Evidence Based Image Selection for 3D Reconstruction

Smita C. Yadavannavar(✉), Varad Vinod Prabhu, Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudengudi

KLE Technological University, Hubli, India
scy15398@gmail.com

Abstract. In this paper, we propose a framework for image selection using evidence theory, towards 3D reconstruction. The process of 3D reconstruction involves image acquisition, image selection, feature extraction, calculating camera parameters, and generation of the point cloud. Of these all, image selection plays a significant role, as it has a significant impact on the final 3D model. However, in large scale 3D reconstruction, image selection based on a single parameter is not sufficient. We measure the similarity between the images using multiple parameters and generate a combined confidence score towards discarding similar images. Experimental results show that the quality of 3D reconstruction is better using images selected by the proposed method.

Keywords: 3D reconstruction · Image selection · Dempster Shafer Combination Rule (DSCR) · SSIM · FLANN based parameter.

1 Introduction

In this paper, we propose a framework for image selection towards 3D reconstruction using evidence theory. 3D reconstruction is the process of the generation of three-dimensional models from multiple images of the object. 3D reconstruction finds its applications in the field of Computer Aided Geometric Design (CAGD), computer graphics, computer vision, computer animation, medical imaging, computational science, virtual reality, digital media, reconstruction of heritage sites, etc.

Typically image acquisition, image selection, evaluation of camera parameters and point cloud generation are the modules in 3D reconstruction. The challenges involved in 3D reconstruction are: less variation in data, variation in lighting conditions, pairwise matching between images is computationally expensive.

To address these challenges, there is a need for efficient image selection algorithm which selects the valid images, improves the reconstruction quality and optimizes the time required for reconstruction. In large scale 3D reconstruction we perform reconstruction of large sites, cities, heritage sites.

Several methods, pipelines have been proposed towards 3D reconstruction using large scale data. Authors in [6] propose a technique for image selection

© Springer Nature Singapore Pte Ltd. 2020
R. V. Babu et al. (Eds.): NCVPRIPG 2019, CCIS 1249, pp. 53–63, 2020.
https://doi.org/10.1007/978-981-15-8697-2_5

using Structure From Motion (SFM) to compute the position and orientation of each camera and the contribution of each image towards 3D reconstruction is computed. Based on the contribution of each image and effect of its contribution to 3D reconstruction a decision is made to select the image. The limitation in this approach is that time taken is more than a traditional approach since each image is checked for its contribution to the final 3D reconstruction. Authors in [3] propose an image-pair selection by creating a bag of visual words based on vector similarity and set similarity. The term frequency-inverse document frequency (tf-idf) weighting based similarity is used as vector similarity and modified Simpson's similarity is used as a set similarity. The image pairs are selected based on the intersection of the results obtained by vector similarity and set similarity. Due to the creation of a bag of visual words for large scale 3D reconstruction, there may be memory issues. The tool used in the paper [2] named Imaging Network Designer (IND) to cluster and select vantage images in a dense imaging network. They require a suitable hardware arrangement.

Researchers have also proposed the images selection based on key-frame selection. The authors in [4] discuss the keyframe selection algorithm. Keyframes are defined as the set of a frame that satisfies the epipolar geometry between the views. The keyframe selection algorithm is based on the "Geometric Robust Information Criterion (GRIC)". In this case, the number of feature points may decrease significantly as the baseline between the current frame and the last key-frame increases. Authors in [5] use a keyframe selection algorithm based on those images which have a large number of feature points and a sufficient baseline between each keyframe. Here the keyframe pairing is done to improve the probability of convergence of bundle adjustment.

Authors in [1] propose an image selection by the removal of redundant images. Images are first sorted in the increasing order of the image size, so that smaller size images are removed first. Then each image is removed to check if the coverage constraint holds well after removal of the image. If the coverage constraint is satisfied, then the image is permanently removed. This process is continued over all the sorted images.

