t+1} (f * \mathcal{G})(x, \tau) \, d\tau \tag{6}$$

The minimization reflects the fact that negative data term values $f < 0$ indicate a favor for an interior label and thus ensures a position that has high probability of being interior. The position of the source node has not much influence on the result, but this choice favors a smoothly temporal change of its position within the data term while maximizing the distance to the surface. An example rendering of a shortest path from a leaf node to the source is shown in Fig. 4a.

Constrained Optimization. The connectivity constraint from [25] is included into the reconstruction process as a monotonicity constraint of the labeling function u with respect to the edges $\mathcal{E}$ in $\mathcal{G}_s$. This monotonicity can be ensured by including inequality constraints on the directional derivative $\delta_e(u(x, t))$ of u along every edge $e \in \mathcal{E}$. Thus, computing a spatio-temporal 3D reconstruction with connectivity constraints can be achieved by computing a minimizer of the constrained optimization problem

$$\min_{u \in \mathcal{BV}(V \times T; \{0,1\})} E(u) \tag{7}$$

$$\text{s. t.} \quad \delta_e(u(x, t)) \leq 0, \quad e \in \mathcal{E}$$

with one constraint for each edge e in the edge set $\mathcal{E}$ of the shortest path tree $\mathcal{G}_s$. $\mathcal{BV}(\cdot)$ denotes the function space of bounded variations [2].

3   Generalized Connectivity Constraints for Objects of Arbitrary Genus

The key idea to generalize the connectivity constraint to objects with arbitrary genus is a modification of the constraints that are defined on the geodesic shortest

path tree. The key ingredient to this modification is to detect loops in the object and to identify parts of these loops with a 'thin' geometry, called handles. This is described in the following.

3.1   Handle and Tunnel Loops

In [8], Dey et al. study arbitrary surfaces represented by a simplicial complex, that is, a hierarchy of p-simplicies with different dimensions p (e.g. $p = 0, \ldots, 2$ corresponding to points, edges, and faces). The surface $\mathbb{M}$ separates the simplicial complex into an interior part $\mathbb{I}$ and an exterior part $\mathbb{E}$, both including the surface, i.e. $\mathbb{I} \cap \mathbb{E} = \mathbb{M}$. Since we want to analyze the topology of the visual hull, these sets will be shorthands for $\mathbb{M} = \partial \mathcal{VH}$, $\mathbb{I} = \mathcal{VH}$ and $\mathbb{E} = (V \setminus \mathcal{VH}) \cup \partial \mathcal{VH}$.

   The authors in [8] define and study cycles of edges ('loops') on the surface which build equivalence classes with respect to contraction or translation of the cycle - like a rubber band which can be moved along the surface, but not above holes in the surface. In this paper will call this equivalence relation $\sim_{\mathbb{M}}$ 'contractible' on the set $\mathbb{M}$, for example, we denote the relation that a loop $l_1 \subset \mathbb{M}$ is contractible to a loop $l_2 \subset \mathbb{M}$ on the set $\mathbb{M}$ as $l_1 \sim_{\mathbb{M}} l_2$. For simplicity we try to define terms and notation on a more intuitive level which should be sufficient to follow the rest of the paper. For mathematically precise definitions based on persistent homology we refer to [8]. Following their work, we now consider loops on the surface with the following properties.

Definition 1 (Handle and tunnel loops). *A **handle loop** $h \subset \mathbb{M}$ is a cycle of edges on the surface that is contractible in the interior ($h \sim_{\mathbb{I}} 0$) and not contractible on the surface ($h \nsim_{\mathbb{M}} 0$). A **tunnel loop** $t \subset \mathbb{M}$ is a cycle of edges on the surface that is contractible in the exterior ($h \sim_{\mathbb{E}} 0$) and not contractible on the surface ($h \nsim_{\mathbb{M}} 0$).*

   With respect to the above mentioned equivalence relation, a closed surface of genus g has exactly g classes of handle loops and g classes of tunnel loops induced by the surface embedding. We consider one representative loop with approximate minimal geometric length per class and denote them as the set of handle loops $\{h_i\}_{i=1}^g$ and the set of tunnel loops $\{t_i\}_{i=1}^g$. Hence, for each surface hole i we have a corresponding pair (h_i, t_i) of representative handle and tunnel loops.

