Soft 3D Reconstruction for View Synthesis

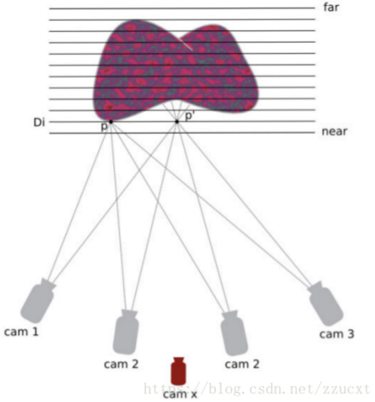
Jiabin CHEN

ECOLE POLYTECHNIQUE

March 13, 2019

Objective

Assuming that our cameras are calibrated and we have known the intrinsic parameters and projection matrix of these cameras, the main task in this paper[2] aims to render a virtual view based on 3D reconstruction obtained from given inputs images.



Overview

1. Soft 3D Reconstruction from given views
   - Initial depth estimation
   - Soft 3D Reconstruction
2. Image Synthesis for virtual view
   - Soft View Synthesis

Initial depth estimation

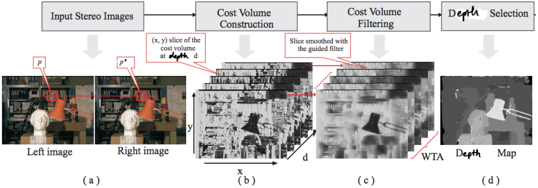


Figure: (a) Input stereo images. (b) Cost volume construction. Each (x, y) slice holds the dissimilarity costs at a given depth value d. (c) Cost volume filtering. Filtering is applied on each of the (x, y) slices using the guided filter[1]. (d) Depth selection. For each pixel, we select the depth of lowest costs in the smoothed volume of (c).

Three main steps:

- Cost Volume Construction
- Cost Volume Filtering
- Depth Selection

Cost Volume Construction

For a reference view with neighborhood views, we construct an array of shape $(rows, cols, DispRange)$, where the value $C(p, d)$ for pixel $p = (x, y)$ at depth d is obtained by measuring the dissimilarity between pixel p of the reference image and pixel p' of neighbor image.

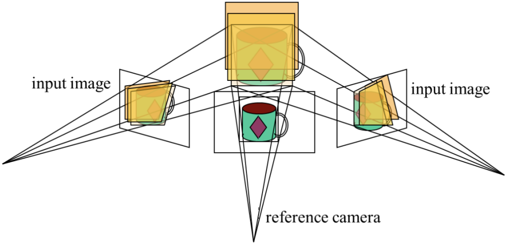


Figure: Sweep Plane

Measuring Dissimilarity

$$C(p,d) = \alpha Min(T_c, M(p,d)) + (1-\alpha) Min(T_g, G(p,d)) \tag{1}$$

where α balances the color and gradient terms, T_c, T_g are color and gradient truncation values. In equation (1), color absolute difference $M(p,d)$ is expressed as:

$$M(p,d) = \sum_{i=0}^{i=3} |I^i_{ref}(p) - I^i_{neighbor}(p^{'})| \tag{2}$$

where $I^i(p)$ denotes the value of the *ith* color channel(RGB)at pixel p. Similarily, gradient absolute difference $G(p,d)$ is computed by:

$$G(p,d) = |\nabla_x(I_{ref}(p)) - \nabla_x(I_{neighbor}(p^{'}))| \tag{3}$$

where $\nabla_x(I(p))$ denotes the gradient in x direction computed at pixel p in image I.

Cost Volume Filtering

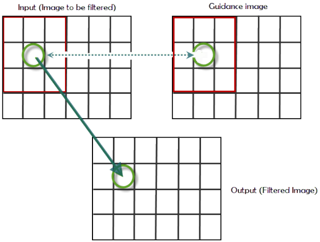


Figure: Guidance image and guided image

$C^{'}(p, d)$, the filtered cost value at pixel p and slice d:

$$C^{'}(p, d) = \sum_{q} W_{p,q}(I)C(q, d) \tag{4}$$

where the guidance image is the reference color image and the guided image is an (x, y) slice of the cost volume.

$W_{p,q}$ is expressed as follows:

$$W_{p,q} = \frac{1}{|\omega|^2} \sum_{k:(p,q)\in\omega_k} (1 + (I_p - \mu_k)^T (\Sigma_k + \epsilon U)^{-1} (I_q - \mu_k)) \quad (5)$$

Here, Σ_k and μ_k are the covariance matrix and mean vector of image I in the window ω_k with size (r, r), centered at pixel k. The number of pixels in this window is $|\omega|$, ϵ is a smoothness parameter, I_i, I_j and μ_k are 3 x 1(RGB) vectors.