Presently people use single parameter or hardware adjustments for 3D reconstruction, but large scale, 3D reconstruction one parameter cannot give the desired results and there cannot be hardware setup for this case. [3] uses two parameters but as a bag of visual words is used it may not work on large scale reconstruction. To overcome this, we propose an approach where we use two parameters and based on an evidence-based technique [9] to perform image selection for 3D reconstruction. Towards this,

- We propose to model a similarity score for the selection of images towards 3D reconstruction.
  - We choose two parametric scores namely Structural Similarity Index (SSIM) and Fast Library for Approximate Nearest Neighbor (FLANN) based parameter to generate a combined score.
  - We propose to combine similarity parametric scores using Dempster Shafer Combination Rule (DSCR) [7] to generate a confidence score.

  - We propose to use the confidence score to eliminate redundant images.
- We demonstrate the results using a heritage dataset and compare the results using state-of-art methods.

2 Evidence Based Image Selection

We demonstrate the proposed framework in Fig. 1 to find the contribution of the images by detecting the similarity between images and find the parametric score of images being similar by using two parameters. Further, combine the two parametric scores using DSCR to obtain the confidence score. We decide if the given image is suitable for 3D reconstruction based on the decision threshold set upon the confidence score.

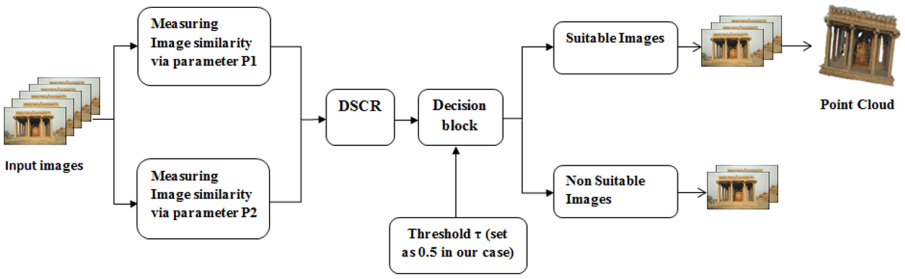

Fig. 1. Evidence based image selection

2.1 Parameters for Image Similarity

In large scale 3D reconstruction, measuring the similarity via a single parameter is not sufficient. This is demonstrated as shown in Table 1. Considering the row 1 of Table 1 column 1 signifies that image 1 was chosen for the image similarity. Flann based method estimated image 1 is similar to image 2 whereas SSIM based similarity estimated image 1 is similar to image 27. The DSCR rule is used to obtain the confidence score based on the two parametric scores and said image 1 was similar to image 27.

We use FLANN (Fast Library for Approximate Nearest Neighbors) based matcher as the first parameter (P_1). FLANN is a library for performing fast approximate nearest neighbor searches in high dimensional spaces. Classical feature descriptors (SIFT [12], SURF [13]) are typically compared and matched using the Euclidean distance (or L2-norm).

$$Euclidean\ distance = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{1}$$

Table 1. Results on the dataset

Image number	Flann Based similar image number	SSIM based similar image number	Our approach based similar image number
1	2	27	27
2	1	15	16
3	1	10	15
4	2	5	35
5	20	33	34
6	5	8	18
7	9	30	3
8	7	14	16
9	1	31	3
10	1	25	31

Fig. 2. Images corresponding to Table 1

Here (x_1, y_1) represents the location of feature 1 in $image_1$ and (x_2, y_2) represents the location of feature 2 in $image_2$. Whereas, binary descriptors (ORB [14], BRISK [15]) are also used for matching using Hamming distance.

$$d_{hamming}(a, b) = \sum_{i=0}^{n-1} (a_i \oplus b_i) \tag{2}$$

Here a and b are binary strings of length n. This distance is equivalent to count the number of different elements for binary strings (population count after applying an XOR operation).

We use the Structural Similarity Index (SSIM) as the second parameter (P_2). SSIM measures the perceptual difference between two similar images. SSIM is based on visible structures of the images. The SSIM index is calculated on various

pairs of a window taken from both the images. Then the measure of SSIM between two windows x of $image_1$ and y of $image_2$ of common size N × N is:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3}$$

where:

μ_x is average of x;
μ_y is average of y;
σ_x^2 is variance of x;
σ_y^2 is variance of y;
$c_1 - (k_1 L)^2, c_2 - (k_2 L)^2$ two variables to stabilize the division with weak denominator;
L the dynamic range of the pixel-values (typically this is $2^{\#\text{ bits per pixel}} - 1$);
$k_1 = 0.01$ *and* $k_2 = 0.03$ by default and kept constant throughout our experimentation.

