# On the Performance Evaluation of 3D Reconstruction Techniques from a Sequence of Images

**Ahmed Eid**

*Computer Vision and Image Processing (CVIP) Laboratory, University of Louisville, Louisville, KY 40292, USA*
*Email: eid@cvip.uofl.edu*

**Aly Farag**

*Computer Vision and Image Processing (CVIP) Laboratory, University of Louisville, Louisville, KY 40292, USA*
*Email: farag@cvip.uofl.edu*

The performance evaluation of 3D reconstruction techniques is not a simple problem to solve. This is not only due to the increased dimensionality of the problem but also due to the lack of standardized and widely accepted testing methodologies. This paper presents a unified framework for the performance evaluation of different 3D reconstruction techniques. This framework includes a general problem formalization, different measuring criteria, and a classification method as a first step in standardizing the evaluation process. Performance characterization of two standard 3D reconstruction techniques, stereo and space carving, is also presented. The evaluation is performed on the same data set using an image reprojection testing methodology to reduce the dimensionality of the evaluation domain. Also, different measuring strategies are presented and applied to the stereo and space carving techniques. These measuring strategies have shown consistent results in quantifying the performance of these techniques. Additional experiments are performed on the space carving technique to study the effect of the number of input images and the camera pose on its performance.

**Keywords and phrases:** evaluation, 3D reconstruction, reprojection, quality assessment, stereo, space carving.

## 1. INTRODUCTION

The design of experimental test beds and methods of analysis in addition to the definition of the ground truth are important components of any performance evaluation system in computer vision. Furthermore, the performance must be quantified to make valuable objective comparisons. Otherwise, the failure of assessing the performance may not only cause unnecessary complexities in the subsequent processes but also may lead to the inability to fulfill the requirements of the application under concern. This reflects the need of a general methodology for testing the performance of similar computer vision techniques. Particularly, in this paper, we address the problem of performance evaluation of 3D reconstruction techniques from a sequence of images.

In fact, this area of research still faces a lack of standardized and widely accepted methods. Yet, the seminal works of Szeliski and Zabih [1, 2] are leading examples in the sense of presenting new metrics and methodologies of performance evaluation of stereo and motion techniques. Another related work [3] has presented a design of an experimental setup for performance evaluation of stereo techniques in telepresence. This setup uses a 3D scanner for providing the ground truth for performance evaluation of stereo techniques.

In this paper, we extend the evaluation domain to include different 3D reconstruction techniques based on common reference, common methodologies, and common quantification measures.

Since, stereo and space carving [4] are standard 3D reconstruction techniques, we group them under common testing methodology to examine their performance [5]. In our research laboratory, we have developed a testing setup that lends itself for generating calibrated sequences of images and concurrently generates a reference 3D model using a 3D laser scanner. This gives the possibility of using the scanner output as ground truth to the evaluation process. However, in this paper, we chose to reduce the dimensionality of the evaluation problem by considering the generated images as ground truth. Of course, it is desirable to test the given 3D data in its original domain provided that the 3D reference data are generated in error-free form [6]. However, this may not be achievable in most cases. For example, the projection of laser in 3D laser scanners is not guaranteed on all surfaces such as hairlike surfaces. In addition, self-occluding objects could not be well reconstructed [7]. Therefore, additional algorithms are necessary to compensate for errors in the reference data.

The testing strategy that we use depends on reprojecting the generated 3D model to similar views of input images, then comparing these reprojections to the input images. The idea of reprojection is not new [3, 8], however, we use this methodology under unified framework for performance evaluation of different 3D reconstruction techniques. This framework includes a general problem formalization, different measuring criteria, and a classification criterion as a first step in standardizing the evaluation process.

To quantify the performance, we used different image quality measures with different assessment philosophies. The dynamic ranges of these measures are modified to facilitate the comparison process.

The rest of this paper is organized as follows. Section 2 describes the data acquisition system we use to generate the input data. Section 3 presents the performance evaluation framework and describes the testing methodology and the quality measures. Section 4 provides the experimental results and Section 5 provides the conclusion and the future work.

