視差マップを用いた 3D モデルの反復生成アプローチ

メヒルダド パナヒプル テヘラニ† 石川 彰夫† 酒澤 茂之† 小池 淳†

†超臨場感通信グループ、KDDI 研究所 〒105-0123 埼玉県ふじみ野市大原 2-1-15

E-mail: †{te-mehrdad, ao-ishikawa, sakazawa, koike}@kddilabs.jp

あらまし 本研究では、視差マップからの 3D モデル生成の問題に取り組む。まず、不正確な 3D モデルを与えて、密な多視点画像を生成する。各ペア間の視差マップはステレオマッチングアルゴリズムを用いて評価する。次に、視差マップを奥行き候補として空間へ射影する。ここで、地面からの候補の各レイヤについて奥行き候補にカーネル・クラシファイアを適用する。そして、さらなる微調整処理のためにより可能性の高い候補を選択する。微調整ステップでは、地面からの各レイヤの中の被写体の境界の最適な位置をダイナミック・プログラミングを用いて決定する。最終的に、3D モデルを生成する。生成された 3D モデルを用いて、密な多視点画像および視差マップを再生成する。生成される 3D モデルの変化が小さくなるまで、同じ処理を複数回反復して行う。実験の結果、少数の反復で、開始時の 3D モデルと比較して 3D モデルの質が高くなっていることが示された。

キーワード 3D モデル, 視差マップ, 多視点画像, カーネル密度関数, 分類, 微調整

An Iterative Approach for 3-D Model Generation Using Disparity Maps

Mehrdad Panahpour Tehrani† Akio Ishikawa† Shigeyuki Sakazawa† Atsushi Koike†

†Ultra Realistic Communications Laboratory, KDDI R&D Laboratories Inc.

2-1-15 Ohara, Fujimino, Saitama, 356-8502 Japan

E-mail: †{te-mehrdad, ao-ishikawa, sakazawa, koike}@kddilabs.jp

Abstract In this research, we address the problem of 3-D model generation from disparity maps. Given an inaccurate 3-D model, dense multiview images are generated. Disparity maps between each pair are estimated using stereo matching algorithm. The disparity maps are projected into space as depth candidates. The kernel classifier is applied to the depth candidates for each layer of candidates from ground, and the candidates with higher probability values are selected for further fine-tuning process. In fine-tuning step, the best location of the border of the objects in each layer from ground is determined using dynamic programming. Finally the 3-D model is generated. Using the generated 3-D model, we regenerate the dense multiview images and disparity maps. We perform the same process for several iterations until the changes in the generated 3-D model are small. Experimental result shows we can eventually enhance the quality of the 3-D model within a few iterations in comparison with the starting 3-D model.

Keyword 3-D model, disparity map, multiview images, kernel density function, classification, fine-tuning

1. Introduction

3-D visual communication, in which a 3-D visual content is used as a communication component, has been highly desired for a long time. For this reason, research on 3-D visual systems is emphasized, and multiview videos rendered by computer graphic algorithms have become more attractive and more efficient, such as FTV [1], and 3D-TV [2].

Image based rendering (IBR) technique allows generation of arbitrary views using ray-space [3] representation of multiview image. Using a multicamera system with known calibration parameters, ray-space data can be generated. In ray-space format, an image can be interpreted as light rays that are captured at a certain time and space. According to plenoptic sampling theory [4] IBR approach requires dense sampling of the scene for alias-free arbitrary viewpoint synthesis, which is costly and inefficient. Therefore, a sub-sampled representation of ray-space is used, in which the geometry compensation to generate a dense ray-space is done by using block matching algorithms [5]. The accuracy of the generated viewpoints using block-matching depends on the maximum disparity between viewpoints. Meanwhile, model based rendering (MBR) techniques can synthesis free viewpoint using sparse camera configuration. In MBR, 3-D model of each object is extracted [6],[7] and free viewpoint is rendered using the multicamera images and depth information of model. The quality of the rendered viewpoints is sensitive to accuracy of calibration and generated model.