   Examples of handle and tunnel loops are shown in Figs. 2, 3, 4c. Dey et al. [8] also propose an algorithm which computes handle and tunnel loops with approximate minimal length that is perfectly suited to process volumetric data. However, this algorithm is considerably slower than a recently published algorithm by Dey et al. [7] which only works for meshes. To this end, we extract an iso-surface mesh of the visual hull to efficiently compute handle and tunnel loops. The speed advantage of the method in [7] stems from the fact that it does not need a 3D tessellation of the scene. In [7], the concept of Reeb graphs is used to estimate an initial set of handle and tunnel loops and their geometric length is shortened in a subsequent refinement step.
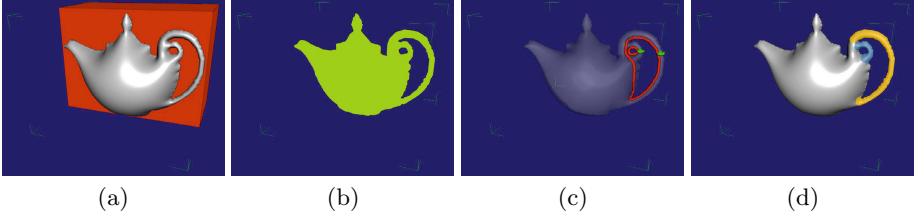
| (a) | (b) | (c) | (d) |

Fig. 2. Various sets defined in this section visualized on a teapot model of genus 2. (a) Exterior $\mathbb{E}$ (red), (b) Interior $\mathbb{I}$ (green), (c) Handle and tunnel loops $\{h_1, h_2\}, \{t_1, t_2\}$ (green+red), (d) Handle segments H_1, H_2 (yellow+blue).

Handle Segmentation. We aim to segment the 'thin' geometric parts around the holes of the surface, called handles. These handle segments will help to make the connectivity constraints adaptive to the data term. For this purpose we introduce the following definitions.

Definition 2 (Handle Segment Surface). *We define the handle segment surface as the connected subset of all points $\boldsymbol{x} \in \mathbb{M}$ for which a handle loop $h_{\boldsymbol{x}}$ exists which is contractible to h_i subject to the additional constraint that the ratio of $\ell(h_{\boldsymbol{x}})$ and $\ell(h_i)$ does not exceed a user given threshold σ:*

$$\mathbb{M}_{H_i} = \left\{ \boldsymbol{x} \in \mathbb{M} \;\middle|\; \exists h_{\boldsymbol{x}} \subset \mathbb{M} : h_{\boldsymbol{x}} \sim_{\mathbb{I}}^{\sigma} h_i \right\} \tag{8}$$

where $h_{\boldsymbol{x}} \subseteq \mathbb{M}$ denotes a handle loop through the surface point $\boldsymbol{x}$ and $h_{\boldsymbol{x}} \sim_{\mathbb{I}}^{\sigma} h_i$ means that handle loop $h_{\boldsymbol{x}}$ is contractible to h_i subject to the constraint $\ell(h_{\boldsymbol{x}}) < \sigma\ell(h_i)$.

Definition 3 (Handle Segment). *Given the handle segment surface $\mathbb{M}_{H_i}$ from the previous definition, we define the corresponding volumetric handle segment $H_i \subseteq \mathbb{I}$ as the set of all points in the visual hull for which the closest point on the visual hull boundary is on the handle segment surface $\mathbb{M}_{H_i}$.*

$$H_i = \left\{ \boldsymbol{x} \in \mathbb{I} \;\middle|\; \operatorname*{arg\,min}_{\boldsymbol{y} \in \mathbb{M}} \operatorname{dist}(\boldsymbol{x}, \boldsymbol{y}) \in \mathbb{M}_{H_i} \right\} \tag{9}$$

where $\operatorname{dist}(\boldsymbol{x}, \boldsymbol{y})$ denotes the Euclidean distance between point $\boldsymbol{x} \in \mathbb{I}$ in the interior and point $\boldsymbol{y} \in \mathbb{M}$ on the surface.