## 2. DATA ACQUISITION SETUP

The experimental setup consists of a 3D laser scanner and a CCD camera mounted on a metal arm of multiple joints that is attached to the scanner head. A monocolor, usually black, screen is attached to the scanner head facing the CCD camera such that the screen appears as a fixed background to the object under reconstruction. The structure of the monocolor screen as well as the motion mechanism of the scanner facilitate the object segmentation task. The shaft over which the scanner head is mounted is controlled in terms of speed and angle of rotation to capture a number of $N_I$ images $I_0, I_2, \ldots, I_{N_I-1}$ at specific locations on a circular path. Meanwhile, the scanner generates a 3D reference model.

The calibration process [9, 10] attempts to estimate the camera parameters at each point in the circular path where the images are acquired. The camera is calibrated at the initial position to determine the projection matrix $\mathbf{P_0}$. Assuming that the camera rotates by a step-rotation angle $\alpha$, the projection matrix can be determined at each position as

$$\mathbf{P}_k = \mathbf{P_0} \begin{bmatrix} \cos(k\alpha) & 0 & \sin(k\alpha) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(k\alpha) & 0 & \cos(k\alpha) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad (1)$$

where $k = 1, 2, \ldots, N_I - 1$. As a result, a sequence of calibrated images is generated. These images are used as inputs to the vision technique under test.

Since the rotation angle of the scanner head is preassumed, subsequent calibration errors could result if the actual rotation of the head did not follow the preassumed rotation. As a result, we put an upper bound on the rotation angle error. The system setup has to be accurate to the limit of this upper bound, otherwise, subsequent errors could affect the accuracy of the evaluation process. This upper limit is a function of the calibration parameters, the rotation angle,

and the 3D coordinates of a given 3D reconstruction [11]. To ensure errors less than $\pm 0.5$ pixel in reprojected images, our experimental results show that an error up to 0.2 rad between the actual rotation and the preassumed rotation is permissible since most commercial 3D laser scanners can achieve 0.1 rad accuracy. For more details about the system, we refer the reader to [11].

## 3. PERFORMANCE EVALUATION FRAMEWORK

In this section, a general formalization of the evaluation problem is presented, followed by a proposed classification of evaluation techniques. In addition, a performance evaluation methodology for 3D reconstruction techniques is described.

### 3.1. Problem formalization

Formally, we want to solve the following problem: given (i) a set $\mathcal{M} \subset \mathbb{R}^3$ of 3D data points of an object generated by a 3D reconstruction technique $\mathbf{X}$ and (ii) a set $\mathcal{G}$ of ground truth data points of the same object, quantify the performance of technique $\mathbf{X}$.

In general, to solve this problem, three main components should be available: (i) an experimental test bed for collecting data, (ii) preevaluation techniques for preparing data for the evaluation process with minimal undesirable effects on the given data, and (iii) a performance evaluation methodology and a measuring criterion. If the ground truth data points are 3D points, then a pre-evaluation 3D registration function or transformation $\mathcal{F}$ is required to minimize the energy function $E$ defined as

$$E = \sum_i d_E^2(m_i \in \mathcal{M}, \mathcal{F}(g_i \in \mathcal{G})), \quad (2)$$

where $d_E$ denotes the Euclidean distance. In the Euclidean space, $\mathcal{F}$ has 6 degrees of freedom (DOF): three for rotation and three for translation. Otherwise, if ground truth data set is a set of 2D points or 1D parameters, then a new data set $\mathcal{D}$ derived from the given data set $\mathcal{M}$ is needed to be matched with $\mathcal{G}$. The data set $\mathcal{D}$ is derived from the given data set $\mathcal{M}$ according to the transformation or criterion $\mathcal{C}$ as

$$\mathcal{D} = \{d : d = \mathcal{C}(m \in \mathcal{M}), d \in \mathbb{R}^n, n \in \{1, 2\}\}. \quad (3)$$

An example of the criterion $\mathcal{C}$ is the camera projection matrix that transforms 3D data points into 2D image points.