In this research, we address the problem of 3-D model generation from disparity maps that satisfies the following goals.

1. Accuracy of model is independent of number of camera.
2. The proposed method is independent of object detection [8] quality.
3. This approach does not need any depth sensor to obtain precise model.
4. This method is robust to the accuracy of camera calibration.

In order to achieve the abovementioned goals, we propose an iterative approach to enhance the generated 3-D model using the multiview images generated from the initial model, and depth maps classification during a fine-tuning step. Evaluation of the proposed method is done using CG multiview images. Result shows significant improvement of 3-D model in comparison with initial model.

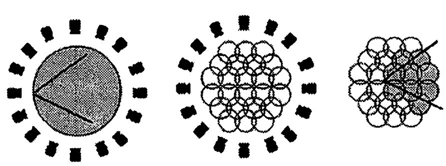


Fig. 1: Multiple Local ray-space representation. (a) local ray-spaces, (b) free viewpoint & walk-through.

2. Proposed Method

2.1. Multiview Generation

Given an inaccurate 3-D model (i.e. initial model), dense multiview images are generated. In our developed system, we generate the multiview images using multiple local ray-space method [9].

In this method, unlike conventional method with one ray-space [3] (Fig. 1a), the space is divided into multiple local ray-spaces, as shown in Fig. 1b for cameras on a circle, surrounding objects. In addition, each object is detected and 3-D depth information of each object is calculated. This helps to specify the occlusion, and makes the walk-through generation possible. For each local space, the sparse rays are recorded. Furthermore, given the depth information of objects, camera parameters, dense local ray-spaces are generated by interpolation. Given the location of required viewpoint in the space, free viewpoint can be generated. The corresponding local ray-spaces in the range of requested viewpoint are read, as shown in Fig. 1c.

2.2. Disparity Maps Generation

Disparity maps between each pair can be generated using any stereo matching algorithm [5], except occluded parts based on a constraint within matching algorithm. By discarding the occluded disparity, we can suppress wrong depth information, and guarantee the accuracy of the subsequent process.

2.3. Classification

2.3.1. Disparity maps to depth candidates

The disparity maps are projected into space as depth candidates, as shown in Fig. 2 for a pair of cameras and disparity map in between.

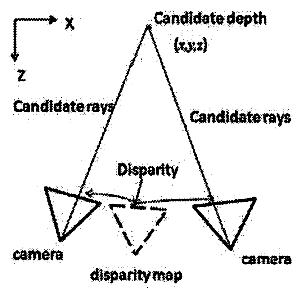


Fig. 2: Depth candidate in 3-D space from disparity map.

We formulate the depth candidates as a set of locations in 3-D space as follows.

$$\mathbf{P}\left(\mathbf{x},\mathbf{y},\mathbf{z}\right)=\left\{\left(x_0,y_0,z_0\right),\left(x_1,y_1,z_1\right),\ldots,\left(x_N,y_N,z_N\right)\right\} \quad (1)$$

2.3.2. Nonparametric modeling

In this step, we build a statistical representation of sub-sampled depth candidates of set "**P**" in the y direction as

$$\mathbf{P}_{\mathbf{y}_\mathbf{i}} = \mathbf{P}\left(\mathbf{x}, y_i \leq \mathbf{y} < y_j, \mathbf{z}\right) \quad (2)$$

The probability density functions (pdfs) are likely to vary for the depth candidates in set "P" for different locations and sub-sampled sets. Furthermore, they will not, in general, have a known parametric form. Thus, a general nonparametric kernel density estimation technique is accordingly used to build these statistical representations of set "P." Without any assumption about underlying distributions, this technique estimates the pdfs directly from the data.

We can use a variety of kernel functions with different properties. However, typically, for continuity, differentiability, and locality properties, Gaussian kernel is used. Note that fitting the distribution to a Gaussian model (normal distribution) differs from choosing Gaussian as a kernel function. Furthermore, kernel density estimation is a more general approach that does not assume any specific shape for density; unlike parametric fitting of a mixture of Gaussian. A discussion on kernel estimation techniques can be found in [10].