In practice, we compute H_i by a breadth first search algorithm on the visual hull. Starting from the handle loop h_i a wavefront is propagated in both directions. Independently for each wavefront, we stop the search if the ratio between the current length of the wavefront and the initial position exceeds the threshold σ.

3.2   Loop Connectivity Constraints

With the handle and tunnel loops of the visual hull we are now able to generalize the connectivity constraint in the presence of loops. By enforcing interior labels

along each tunnel loop t_i we can assure that loops in the visual hull are preserved in the final segmentation. However, in order add a minimum amount of costs to the energy (7) when enforcing loop connectivity, we need to find corresponding loops that respect the costs of the data term. We approximate these geodesics shortest loops by computing corresponding loops $t_i^{\mathcal{G}_s} \subset \mathbb{I}$ on the precomputed geodesic shortest path tree $\mathcal{G}_s$ which are contractible to the original tunnel loop on the surface, i.e. $t_i^{\mathcal{G}_s} \sim_{\mathbb{I}} t_i$. The computation of $t_i^{\mathcal{G}_s}$ is discussed later in this section. For each tunnel loop t_i of the visual hull we define a *loop preserving* constraint as

$$\forall i \in [1, \ldots, g] : \quad \left\{ \forall \boldsymbol{x} \in t_i^{\mathcal{G}_s} : u(\boldsymbol{x}) = 1 \right\}. \tag{C0}$$

Proposition 1. *The constraint* (C0) *preserves the handle and tunnel loops and thus all holes of the visual hull in the reconstructed object. The topological genus of the reconstructed object is larger or equal to the one of the visual hull.*

Proof. Let us assume that the proposition does not hold. To let the genus of the reconstructed object decrease, either (i) at least one hole of the visual hull needs to be filled or (ii) at least one tunnel loop has to be disconnected in the reconstructed object. Because the domain of the reconstructed object is restricted to the visual hull, (i) cannot be fulfilled. By construction, (ii) is fulfilled if (C0) is fulfilled. Therefore the genus of the reconstructed object has to be larger or equal to the genus of the visual hull.

Note that, depending on the data term f the reconstructed object is allowed to have more holes than the visual hull. In some cases, it is not desirable to exactly preserve all holes and corresponding handles of the visual hull. A possible scenario is depicted in Fig. 3 where aliasing artifacts of the visual hull lead to spurious handle loops which should not be preserved in the final reconstruction. Therefore we propose to relax the loop preserving constraint (C0) such that either the connectivity of a handle is preserved in the final reconstruction or, in case the photometric support via f is not strong enough, the handle segment H_i is suppressed completely. We define the *generalized connectivity* constraint as

$$\forall i \in [1, \ldots, g] : \quad \left\{ \forall \boldsymbol{x} \in t_i^{\mathcal{G}_s} \cap H_i : \frac{d}{ds} u(\boldsymbol{x}) = 0 \right\} \tag{C1}$$

where $\frac{d}{ds}$ is the directional derivative along the loop $t_i^{\mathcal{G}_s}$.