Considering the evaluation problem as a matching problem [12], we can define the error criterion $\mathcal{E}(\mathcal{T}, d, g)$ as

$$\mathcal{E}(\mathcal{T}, d, g) = \begin{cases} 0 & \text{if } d \text{ matches } g, \\ 1 & \text{otherwise,} \end{cases} \quad (4)$$

where $\mathcal{T} : \mathcal{G} \to \mathcal{D}$ is a matching criterion. The error ratio $(E_r)$ is defined as

$$E_r = \frac{\sum_{i=0}^{\text{card}(\mathcal{E})-1} \mathcal{E}(\mathcal{T}, d_i, g_i)}{\text{card}(\mathcal{E}(\mathcal{T}, d, g))}, \quad (5)$$

where card($\mathcal{E}$) is the cardinality of $\mathcal{E}$. The matching criterion $\mathcal{T}$ could be one-to-one mapping or many-to-one mapping. An example of the later case is the closest point criterion where a point in the measured data could be matched to many points in the ground truth data because the measured point could be the closest point to such ground truth points. A one-to-many matching case could happen if the cardinality of the measured data is greater than the cardinality of the ground truth data. In this case, false matching and hence nonzero errors are expected. The following proposition states this case.

**Proposition 1.** *If* card($\mathcal{D}$) > card($\mathcal{G}$) *and if* $\mathcal{T}$ *exists such that* $\mathcal{T} : \mathcal{G} \to \mathcal{D}$*, then* $E_r \neq 0$.

*Proof.* Since card($\mathcal{D}$) > card($\mathcal{G}$), then $\mathcal{T}$ is not an injective mapping, then at least one point $g_i \in \mathcal{G}$ is matched to a subset $\mathcal{D}_j \subset \mathcal{D}$ where card($\mathcal{D}_j$) > 1, then false matching is guaranteed, hence the nonzero error result. $\square$

The above proposition provides a preliminary evaluation method since it expects errors if the cardinality of the data under test is greater than the cardinality of the ground truth data.

### 3.2. Classification of evaluation techniques

Seeking for a standardization to the evaluation problem components, we propose a classification criterion based on which we can classify different tests that measure the performance of 3D reconstruction techniques. Based on this classification, it will be easy to qualify new proposed tests and identify the goals and benefits of applying such tests to vision techniques. In addition, the classification could lead to a clue about the importance of applying these tests.

The proposed classification is based on four sets: the operating conditions set **A**, the complexity of data analysis set **B**, generality of measures set **C**, and the position of the test point set **D**.

#### Operating conditions set

Based on the operating conditions, we can identify two types of tests.

(i) *Dynamic tests:* in this type, the test is performed under different conditions of lighting, interference, calibration, and object complexity. These tests should measure the immunity of the vision technique to variations.

(ii) *Static tests:* in this type, the test is performed under constant conditions. Actually these tests investigate the basic functionality of the vision technique.

#### Complexity of data analysis set

Tests could be quantitative or qualitative.

(i) *Quantitative tests:* massive data are analyzed by these tests, statistical analysis can be a part of these tests. A test is said to be quantitative if the data set under test $\mathcal{M}_j$, where $\mathcal{M}_j \subset \mathcal{M}$, has cardinality > $\beta$ card($\mathcal{M}$) where $\beta$ > 0.5.

(ii) *Qualitative tests:* the objective of these tests is to provide a quick figure of merit of the performance of the vision technique under test. In this case, $\mathcal{M}_j$ has cardinality < $\gamma$ card($\mathcal{M}$) where $\gamma$ < 0.5.

#### Generality of the measure set

Measures could be global or local, hence we have two type of tests.

(i) *Global tests:* these tests provide a single measure of the overall performance of the vision technique under test. Such types of tests are of great importance because they give a final decision on the technique's performance.

(ii) *Local tests:* these tests investigate the local errors provided by the vision technique. Using local measures provided by the test, enhancement of the technique's performance could be possible.

#### Position of the test point set

Data can be tested in a form of 3D data, a form that results after applying a certain transformation to the 3D data, or a form that requires a certain transformation or criterion to get the 3D data form. Based on this forms, we have three types of tests

(i) *Type I tests:* these tests are applied directly to the data set $\mathcal{M}$. This means that transformation $\mathcal{C}$ is unity. These tests are highly trusted because they work directly on 3D data sets avoiding errors introduced by such transformations.