Given $\mathbf{P}_{yi}$ a 2-D kernel density function can be generated by the product of 1-D kernels [10]. Therefore, we can estimate $\hat{p}(x_n, z_n)$ density at point $(x_n, y_i \leq y < y_j, z_n)$ using

$$\hat{p}(x_n, z_n) = \frac{1}{M}\sum_{i=1}^{M} G(x_i, \sigma) G(z_i, \sigma) \tag{3}$$

where the same kernel function (Gaussian) is used with a suitable bandwidth σ for each dimension. Hence, we can represent the depth candidates by a 2-D Gaussian kernel density function.

2.3.3. Tuning

The generated probability model using kernel density function is later divided into two regions based on the probability value of the model. Areas in which are detected with probability more than a threshold are assigned as unknown as shown in Fig. 3a. The object border is located in the detected areas. Note that the thresholds are automatically calculated for each model using the distribution of the model, where normally has the ring shape. It is done by the fact that the probability values have lower value in the middle of the object.

2.4. Model Generation – Fine Tuning

In order to detect the border, as shown in Fig. 3a, the fine-tuning step is performed. In this process, the best location of the border (i.e. contour **C**) in each sub-sampled "P" is determined using a dynamic programming.

$$\mathbf{C} = \arg\min_t \left(\mathbf{E}(\mathbf{M}) + \lambda_1 \mathbf{E}(\mathbf{d}) + \lambda_2 \mathbf{E}(\Delta)\right) \tag{4}$$

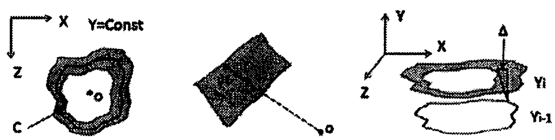
The energy functions for minimization are defined as

$$\mathbf{E}(\mathbf{M}) = \sum_n \hat{p}(x_n, z_n) \tag{5}$$

$$\mathbf{E}(\mathbf{d}) = \sum_n |d_1 - d_2|^2 \tag{6}$$

$$\mathbf{E}(\Delta) = \sum_n \Delta^2 \tag{7}$$

where **E(M)** the energy function corresponds to the generated nonparametric model. As shown in Fig. 3b and Fig. 3c, **E(d)** and **E(Δ)** are regulation energy functions, which are enforcing the contour **C** (i.e. edge of the object) aligns in the middle of the unknown area and does not deviate from previously calculated contour, respectively. The minimization can be also done using dynamic programming. Finally the 3-D model is generated. Values for λ_1 and λ_2 are experimentally set.



(a) (b) (c)

Fig. 3: Fine-tuning; (a) Unknown region and contour C, (b) Calculation of **E(d)**, (c) Calculation of **E(Δ)** .

2.5. Iteration

Using the generated 3-D model, we regenerate the dense multiview images. This increases the quality of multiview images. We perform the same process for several iterations until the change in the generated 3-D model is small.

3. Experiment

In order to evaluate the proposed method, we used CG generated multiview images.

Fig. 4a shows a view of the CG generated free viewpoint and its ground truth model. For this experiment, 30 multiview images on the circle are generated around the CG object. Object mask are manually detected for each viewpoint and 3-D model is generated [7]. Multiview images are generated using multiple local ray-space method [9]. Further, we generated one disparity map between each pair [11]. In the following, we applied our proposed iterative method to generate the 3-D model from disparity maps as it explained in section 2. Fig. 4b shows the rendered arbitrary viewpoint and the initial model. Fig. 4c and Fig. 4d show the same viewpoint and the generated 3-D model in two iterations using our proposed method. As it can be seen the quality of the synthesized viewpoint is significantly improves that verifies the accuracy of the generated 3-D model using our approach. In addition, Table 1 shows the accuracy of the generated model in different steps, where the error decreases during iterations.