Finding the optimal connected loop $t_i^{\mathcal{G}_s}$. For objects of genus 0, the use of the shortest path tree in the connectivity constraint is motivated by the optimal connecting path, that adds the minimum cost to the final segmentation result. In case of objects with higher genus, we wish to preserve the connectivity with respect to loops in the final segmentation. Therefore a loop through each handle needs to be found, which is optimal in the same way, i.e. that it also adds the minimum cost to the final segmentation. Using the already computed shortest path tree $\mathcal{G}_s$, we can find the shortest loop $t_i^{\mathcal{G}_s}$ with respect to $\mathcal{G}_s$ for each
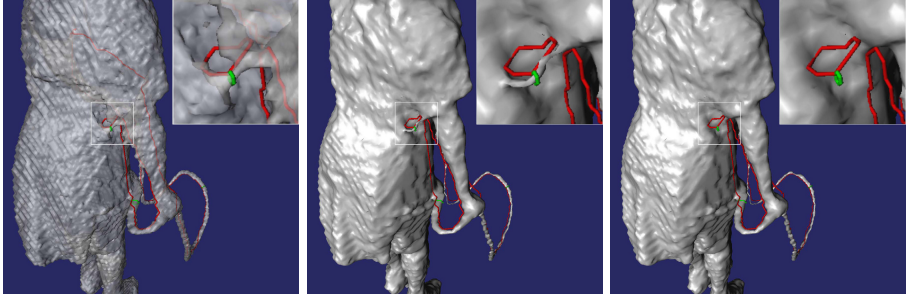
(a) Visual Hull   (b) Result with strict loop   (c) Result with generalized
                         connectivity (C0)   connectivity (C1)

Fig. 3. (a) In some cases artifacts of the visual hull can lead to spurious handle loops which should not be preserved in the final reconstruction. (b) The constraint C0 strictly preserves all loops in the solution. (c) Relaxing the topology preserving constraint to our generalized connectivity constraint allows to suppress handles where the photoconsistency is not strong enough. The rope, where the support of the photoconsistency is sufficient, is still completely preserved. Handle and tunnel loops are depicted in green and red, respectively.

handle i by the following steps: With a depth first search on $\mathcal{G}_s$, starting from the boundary of a handle segment H_i, we compute the partitions $H_i^1 \cup H_i^2 = H_i, H_i^1 \cap H_i^2 = \emptyset$ which are disconnected on the shortest path tree $\mathcal{G}_s$. These partitions are shown in Fig. 4d. If one of these partitions is empty, i.e. all points in the handle segment H_i are connected on $\mathcal{G}_s$, then no further constraints need to be added in order to preserve handle segment H_i. Otherwise, we compute an optimal pair of points

$$(\boldsymbol{p}, \boldsymbol{q}) = \arg\min_{(\boldsymbol{x} \in H_i^1, \boldsymbol{y} \in H_i^2, \boldsymbol{y} \in \mathcal{N}(\boldsymbol{x}))} \mathcal{D}_{\boldsymbol{s}}(\boldsymbol{x}) + \mathcal{D}_{\boldsymbol{s}}(\boldsymbol{y}) \qquad (10)$$

which are leaf-nodes in $\mathcal{G}_s$. The set $\mathcal{N}(\boldsymbol{x})$ denotes the local spatial neighborhood of a point $\boldsymbol{x} \in V$. The optimal path through the handle is computed by tracing the path backwards along the predecessors of both nodes $\boldsymbol{p}, \boldsymbol{q}$ in $\mathcal{G}_s$, resulting in the path with minimum costs through the handle (Fig. 4e).

While the tree connectivity constraint resulted in an inequality constraint on the derivative of the label function, the loop connectivity is preserved by adding the equality constraints

