Quality Improvement Of 3-D Voxel Models Based On Histograms Of Model Reprojection

Paweł Gardziński, Krzysztof Kowalak, Sławomir Maćkowiak, Łukasz Kamiński

Poznań University of Technology, Polanka 3, 60-965 Poznań Poland
{pgardzinski, kkowalak, smack, lkaminski }@multimedia.edu.pl

***Abstract* – This paper focuses on a method to improve the quality of the reconstruction of volumetric computer 3-D models calculated from multiple camera views. In particular, the idea of correcting some internal and external camera parameters, in order to obtain the estimated volumetric model with improved quality has been proposed. The authors proposed a solution which is built into volumetric reconstruction procedures. The proposed approach is based on an analysis of histograms of input images and their corresponding images of estimated model reprojection into the image planes. The proposed method utilizes cross-correlation to estimate corrections and characterizes itself as a non-optimization approach as opposed to the most common techniques.**

The effectiveness of the proposed approach has been tested on a widely used '*Temple*' model from *Middlebury dataset*. The authors have added a known error values to the camera parameters to test the approach. The experimental results prove high effectiveness of proposed approach.

***Keywords* – voxel reconstruction; camera calibration; histograms of model reprojection; quality improvement; cross-correlation**

I. INTRODUCTION

The reconstruction of photo-realistic 3-D models of real objects from a set of images or a video has become very popular in recent times. 3-D computer models of objects are key components in many 3-D multimedia systems. The quality of these models often has a large impact on the acceptability of output product e.g. the objects made by three-dimensional printers.

The most popular methods of the 3-D model objects reconstruction based on images are: techniques based on the estimation of depth maps from two or more views of the object [5,8], reconstructions based on the matching characteristic features between the images [9,10] and volumetric reconstruction techniques from multiple camera views [6,7]. The authors have focused on this third technique of the 3-D model reconstruction.

During the conducted research on the reconstruction of 3-D models of objects using volumetric reconstruction algorithm, the authors came across a problem with the accuracy of the internal and external camera parameters and the effect of these parameters on the quality of the estimated 3-D model. These camera parameters have a significant impact on the quality of the estimated model. Camera parameters given by the content creators: position, rotation and zoom of the camera, can have small errors, which leads to significant changes in the estimated model during the volumetric reconstruction algorithm. In many situations it is not possible to measure these parameters again more precisely, because the content (video or images) have already been created and delivered. These errors in camera parameters can be observed by comparing one of the input (reference) image with the plane image of the created 3-D model which is back projected to that particular view.

Commonly used method for determining the internal and external camera parameters between all views is a global parameters estimation technique - *bundle adjustment* algorithm. This is an optimization algorithm that requires a large set of local features, that are often generated by scale-invariant feature transform (SIFT [1]). Unfortunately, determination of these local features is related to the high computational complexity of the process. Moreover, it is not always possible to find a sufficient number of features for the algorithm to work properly. The bundle adjustment algorithm is an iterative algorithm that minimizes the global error of the camera parameters, using the Levenberg-Marquardt algorithm [2], which itself has a high computational complexity.

In this paper an approach that allows an easy adjustment of the camera parameters estimation has been proposed. The authors proposed the solution which is built into volumetric reconstruction procedures instead of global estimation of camera parameters. Simplicity of this solution lies in the use of histograms of images of 3-D model reprojection into image plane for determining the correction of the camera parameters of rotation and zoom of the camera. This method requires a preliminary camera parameters which may have a small error. Thus, the internal and external camera parameters do not need to be estimated again, as it is done repeatedly in the bundle adjustment algorithm. The proposed approach, compared to bundle adjustment algorithm is not iterative, making it possible to set the required camera parameters in a very short period of time, which is an advantage of the proposed method.

The solution is not intended to improve the actual camera parameters, but only to adjust the parameters of those input images that poorly match the model that is being reconstructed.