(ii) *Type $I^+$:* unlike type $I$, these tests are applied to the data set $\mathcal{D}$ generated by applying the transformation $\mathcal{C}$ to data set $\mathcal{M}$. Errors should be predicted due to this additional transformation step. As a result, these tests may underestimate the performance of the given technique under test. An example of this type is testing the data in the form of 2D intensity images.

(iii) *Type $I^-$:* like type $I^+$, these tests are applied to the measured data set, however, a step before getting the data set $\mathcal{M}$. Overestimation of the performance is predicted when using this type of tests because we test data in a form prior to the 3D form. An example of this type is testing the data in the form of disparity maps, the form of data that needs further transformation or criterion to get the 3D data form.

Based on the preceding classification a number of

$$\text{card}(\mathbf{A}) \times \text{card}(\mathbf{B}) \times \text{card}(\mathbf{C}) \times \text{card}(\mathbf{D}) \tag{6}$$

different tests can be accomplished under this classification. The next proposition generalizes the above formula.

**Proposition 2.** *For disjoint test sets* $\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_k$*, there exists*

$$\text{card}(\mathcal{X}_1) \times \text{card}(\mathcal{X}_2) \times \cdots \times \text{card}(\mathcal{X}_k) \text{ number of tests.} \tag{7}$$

*Proof.* It is a generalization to the above formula. □

According to the above classification, there are $2 \times 2 \times 2 \times 3 = 24$ different types of tests.

### 3.3. Image reprojection testing methodology and measuring criteria

Different performance evaluation techniques could be proposed to investigate the performance of a given 3D reconstruction technique. In our previous work [6], we introduced a technique for local quality assessment of 3D reconstructions. Using this technique, we analyze the performance in the 3D domain with ground truth data being generated using the setup presented in Section 2. This type of test is classified as quantitative, dynamic, local, and type $I$. Since type $I$ tests manipulate 3D data and assume the availability of 3D ground truth data and accurate methods for 3D data registration [13], their use could be limited. Types $I^+$ and $I^-$ tests could be alternatives to avoid such difficulties related to type $I$ tests.

Here we introduce an example of type $I^+$ tests, the image reprojection test. This test uses calibrated sequence of images captured for the object under reconstruction as ground truth data. A corresponding sequence of images is generated by reprojecting the 3D reconstruction under test to the same views as of ground truth images. Different measuring criteria are presented to quantify the similarities/dissimilarities between such sets of images. Three measures are presented: the signal-to-noise ratio (SNR), the inverse fuzzy image metric (IFIM), and the modified image quality index ($Q_m$).

#### Signal-to-noise ratio

SNR is a mean-squared ($l_2$-norm) error measure [14]. SNR is defined as the ratio of average signal power to average noise power. Assuming that $\mathcal{G}$ and $\mathcal{D}$ are two $M \times N$ images representing a ground truth image and a data image, respectively, the signal-to-noise ratio (SNR) is defined as

$$\text{SNR(dB)} = 10 \log_{10} \left( \frac{\sum_{i,j} g(i,j)^2}{\sum_{i,j} \left( g(i,j) - d(i,j) \right)^2} \right) \quad (8)$$

for $0 \le i \le M - 1$ and $0 \le j \le N - 1$, where $g(i,j)$ denotes the intensity of pixel $(i,j)$ of the standard image and $d(i,j)$ denotes the intensity of pixel $(i,j)$ of the data image.

#### Inverse fuzzy image metric

A fuzzy image metric (FIM) [15], defined based on Sugeno's fuzzy integral, can be used as a quality metric instead of SNR measure. While the SNR measure is commonly used in evaluating image quality to certain extent, it fails to be consistent with human visual perception. On the other hand, the fuzzy image metric (FIM) has the ability to reflect the human visual perception [15].

The two images $\mathcal{D}$ and $\mathcal{G}$ can be written as 1D arrays denoted by

$$\mathcal{D} = (d_1, d_2, \ldots, d_{M \times N}), \qquad \mathcal{G} = (g_1, g_2, \ldots, g_{M \times N}), \quad (9)$$

where $0 \le g_i, d_i \le 1$ (after normalization).