$$\delta_e (u(\boldsymbol{x}, t)) = 0, \quad e \in \mathcal{E}_=. \qquad (11)$$

to the optimization problem (7), where $\mathcal{E}_=$ is the set of edges of the optimal path through the handle.
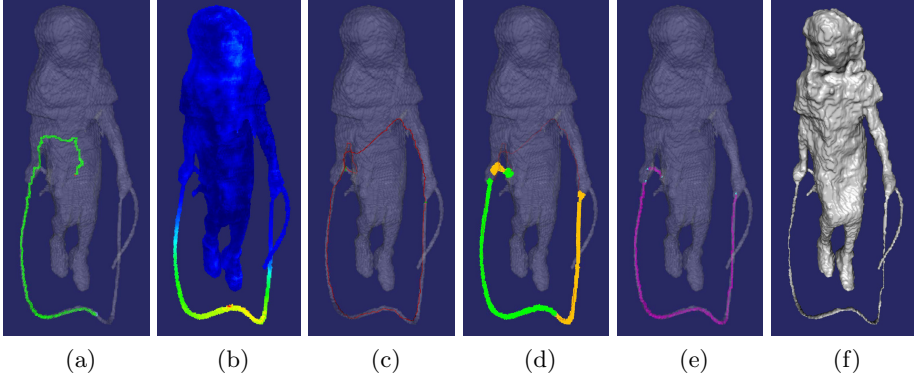
Fig. 4. Visualization of various properties that we compute based on the shape of the visual hull (genus 2 in this case) and the data term. (a) Example shortest path from a leaf node to the source node s (red); (b) color-coded geodesic distance map $\mathcal{D}_s$ with respect to the source node s; (c) handle (green) and tunnel (red) loops; (d) handle segmentations $H_i = H_i^1 \cup H_i^2$ (green+orange), the coloring shows disconnected parts within the handle with respect to the geodesic path tree $\mathcal{G}_s$. (e) shortest path through the handle for which the equality constraints (C1) are imposed; (f) final reconstruction result.

4   Numerical Optimization

To minimize energy (7) using convex optimization we first relax the discrete image function to the continuous interval $[0, 1]$. The constraints defined on the derivative of the image function remain the same as in the discrete setting.

Because the total variation norm is non-differentiable, we introduce a dual variable $p : V \times T \mapsto \mathbb{R}^4$ and reformulate the optimization problem Eq.(7) as the equivalent saddle-point problem

$$\min_{u} \max_{\|p\| \leq 1} \int_{V \times T} \langle u, -\operatorname{div}(p) \rangle \ d\boldsymbol{x} dt + \lambda \int_{V \times T} fu \ d\boldsymbol{x} dt \ . \tag{12}$$

$$\text{s. t.} \qquad \delta_e\left(u(\boldsymbol{x}, t)\right) \leq 0, \quad e \in \mathcal{E}$$

$$\delta_e\left(u(\boldsymbol{x}, t)\right) = 0, \quad e \in \mathcal{E}_=$$

The constraints on u over the edge sets $\mathcal{E}$ and $\mathcal{E}_=$ are included in the optimization using Lagrangian multipliers β and γ. The Lagrangian associated to problem (12) becomes

$$\min_{u} \max_{\substack{\|p\| \leq 1, \\ \beta \geq 0, \\ \gamma}} \int_{V \times T} \langle u, -\operatorname{div}(p) \rangle \ d\boldsymbol{x} dt + \lambda \int_{V \times T} fu \ d\boldsymbol{x} dt \tag{13}$$

$$+ \int_{T} \left\{ \sum_{e \in \mathcal{E}} \beta_e \, \delta_e\left(u\right) + \sum_{e \in \mathcal{E}_=} \gamma_e \, \delta_e\left(u\right) \right\} dt \ .$$

This saddle point problem is optimized using the preconditioned primal-dual algorithm by Pock and Chambolle [23]. The algorithm results in an iterative update scheme with a gradient ascent in the dual and a gradient descent in the primal variable

$$
\begin{aligned}
p^{n+1} &= \Pi_C \left[p^n + \sigma \nabla \bar{u}^n \right] \\
\beta_e^{n+1} &= \Pi_{\geq 0}(\beta_e^n + \mu \, \delta_e \, (\bar{u}^n)) \\
\gamma_e^{n+1} &= \gamma_e^n + \nu \, \delta_e \, (\bar{u}^n) \\
u^{n+1} &= \Pi_{[0,1]} \left[u^n + \tau \Big(\operatorname{div} p^{n+1} + \operatorname{div} \beta^{n+1} + \operatorname{div} \gamma^{n+1} - \lambda f \Big) \right] \\
\bar{u}^{n+1} &= 2u^{n+1} - u^n
\end{aligned}
\tag{14}
$$