Let N be the total number of images in the dataset for 3D reconstruction. We perform one versus all similarity check. Let $image_i$ be the current query image. It is compared with all the other $(N - 1)$ images in the dataset, based on this parametric score P_1 and P_2 are calculated. The Flann based parametric score P_1 is obtained by the number of valid matches between the image pair divided by the total number of features detected for one image (in our case the total number of features to be detected for one image is fixed to 500). The SSIM based parametric score P_2 is obtained by the SSIM score between the image pairs. The obtained parametric scores are combined using DSCR.

2.2 Confidence Score Using DSCR

We combine the two parametric scores using the Dempster Shafer Combination Rule (DSCR) to obtain the confidence score. Confidence score is obtained using DSCR based on the two parametric scores. We decide if the given query image is suitable for 3D reconstruction based on the decision threshold set upon the confidence score to maximize the probability of images being similar. Let P_1 and P_2 be the parametric score to be combined. DSCR combines two hypotheses consisting of three parameters, mass of belief, the mass of disbelief and mass of uncertainty rather than two parametric scores. We construct a hypothesis, hyp_1, and hyp_2 as a set of the mass of belief $(m(b))$, disbelief $(m(d))$ and uncertainty. $(m(u))$ respectively. We set the mass of belief $(m_1(b))$ for hyp_1 as P_1 and mass of belief $(m_2(b))$ for hyp_2 be P_2. We assume the mass of disbelief $(m_1(d))$ for hyp_1 and hyp_2 to be 0 and mass of uncertainty $(m_1(u)$ *and* $m_2(u))$ for hyp_1 and hyp_2 as $1 - P_1$ and $1 - P_2$ respectively. We combine hyp_1 and hyp_2 using a combination table as shown in Table 2.

In the above combination table, the product of the mass of belief of one hypothesis and mass of disbelief of other hypothesis gives rise to conflict and is represented by ϕ. The product of the mass of belief and mass of belief or the

Table 2. Combinational table

$\cap$	m_1^{belief}	$m_1^{disbelief}$	$m_1^{ambiguity}$
m_2^{belief}	ψ_1	ϕ	ψ_1
$m_2^{disbelief}$	ϕ	ψ_2	ψ_2
$m_2^{ambiguity}$	ψ_1	ψ_2	Ω

product of the mass of belief and mass of uncertainty represents joint belief and is represented by ψ_1. Similarly, ψ_2 represents the joint disbelief. The combined belief of two pieces of evidence is considered as confidence score and is given by:

$$Confidence\ score = \frac{\sum \psi_1}{1 - \sum \phi} \tag{4}$$

We decide if the given frame is suitable for 3D reconstruction based on the decision threshold τ set upon the confidence score [8]. The advantage of using DSCR for combining the two parametric scores is that it emphasizes the fact that if P_1 is the probability of the image being suitable for 3D reconstruction, then $1 - P_1$ need not be the probability of image not being suitable for 3D reconstruction. It can be uncertainty as well. Heuristically the value of decision threshold τ is set to 0.5 based on various heritage sites.

2.3 Algorithm

Algorithm 1. Evidence-based image selection

Input: *Images* $(i_1 \ldots i_N)$
Output: *Selected Images*

for $i \leftarrow 1$ to N **do**
  for $j \leftarrow 1$ to N **do**
    if i is not j **then**
      $P_1 \leftarrow FLANN\_Based\_parameter(image_i, image_j)$
      $P_2 \leftarrow SSIM(image_i, image_j)$
      $Confidencefactor \leftarrow DSCR(P_1, P_2)$
      if $Confidencefactor$ greater than 0.5 **then**
        Image is rejected
      else
        Image is selected
      end if
    end if
  end for
end for

By using Algorithm 1 we perform evidence-based image selection on heritage sites.