The FIM is defined as

$$\text{FIM} = \max_{0 \le i \le 255} \min \left( \frac{i}{255}, \mu \left( N_{i/255} \left( |\mathcal{G} - \mathcal{D}| \right) \right) \right), \quad (10)$$

where $\mu(\{\cdot\}) = \text{card}(\{\cdot\})/M \times N$ and $N_l(f) = \{x \mid f(x) \ge l\}$.

The FIM measure has a dynamic range of $[0, 1]$ where the smaller values indicate high-quality images. We propose the inverse fuzzy image measure (IFIM) which has better dynamic range than FIM for better comparison with SNR measure. The IFIM is defined as follows:

$$\text{IFIM} = 10 \log \frac{1}{\text{FIM}}. \quad (11)$$

The IFIM has a lowest value of 0 dB with higher values indicating better quality images.

#### Image quality index measure

Wang et al. [16] proposed a quality index which models image degradation as structural distortion instead of errors. This quality index is defined as

$$Q = \frac{4\sigma_{gd}(\overline{g})(\overline{d})}{(\sigma_g{}^2 + \sigma_d{}^2)((\overline{g})^2 + (\overline{d})^2)}, \quad (12)$$

where

(i) $\overline{g}$ and $\overline{d}$ are the means of $\mathcal{G}$ and $\mathcal{D}$ images, respectively,

(ii) $\sigma_g{}^2$ and $\sigma_d{}^2$ are the variances of $\mathcal{G}$ and $\mathcal{D}$ images, respectively,

(iii) $\sigma_{gd}$ is covariance between $\mathcal{G}$ and $\mathcal{D}$ images.

The dynamic range of $Q$ is $[-1, 1]$. The best value 1 is achieved if and only if $g_i = d_i$ for $i = 1, 2, \ldots, M \times N$. This quality index models any distortion as a combination of three different factors: loss of correlation, mean distortion, and variance distortion. To be consistent with above measures, the $Q$ measure is modified as

$$Q_m = 10 \log(2 + Q), \quad (13)$$

where $Q_m$ is the modified quality index in decibels.

The above measures are applied to space carving and stereo reconstructions to quantify their performance based on the reprojection methodology. The procedure of the reprojection test is outlined as follows:

(i) apply the vision algorithm under test to the acquired sequence of images, the ground truth sets $\mathcal{D}$, to generate the 3D data set $\mathcal{M}$,

(ii) apply (1) to generate sets $\mathcal{D}$ of reprojected images,

(iii) apply (8), (11), or (13).

Based on the classification presented in the previous section, the reprojection test is a quantitative, dynamic, global, and type $I^+$ test.
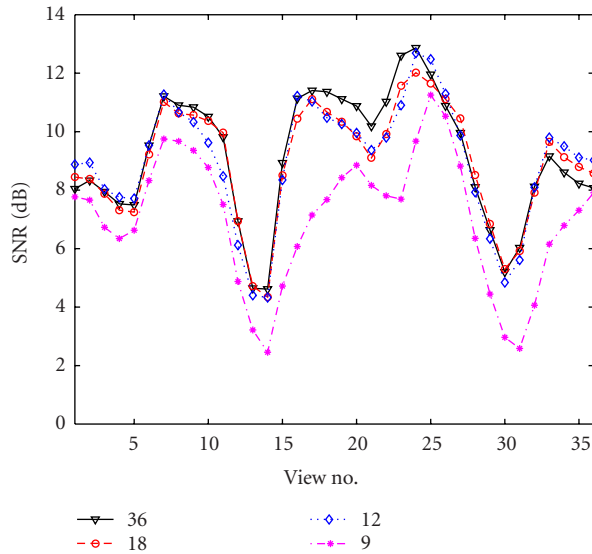
FIGURE 1: The SNR values at different views when the number of input images to the space carving is changed.