where $\Pi_{[0,1]}$ is the projection of u onto the unit interval $[0,1]$ and $\Pi_{\geq 0}$ onto positive values. The projection onto the set $C = \{ q = (q_{\boldsymbol{x}}, q_t)^T : V \times T \mapsto \mathbb{R}^4 \,|\, \|q_{\boldsymbol{x}}\| \leq 1, |q_t| \leq 1 \}$ is a projection on a 4D hyperball and can be done as follows:

$$
\Pi_C(q) = \left(\frac{q_{\boldsymbol{x}}}{\max(1, \frac{\|q_{\boldsymbol{x}}\|}{\rho})}, \max \big(- g_t, \min(g_t, q_t) \big) \right)^T
\tag{15}
$$

The step sizes τ, σ, μ and ν are chosen as suggested in [23]. Because our energy model is convex and the linear constraints preserve convexity of the optimization problem, the update scheme (14) converges to a global minimum of the relaxed energy (7). An optimal binary labeling can be found by thresholding the relaxed solution [23].

Implementation. The proposed iterative scheme for minimal surface reconstruction with connectivity constraints (14) allows a high degree of parallelization and is implemented using the CUDA programming framework. The connectivity graph precomputation is more difficult to parallelize and therefore is implemented on the CPU.

5 Experiments

We evaluated our method on several spatio-temporal multi-view data sets provided by the INRIA 4D repository [16]. All scenes were synchronously recorded by 16 cameras in a green room environment.

In the experiments we mainly focus on comparing reconstruction results with and without connectivity constraints. Since no other 4D reconstruction methods are publicly available, we compare our results with the ones of the state-of-the-art 3D reconstruction methods by Jancosek and Pajdla [17] and the combination of Furukawa et al. (PMVS) [10] and Poisson surface reconstruction [18].

Approximate silhouette information was used for all methods except of the method by Jancosek and Pajdla [17] for which it cannot be used. We used the 6-neighborhood for the computation of the geodesic shortest path tree $\mathcal{G}_{\boldsymbol{s}}$. In this setting, the generalization to arbitrary genus by using equality constraints does not increase the number of dual variables (Lagrange multipliers), because some inequality constraints are exchanged by equality constraints.
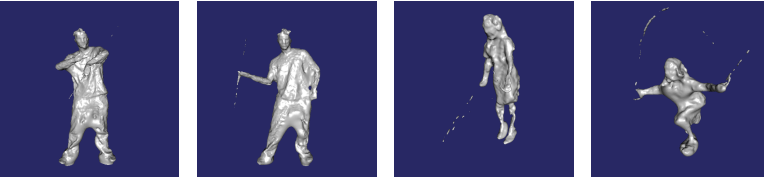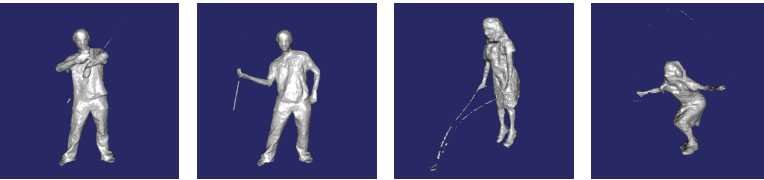
1 of 16 Input Images
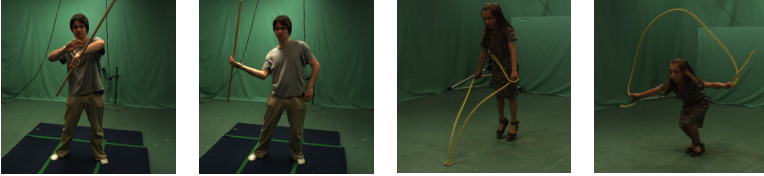
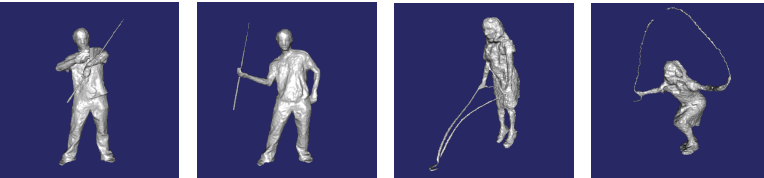Jancosek and Pajdla [17]
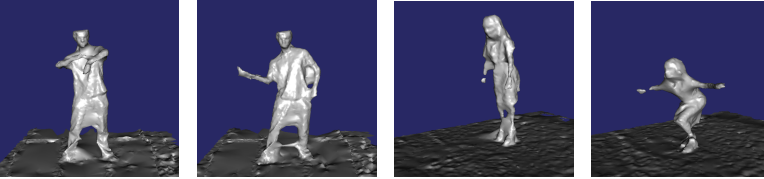
Furukawa et al. (PMVS) [10] + Poisson surface reconstruction [18]