The paper is organized as follows. Section II presents the idea of camera parameters correction using image histograms. Firstly, an approach for correcting the rotation matrix camera parameters (expressed by the Euler angles: α, β and γ) is presented and then the correction of the focal length of the camera is presented. Section III presents the results of the experiments. Section IV provides a summary and a proposal for further work.

II. Histogram Based Correction

In our approach we utilize *Voxel Reconstruction*, also known as *Voxel Carving* algorithm that has already been described in [3], specifically we use an approach presented in [4]. Generally it is possible to use any other 3-D modelling technique with a requirement that it is able to generate a reprojection that correspond to the input image to be able to conduct histogram-based correction. Figure 1 shows parameterization of rotation matrix used in the proposed algorithm (usually included in extrinsic rotation parameters matrix).

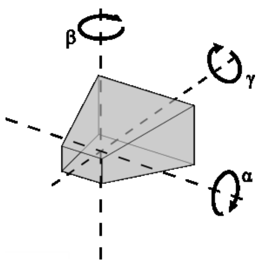

Figure 1. Rotation parameters designation in used frame of reference.

It is known that given error of a certain parameter will result in a known distortions in the model reprojection corresponding to the given input image. Figure 2 shows an example of resulting distortions in model reprojection (*Temple* – test dataset from [5]). Those distortions were best to see on the image histograms (calculated as a sum of white pixels either in rows or columns). One may already see that each parameter error is visible on their respective histograms e.g. an error in α angle is visible as a shift of vertical histogram (rotation invariant), an error in β angle is visible as a shift of horizontal histogram etc. The proposed methods to calculate each parameter correction are described in details in the following sections.

A. Correction of error in α and β angles

To calculate the $\Delta\alpha$ angle rotation of i-th input image (i-th view), firstly we calculate vertical histograms for both the model reprojection image and the i-th input image. Then, we calculate cross-correlation between both histograms which allows to calculate *gap* between them. Finally, we utilize the pinhole camera model to calculate the angle rotation that corresponds to *gap* from trigonometric equation:

$$\Delta\alpha = arctan(\frac{gap}{f}) \tag{1}$$

where f – focal length, *gap* – calculated shift of the histograms.

In case of the horizontal shift calculation that corresponds to a $\Delta\beta$ angle rotation of i-th input image we follow the steps described for $\Delta\alpha$ calculation but we use horizontal histograms instead.

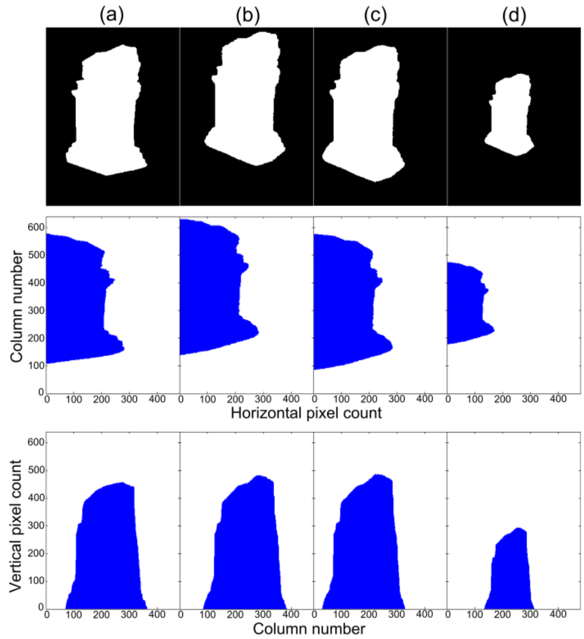

Figure 2. Visualization of known distortion of certain camera parameters. In columns: (a) – input image, (b) – reprojection with distorted α, (c) – reprojection with distorted β, (d) – reprojection with distorted f.