3 Results and Discussions

In this section, we demonstrate our results using real-time heritage datasets. We compare our results with the state-of-art [10,11] methods using the number of images retained after the image selection process. We have experimented with two heritage sites of Karnataka 1) Sasivekalu Ganapati of Hampi, 2) Mahadeva Temple of Koppal and one dataset generated by us bell dataset with no redundant images and which consists of 48 non-redundant images taken around a bell. We analyze our results quantitatively and qualitatively.

3.1 Quantitative Analysis

Table 3. Results on the dataset

Dataset	Number of images in dataset	Number of redundant images	Number of images retained using our approach	Time taken for 3D reconstruction (using all images) using MVG-MVS pipeline	Time taken for 3D reconstruction (our approach)
Sasivekalu Ganapati of Hampi	129	49	80	4 h 57 min	2 h 51 min
Bell	48	0	48	1 h 25 min	1 h 40 min
Mahadeva Temple of Koppal	301	75	226	2 days 43 min (for texture memory insufficient)	1 day 1 h 21 min

The proposed framework is implemented on the HP workstation with 64 GB of RAM.

Table 3 shows the results with and without using our framework. The Sasivekalu Ganapati of Hampi dataset consists of 129 images, 49 images were redundant according to our framework. The point cloud obtained by our framework is comparable with the state of art. The point cloud of the Sasivekalu Ganapati of Hampi is shown in Fig. 3.

We also experimented with the framework on bell dataset results of which are shown in Table 3. The bell dataset contains no redundant images. Our framework took 25 minutes extra when compared to that of the state of art (MVG-MVS pipeline) [10,11]. Since we are working on large scale 3D reconstruction there is a very low possibility of fewer redundant images. The point cloud obtained on the bell dataset is shown in Fig. 4.

Since our objective was to work on large scale 3D reconstruction we experimented with our approach on large datasets like the Mahadeva Temple of Koppal which consists of 301 images, in which there were redundant. With our framework 226 images were selected and were able to generate a point cloud with texture

(a) Using MVG-MVS pipeline (b) With our framework

Fig. 3. Point cloud of Sasivekalu Ganapati of Hampi.

(a) Using MVG-MVS pipeline (b) With our framework

Fig. 4. Point cloud of Bell dataset.

on the implemented system but, without our approach, point cloud with texture was not generated due to insufficient memory for rendering. Figure 5 shows the point cloud with and without our approach. Figure 5a shows the dense point cloud on Mahadeva Temple of Koppal without our framework. Figure 5b shows the dense point cloud on Mahadeva Temple of Koppal with our framework.

3.2 Qualitative Analysis

Figure 6 depicts the subjective analysis of the experimentation. The graph values indicate the number of people who were satisfied with the quality of the point cloud obtained with our approach or using state-of-art (MVG-MVS pipeline). The higher the value better is the quality of the result obtained. The survey was done with 100 people. It can be inferred from the graph that our approach has performed better.

(a) Using MVG-MVS pipeline (b) With our framework

Fig. 5. Point cloud of Mahadeva Temple of Koppal.

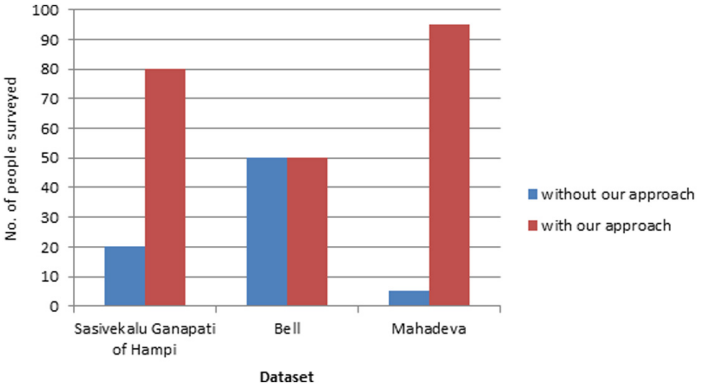

Fig. 6. Qualitative analysis

4 Conclusions

In this paper, we demonstrated the results in real-time for image selection for 3D reconstruction. We detected the similarity between images using two parameters and combined the parametric score of images being similar using DSCR (Dempster Shafer Combination Rule). We modeled 3D reconstruction as a combination of unique images. We have demonstrated the results of real-time heritage datasets. We have used the time taken for 3D reconstruction and the number of images selected for 3D reconstruction as a quantitative parameter and subjective analysis as a qualitative parameter to prove the superiority of the proposed method over the other algorithms.