Without Connectivity Constraint [22]

With Connectivity Constraint [25]+[22]

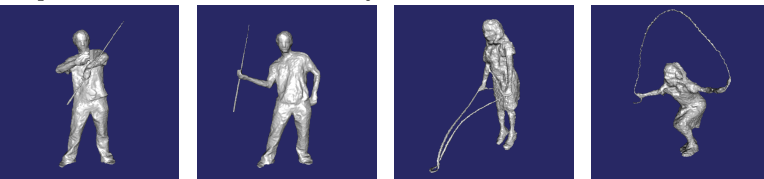

Proposed Generalized Connectivity Constraint

Fig. 5. Comparison of different reconstruction methods: Existing state-of-the art approaches [17,10,18] fail to recover thin structures like the stick and the rope. The connectivity constraint allows to preserve the stick, but for the rope-jump scene with higher genus, it does not completely preserve the connection of the rope. Our proposed generalized connectivity constraint allows to correctly reconstruct both scenes (volume resolution $|V| = 384^3$).

Runtime and Memory Resource Evaluation. The memory footprint of the suggested implementation increases only by $|V \times T|$ bytes in comparison to the original approach. The numerical optimization runtime per iteration remains almost unchanged, but depending on the scene structure more iterations are needed for sufficient convergence. All experiments were run on a Linux-based Intel Xeon E5520 PC with 24GB RAM and NVidia GTX Titan graphics card. For the genus 0 connectivity [25] the precomputation time per frame was about 20 sec for computing the tree of the tree-shaped connectivity constraints. For the generalized connectivity constraints the precomputation time was about 1 min for handle and tunnel loop detection, handle segmentation and computation of the tree. The optimization needs about 3 min per frame resulting in a total runtime of about 4 minutes per frame when using the generalized connectivity constraints.

6   Conclusion

In this paper we introduced tree-shaped connectivity constraints into spatio-temporal multi-view 3D reconstruction. By detecting loops in the object we are able to generalize the connectivity constraint to objects with non-tree structure of arbitrary genus. In several experiments, we demonstrated that the proposed connectivity constraints significantly improve the reconstruction quality in the presence of fine elongated structures.

To the best of our knowledge, apart from the work in [3], which uses a strong simplification of a connectivity prior and essentially is a 2.5D method, this is the first work which imposes connectivity constraints in a multi-view 3D reconstruction setup.

The connectivity constraint is especially useful in 4D multi-view settings, for which exact silhouettes are usually not available and exact silhouette constraints are not applicable. Assuring temporal coherence of the connectivity constraints would need explicit modeling of the occupancy flow and remains for future work.