B. Correction of error in γ angle

To calculate the γ angle rotation of i-th input image we cannot use the histograms mentioned in point A, because

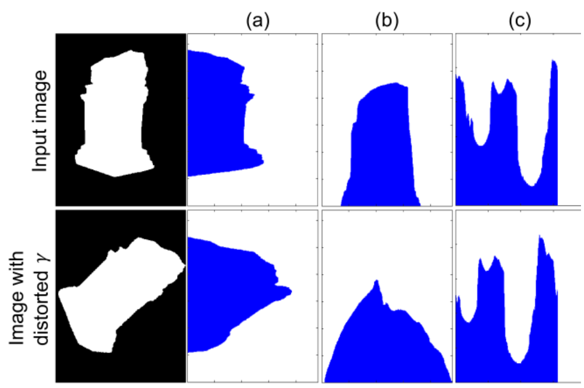

Figure 3. Visualization of histogram shape variation due to the γ rotation of i-th input image. (a) – vertical histograms, (b) – horizontal histograms, (c) – radial histograms depicted as a function Distance(angle[*rad*]).

as one may have already noticed the point distribution in lines has changed dramatically. Figure 3 shows the variation in histogram shape resulting from $\Delta\gamma$ rotation of input image that prevents the use of horizontal and vertical histograms in the algorithm .

This is why, a radial histogram has been used in the process of calculating the rotation $\Delta\gamma$ of *i*-th input image. As shown in Figure 3(c), the radial histogram is rotation invariant for γ angle and the rotation is seen as a shift. The radial histogram is calculated as a biggest distance of an object contour from the camera optical center given in pixel coordinates (usually denoted as a pair (C_x, C_y) in the intrinsic parameters matrix) in each direction from the optical center point (C_x, C_y). In our experiments we quantized the directions into 360 intervals.

At this point the procedure of calculating the $\Delta\gamma$ rotation of *i*-th input image is conducted in a similar way as described in section *A*. We compute the cross-correlation and *gap*, which describes the quantized rotation $\Delta\gamma$.

C. Correction of focal length

In order to calculate focal length correction Δf we need to follow a two-step approach, which has been described below.

Firstly, we calculate the radial histograms for both the input images and images of 3-D model reprojection into image plane (with input focal length f_1). Then, we calculate their trimmed means y_1, y_3 (also known as *truncated mean*) in order to estimate an average pixel distance $|y_1 - y_3|$ between the histograms. It is necessary, because the correction of parameters is often conducted during the reconstruction process therefore there are some voxels that has not been carved yet and they add an overflow value into the reprojection image histogram. The calculated trimmed means y_1, y_3 are shown on Figure 4, where a pinhole camera model with a highlight of image plane Y and optical axis F has been presented.

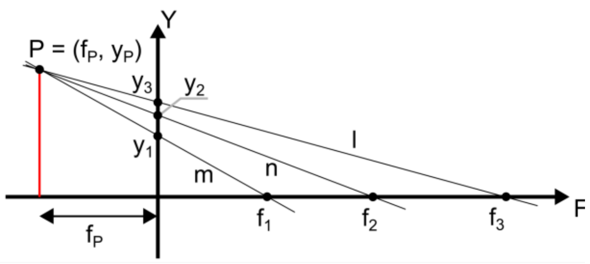

Figure 4. Visualization of 3 reprojection lines m, n, l that cast the point P into image plane with respect to 3 focal lengths f_1, f_2, f_3. Visualization of system of equations to calculate corrected focal length f_3.