References

1. Furukawa, Y., Curless, B., Seitz, S.M., Szeliski, R.: Towards internet-scale multi-view stereo. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1434–1441, June 2010. https://doi.org/10.1109/CVPR.2010.5539802
2. Hosseininaveh, A., Yazdan, R., Karami, A., Moradi, M., Ghorbani, F.: Clustering and selecting vantage images in a low-cost system for 3D reconstruction of texture-less objects. Measurement **99**, 185–191 (2016). https://doi.org/10.1016/j.measurement.2016.12.026
3. Kato, T., Shimizu, I., Pajdla, T.: Selecting image pairs for SFM on large scale dataset by introducing a novel set similarity. In: 2017 6th ICT International Student Project Conference (ICT-ISPC), pp. 1–4, May 2017. https://doi.org/10.1109/ICT-ISPC.2017.8075347
4. Repko, J., Pollefeys, M.: 3D models from extended uncalibrated video sequences: addressing key-frame selection and projective drift, pp. 150–157, July 2005. https://doi.org/10.1109/3DIM.2005.4
5. Thormählen, T., Broszio, H., Weissenfeld, A.: Keyframe selection for camera motion and structure estimation from multiple views. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 523–535. Springer, Heidelberg (2004). https://doi.org/10.1007/978-3-540-24670-1_40
6. Yang, C., Zhou, F., Bai, X.: 3D reconstruction through measure based image selection. In: 2013 Ninth International Conference on Computational Intelligence and Security, pp. 377–381, December 2013. https://doi.org/10.1109/CIS.2013.86
7. Tabib, R.A., Patil, U., Ganihar, S.A., Trivedi, N., Mudenagudi, U.: Decision fusion for robust horizon estimation using dempster shafer combination rule. In: 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), Jodhpur, pp. 1–4 (2013). https://doi.org/10.1109/NCVPRIPG.2013.6776247
8. Tabib, R.A., Patil, U., Naganandita, T., Gathani, V., Mudenagudi, U.: Dimensionality reduction using decision-based framework for classification: sky and ground. In: Sa, P.K., Sahoo, M.N., Murugappan, M., Wu, Y., Majhi, B. (eds.) Progress in Intelligent Computing Techniques: Theory, Practice, and Applications. AISC, vol. 519, pp. 289–298. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-3376-6_32
9. Patil, U., Tabib, R.A., Konin, C.M., Mudenagudi, U.: Evidence-based framework for multi-image super-resolution. In: Sa, P.K., Bakshi, S., Hatzilygeroudis, I.K., Sahoo, M.N. (eds.) Recent Findings in Intelligent Computing Techniques. AISC, vol. 709, pp. 413–423. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-8633-5_41
10. Moulon, P., Monasse, P., Perrot, R., Marlet, R.: OpenMVG: open multiple view geometry. In: Kerautret, B., Colom, M., Monasse, P. (eds.) RRPR 2016. LNCS, vol. 10214, pp. 60–74. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56414-2_5
11. OpenMVS. https://github.com/cdcseacave/openMVS
12. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
13. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-Up Robust Features (SURF). Comput. Vis. Image Underst. **110**(3), 346–359 (2008). https://doi.org/10.1016/j.cviu.2007.09.014

14. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: ORB: an efficient alternative to SIFT or SURF. In: Proceedings of the 2011 International Conference on Computer Vision (ICCV 2011), pp. 2564–2571. IEEE Computer Society, Washington, DC (2011). https://doi.org/10.1109/ICCV.2011.6126544
15. Leutenegger, S., Chli, M., Siegwart, R.Y.: BRISK: binary robust invariant scalable keypoints. In Proceedings of the 2011 International Conference on Computer Vision (ICCV 2011), pp. 2548–2555. IEEE Computer Society, Washington, DC (2011). https://doi.org/10.1109/ICCV.2011.6126542