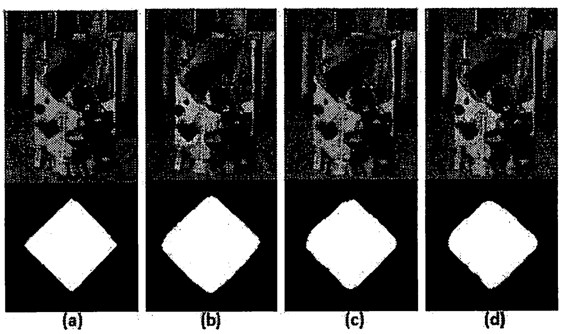


Fig. 4: Evaluation of our proposed method (a) CG generated free viewpoint and model (b) generated free viewpoint using the initial model, (c) and (d) generated free viewpoint and model in the first and the second iterations.

Table 1: Error in different stages in comparison with the ground-truth 3-D model

Data	**Original**	**Start**	**Itr. 1**	**Itr. 2**
Error (%)	0	20.94	4.88	0.61

4. Conclusion

In this paper, we proposed a novel method to generate a precise 3-D model using an iterative approach. Using the proposed method the accuracy of the generated model is significantly improves in comparison with initial model. Experimental results verify the effectiveness of the proposed method.

In our future research, we focus on improvement of a coarse 3-D model, as an initial model.

References

[1] P. Na Bangchang, M. Panahpour Tehrani, T. Fujii, M. Tanimoto, “Realtime System of Free Viewpoint Television,” *Journal of the Institute of Image Information and Television Engineers (ITE)*, 59 (8): 63–70, Aug. 2005I.

[2] W. Matusik, H. Pfister, "3-D TV: A Scalable System for Realtime Acquisition, Transmission, and Autostereoscopic Display of Dynamic Scenes," *Proc. SIGGRAPH, ACM,* Los Angeles, 2004, 23(3): 814–824.

[3] T. Fujii, T. Kimoto, and M. Tanimoto: “Ray Space Coding for 3D Visual Communication”, Proc. PCS’96, Picture Coding Symposium, pp. 447-451, March 1996.

[4] J.X. Chai, X. Tong, S.C. Chan and H.Y. Shum: “Plenoptic Sampling”, Computer Graphics (SIGGRAPH’00), pp. 307-318, July 2000.

[5] D. Scharstein and R. Szeliski, “Ataxnomy and Evaluation of dense two-frame stereo correspondence algorithm”, *International Journal of Computer Vision*, 47(1/2/3):7-42, April-June 2002.

[6] W. Niem and R. Buschmann, “Automatic Modelling of 3D Natural Objects from Multiple Images”, Proc. *European Workshop on Combined Real and Synthetic Image Processing for Broadcast and Video Production*, 1994.

[7] T.Matsuyama, X. Wu,T. Takai, and T. Wada, “Real-Time Dynamic 3D Object Shape Reconstruction and High-FidelityTexture Mapping for 3D Video,” *IEEE Trans. on Circuits and Systems for Video Technology*, Vol.CSVT-14, No.3, pp.357-369, 2004.

[8] Y. Satoh, and K. Sakaue: “Robust Background Subtraction based on Bi-polar Radial Reach Correlation,” TENCON 2005, pp. 1-6, IEEE Region 10, Nov. 2005.

[9] A. Ishikawa, M. Panahpour Tehrani, S. Naito, S. Sakazawa, and A. Koike, “Free viewpoint video generation for walk-through experience using image-based rendering”, *Proc. ACM Multimedia* 2008, Vancouver, Canada, Oct.-Nov.2008.

[10] D. W. Scott, *Multivariate Density Estimation.* New York: Wiley-Interscience, 1992.

[11] M. Droese, T. Fujii, M. Tanimoto: “Ray-Space Interpolation based on Filtering in Disparity Domain”, *Proc. of 3D Image Conference* 2004, Tokyo, Japan, pp. 29-30, 2004.

[12] M. P. Tehrani, A. Ishikawa, S. Sakazawa, A. Koike, “Enhanced Multiple Local Ray-spaces Method for Walk-through View Synthesis”, To be appeared in *Proc. of IEEE Computer Society, International Symposium Universal Communication,* ISUC 2008, Osaka, Japan, Dec. 2008.