Secondly, we calculate Δf by estimating the focal length f_3 according to the following equations:

$$\begin{cases} m: y_P = \dfrac{y_1}{f_1} f_P + y_1 \\ n: y_P = \dfrac{y_2}{f_2} f_P + y_2 \end{cases} \tag{2}$$

Using (2) we derive an equation for the position of P:

$$f_P = \frac{(y_1 - y_2) f_1 f_2}{(y_1 f_2 - y_2 f_1)} \tag{3}$$

where

$$f_2 = f_1 * \frac{y_1 + y_3}{2y_1} \tag{4}$$

Equation (4) describes a sampling focal length f_2 used to generate a help model reprojection, for which we calculate a trimmed mean y_2 over its radial histogram. Which is used in (5)

$$l: 0 = \frac{y_P - y_3}{f_P} f_3 + y_3 \tag{5}$$

to derive a final equation for desired focal length f_3:

$$f_3 = \frac{-y_3 f_P}{y_P - y_3} \tag{6}$$

III. EXPERIMENTS

The experiments has been conducted on a test model *Temple* from the Middlebury dataset [5]. The results have been averaged over all 47 views of aforementioned model. Each of the parameters error correction (rotation parameters and focal length) has been tested separately.

Firstly, an undistorted input parameters were taken and some of them were modified by a known value. Then a voxel reconstruction has been calculated in order to generate a distorted model. In the next step, an approach described in section 2 has been utilized to estimate parameters correction. Finally, the parameters correction was evaluated by comparison of the model reprojection image with the reference image (input image). They were used to calculate the following two metrics:

$$match = \frac{mutual\_area}{reference\_area} \tag{7}$$

$$overflow = \frac{reproj\_area}{reproj\_area + mutual\_area} \tag{8}$$

where $reference\_area$ is an area of the model in the reference image, $mutual\_area$ is a mutual area of both reference and reprojection image and $reproj\_area$ is an area of the reprojection image that is not mutual with reference image. The aim of the correction is to reach $match = 1$ and $overflow = 0$.

It is worth to mention that these two metrics are introduced, because the commonly used method of comparing the reconstructed model with a reference model (e.g. the one provided along with the input images from Middlebury dataset) does not work before reaching the end of reconstruction process.

TABLE I. EXPERIMENTAL RESULTS FOR SELECTED PARAMETERS.

Parameter	Before correction		After correction	
	match	***overflow***	***match***	***overflow***
$\alpha = \alpha - 0.5°$	0.9847	0.0925	0.9957	0.0822
$\alpha = \alpha + 0.5°$	0.9663	0.1093	0.9951	0.0828
$\alpha = \alpha - 1°$	0.9543	0.1208	0.9953	0.0824
$\alpha = \alpha + 1°$	0.9244	0.1478	0.9944	0.0833
$\alpha = \alpha - 2°$	0.8757	0.194	0.9959	0.0422
$\alpha = \alpha + 2°$	0.8465	0.2203	0.9938	0.0841
$\beta = \beta - 0.5°$	0.9903	0.0865	0.999	0.0787
$\beta = \beta + 0.5°$	0.8687	0.1992	0.999	0.0792
$\beta = \beta - 1°$	0.9862	0.0913	0.999	0.0793
$\beta = \beta + 1°$	0.8035	0.2588	0.9989	0.0782
$\beta = \beta - 2°$	0.8653	0.2028	0.9981	0.0391
$\beta = \beta + 2°$	0.678	0.3753	0.999	0.079
$\gamma = \gamma - 5°$	0.968	0.1075	0.9969	0.0806
$\gamma = \gamma + 5°$	0.9744	0.1015	0.99	0.0871
$\gamma = \gamma - 10°$	0.9318	0.1409	0.9969	0.0806
$\gamma = \gamma + 10°$	0.9342	0.1387	0.99	0.0871
$\gamma = \gamma - 15°$	0.8941	0.1757	0.9969	0.0806
$\gamma = \gamma + 15°$	0.8933	0.1865	0.99	0.0873
$f = f - 100$	0.9312	0.0161	0.9847	0.0397
$f = f + 100$	0.9987	0.1892	0.9834	0.0384
$f = f - 200$	0.8165	0.0011	0.9815	0.037
$f = f + 200$	0.9925	0.2844	0.939	0.1497
$f = f - 300$	0.6985	0.0006	0.9809	0.0362
$f = f + 300$	0.9989	0.3575	0.9819	0.0372

Reaching the ideal value of $match = 1$ is usually impossible due to the fact, that there is a limited number of voxels in a given area which is lower than the image resolution. It means that a single voxel reprojection image would consist of more than one pixel. It can be interpreted as a 3 dimensional quantization error. On the other hand, the ideal value of $overflow = 0$ is usually unreachable, because it is calculated throughout the reconstruction process while there are still some *uncarved* voxels as it was described in section *C*. However, the $overflow$ value drops with each input image added to the carving process.