Depth Selection

Once the cost volume slices are filtered as $C^{'}$, for each pixel, we select the depth of lowest costs in the smoothed volume $C^{'}$:

$$d_p = argmin_{d \in D} C^{'}(p, d) \tag{6}$$

where D represents the set of all allowed depths.

Soft 3D Reconstruction

Problems:

- One type of uncertainty is between neighboring pixels within a single image.
- Another type of uncertainty is across different views, e.g, multiple depth maps from different views do not always agree with each other about commonly visible points.

Solutions:

- Consider each pixels depth as a distribution originating from its neighborhood in the depth map rather than a single depth.
- Construct a vote volume to aggregate the consensus at each point in space.

$$VoteVal_{\mathrm{raw}}(x, y, z) = \begin{cases} 1 & z = D(x, y) \\ 0 & otherwise \end{cases}$$

$$VoteConf_{\mathrm{raw}}(x, y, z) = \begin{cases} 1 & z <= D(x, y) \\ 0 & otherwise. \end{cases}$$

Consensus and Visibility Function

- **Filtered vote volume**

$$VoteVal(x,y,z) = \sum_{(\hat{x},\hat{y})\in W(x,y)} w(x,y,\hat{x},\hat{y})\, VoteVal_{\text{raw}}(\hat{x},\hat{y},z)$$

$$VoteConf(x,y,z) = \sum_{(\hat{x},\hat{y})\in W(x,y)} w(x,y,\hat{x},\hat{y})\, VoteConf_{\text{raw}}(\hat{x},\hat{y},z),$$

- **Consensus function(array)**

$$Consensus'(x,y,z) = \frac{\sum_{k\in M} VoteVal_k(x'_k,y'_k,z'_k)}{\sum_{k\in M} VoteConf_k(x'_k,y'_k,z'_k)}$$

- **Soft Visibility Function(array)**

$$SoftVis(x,y,z) = max(0, 1 - \sum_{\hat{z}\in Z, \hat{z}<z} Consensus(x,y,\hat{z})).$$

Soft View Synthesis

- Calculate consensus volumes and then soft visibility volumes for all input views.
- Calculate interpolation weight $W_k()$ for each input view based on the geometry relation with synthesized view.
- In the synthesized view's reference, the color at a 3D point is interpolated from the input images $I_k()$:

$$Color_{\text{synth}}(x, y, z) = \frac{\sum_k SoftVis_k(x'_k, y'_k, z'_k) W_k(x'_k, y'_k) I_k(x'_k, y'_k)}{\sum_k SoftVis_k(x'_k, y'_k, z'_k) W_k(x'_k, y'_k)}$$

Soft View Synthesis

- The geometry(consensus) in front of the synthesized view is smoothly interpolated from our input views using our weights $W_k()$:

$$Consensus_{\text{synth}}(x, y, z) = \sum_k W_k(x'_k, y'_k) Consensus_k(x'_k, y'_k, z'_k)$$

- Once we have $Consensus_{synth}$, we then obtain $SoftVis_{synth}$ by:

$$SoftVis(x, y, z) = max(0, 1 - \sum_{\hat{z} \in Z, \hat{z} < z} Consensus(x, y, \hat{z}))$$

- Define $CC()$ as consensus clamped to the remaining ray visibility:

$$CC(x, y, z) = min(Consensus_{\text{synth}}(x, y, z), SoftVis_{\text{synth}}(x, y, z))$$

- Synthesized image

$$I_{\text{synth}}(x, y) = \frac{\sum_{z \in Z} Color_{\text{synth}}(x, y, z) CC(x, y, z)}{\sum_{z \in Z} CC(x, y, z)}$$

Conclusion

Contributions:

- A fast, local stereo methods can produce high quality view synthesis results.
- Vote-based representation is proposed, which provides visibility estimates to perform per-pixel view selection and improve depth edges in depth estimation, and provide soft visibility weights in synthesis.

Limitation: Geometry is reconstructed directly in volumes, so memory consumption increase linearly with depth precision, which limits the amount of free view-point movement away from the source views.

References

Kaiming He, Jian Sun, and Xiaoou Tang.
Guided image filtering.
In *Proceedings of the 11th European Conference on Computer Vision: Part I*, ECCV'10, pages 1–14, Berlin, Heidelberg, 2010. Springer-Verlag.

Eric Penner and Li Zhang.
Soft 3d reconstruction for view synthesis.
ACM Transactions on Graphics (Proc. SIGGRAPH Asia), 36, 2017.