References

1. Aganj, E., Pons, J.P., Ségonne, F., Keriven, R.: Spatio-temporal shape from silhouette using four-dimensional delaunay meshing. In: ICCV, pp. 1–8 (2007)
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. Oxford Mathematical Monographs. The Clarendon Press Oxford University Press, New York (2000)
3. Bleyer, M., Rother, C., Kohli, P., Scharstein, D., Sinha, S.: Object stereojoint stereo matching and object segmentation. In: CVPR, pp. 3081–3088. IEEE (2011)
4. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE TPAMI 23(11), 1222–1239 (2001)
5. Chen, C., Freedman, D., Lampert, C.H.: Enforcing topological constraints in random field image segmentation. In: CVPR, pp. 2089–2096 (2011)
6. Cremers, D., Kolev, K.: Multiview stereo and silhouette consistency via convex functionals over convex domains. IEEE TPAMI 33, 1161–1174 (2011)

7. Dey, T.K., Fan, F., Wang, Y.: An efficient computation of handle and tunnel loops via reeb graphs. ACM Trans. Graph. 32(4), 32 (2013)
8. Dey, T.K., Li, K., Sun, J., Cohen-Steiner, D.: Computing geometry-aware handle and tunnel loops in 3d models. ACM Trans. Graph. 27(3) (2008)
9. Esteban, C.H., Schmitt, F.: Silhouette and stereo fusion for 3d object modeling. CVIU 96(3), 367–392 (2004)
10. Furukawa, Y., Ponce, J.: Accurate, dense, and robust multiview stereopsis. IEEE TPAMI 32(8), 1362–1376 (2010), http://dx.doi.org/10.1109/TPAMI.2009.161
11. Goldluecke, B., Ihrke, I., Linz, C., Magnor, M.: Weighted minimal hypersurface reconstruction. IEEE TPAMI 29(7), 1194–1208 (2007)
12. Goldluecke, B., Magnor, M.: Space-time isosurface evolution for temporally coherent 3D reconstruction. In: CVPR, vol. I, pp. 350–355 (July 2004)
13. Guillemaut, J.Y., Hilton, A.: Space-time joint multi-layer segmentation and depth estimation. In: 3DIMPVT, pp. 440–447 (2012)
14. Gulshan, V., Rother, C., Criminisi, A., Blake, A., Zisserman, A.: Geodesic star convexity for interactive image segmentation. In: CVPR, pp. 3129–3136. IEEE (2010)
15. Han, X., Xu, C., Prince, J.L.: A topology preserving level set method for geometric deformable models. IEEE TPAMI 25(6), 755–768 (2003)
16. Institut national de recherche en informatique et en automatique (INRIA) Rhône Alpes: 4d repository, http://4drepository.inrialpes.fr/
17. Jancosek, M., Pajdla, T.: Multi-view reconstruction preserving weakly-supported surfaces. In: CVPR, pp. 3121–3128 (2011)
18. Kazhdan, M.M., Bolitho, M., Hoppe, H.: Poisson surface reconstruction. In: Symposium on Geometry Processing, pp. 61–70 (2006)
19. Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous global optimization in multiview 3d reconstruction. IJCV 84(1), 80–96 (2009)
20. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3d surface construction algorithm. SIGGRAPH Comput. Graph. 21, 163–169 (1987)
21. Nowozin, S., Lampert, C.H.: Global connectivity potentials for random field models. In: CVPR, pp. 818–825. IEEE (2009)
22. Oswald, M.R., Cremers, D.: A convex relaxation approach to space time multi-view 3d reconstruction. In: ICCV - Workshop on Dynamic Shape Capture and Analysis (4DMOD) (2013)
23. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: ICCV, Washington, DC, USA, pp. 1762–1769 (2011)
24. Starck, J., Hilton, A.: Surface capture for performance-based animation. IEEE Computer Graphics and Applications 27(3), 21–31 (2007)
25. Stühmer, J., Schröder, P., Cremers, D.: Tree shape priors with connectivity constraints using convex relaxation on general graphs. In: ICCV, Sydney, Australia (December 2013)
26. Vicente, S., Kolmogorov, V., Rother, C.: Graph cut based image segmentation with connectivity priors. In: CVPR (2008)
27. Zeng, Y., Samaras, D., Chen, W., Peng, Q.: Topology cuts: A novel min-cut/max-flow algorithm for topology preserving segmentation in n–d images. Computer Vision and Image Understanding 112(1), 81–90 (2008)