Finally, it is worth to mention that the approach presented in this paper works for relatively small angle errors. As one may imagine a shift of $90°$ in α would result in shifting the reprojected model out of reprojection image, so the solution works until the object “hits” the image boundary.

IV. CONCLUSIONS

In this paper, authors showed an approach for improvement of a 3-D voxel model quality based on histograms of model reprojection. During the research, the authors noticed the voxel reconstruction’s susceptibility to even the smallest parameters variation and have examined their influence on the resulting model.

Based on the experimental results, the authors proposed a non-optimization approach to improve the voxel reconstruction quality. The approach utilizes an information that the input data for volumetric reconstruction may contain some minor errors, that can be corrected without the need of using complex methods. The approach has been evaluated using the match and overflow metrics, that showed its high effectiveness.

Moreover, an additional experiments concerning combinations of errors (e.g. simultaneous α and β error) have been carried out. Firstly, they proved that the proposed histograms’ character for their respective angles remain invariant from rotation, while retaining information about the rotation itself as a shift. Secondly, combinations of individual rotation errors shown in Table 1 were successfully corrected. However, the research concerning all possible combinations requires more work and a separate analysis. Lastly, the additional problem of correcting the transition matrix parameters (X, Y, Z) is even more complex as it leads to perspective changes in the reprojection image and requires more research as well.

V. ACKNOWLEDGMENT

This work has been supported by the public funds as a DS research project for The Faculty of Electronics and Telecommunications (FET) at Poznan University of Technology.

REFERENCES

[1] D.G. Lowe, "Object recognition from local scale-invariant features", *Proceedings of the International Conference on Computer Vision,* pp. 1150–1157, 1999.

[2] K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares", *Quarterly of Applied Mathematics*: pp. 164–168, 1944.

[3] A.O. Balan, “Voxel Carving and Coloring – Constructing a 3D Model of an Object from 2D Images”, Brown University, December 2003.

[4] S. Maćkowiak, P. Gardziński, Ł. Kamiński, K. Kowalak, "Human Activity Recognition in Multiview Video", *11th IEEE International Conference on Advanced Video and Signal-Based Surveillance*, South Korea, pp. 148-153, 2014.

[5] S.M. Seitz, B. Curless, J. Diebel, D. Scharstein, R. Szeliski, “A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms”, *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 519-528, 2006.

[6] B. Goldluckeand, M. Magnor, "Real-time, free-viewpoint video rendering from volumetric geometry", *Proc. Visual Computation and Image Processing (VCIP)*, Lugano, Switzerland, 5150(2): pp. 1152-1158, 2003.

[7] H. Baker, "Three-dimensional modelling", *Proc. of the 5th International Joint Conference on Artificial Intelligence*, Cambridge, MA, USA, pp. 649-655, 1977.

[8] J. Xiao, J. Chen, D.-Y. Yeung and L. Quan, “Learning twoview stereo matching”, *In Proceedings of the 10th European Conference on Computer Vision: Part III, ECCV ’08*, Springer-Verlag, Berlin, Heidelberg, pp. 15–27, 2008.

[9] N. Snavely, S. Seitz, and R. Szeliski, "Photo tourism: exploring photo collections in 3d", *In ACM Transactions on Graphics (TOG)*, volume 25, ACM, pp. 835–846, 2006.

[10] N. Snavely, S. Seitz, and R. Szeliski, "Modeling the world from internet photo collections", *International Journal of Computer Vision*, 80(2): pp. 189–210, 2008.