TABLE 1: The effect of the number of input images on the performance of space carving.

| No. of input images | Mean (dB) | No. of voxels |
|---|---|---|
| 36 | 9.190 | 81 535 |
| 18 | 9.012 | 83 233 |
| 12 | 9.000 | 83 076 |
| 9 | 7.069 | 87 591 |

## 4. EXPERIMENTAL RESULTS

In this section, we introduce examples of performance evaluation of two common 3D reconstruction techniques: space carving and stereo. First, the performance of space carving is examined when the number of input images and the camera pose are changed. Second, we provide a preliminary comparison of stereo and space carving techniques using different measuring criteria.

### 4.1. Performance evaluation of space carving

In this section, we study the effect of the number of input images on the performance of space carving approach. A number of 36 ($\alpha = 10°$) images are acquired for a house object. The space caving approach is applied to the acquired set of input images. The image reprojection testing criterion is applied to the output reconstruction, then the SNR is computed at each view. The experiment is repeated with different numbers of input images: 18, 12, and 9.

Figure 1 shows the SNR values for each view for different numbers of input images. As shown in this figure, the SNR values are almost the same for 36, 18, and 12 cases, however, they are lower in the case of 9 images. Table 1 shows the arithmetic mean of SNR values in each case. It shows that the mean of the 9-input case has the lowest value among the other cases.
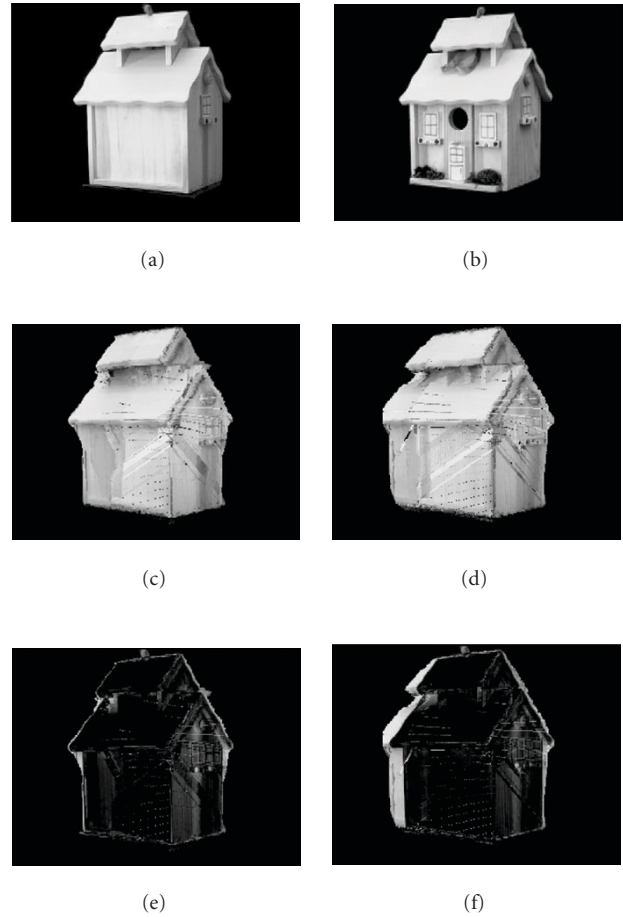


FIGURE 2: (a) and (b) Two input images of the house object. Reprojected images of a 3D reconstruction by space carving given (c) 12 and (d) 9 input images. Difference images between the reprojections of 3D reconstruction and the input images at the same view given (e) 12 and (f) 9 input images. The fattening effect is clearly manifested in (f).

It can also be noted that the number of voxels in the output reconstruction of the 9-input case is much higher than the other cases. This indicates that 9-input reconstruction is much fatter than the others, hence the degradation in the SNR values. Figure 2 shows visual results of this fattening problem. Figures 2a and 2b show two input images. Two reprojected images at the same view as in Figure 2a are shown in Figures 2c and 2d for the 12-input and 9-input cases, respectively. These reprojected images are subtracted from the original image in Figure 2a. The difference images are shown in Figures 2e and 2f for 12- and 9-input images cases, respectively. As shown in Figure 2f, the 9-input reconstruction is fatter than the 12-input case. Actually, this is an inherent problem in the space carving approach when the reconstructed scene has large homogenous (same intensity) areas. To overcome this fattening problem, the number of input images should be increased to put additional constraints on the reconstructed shape.
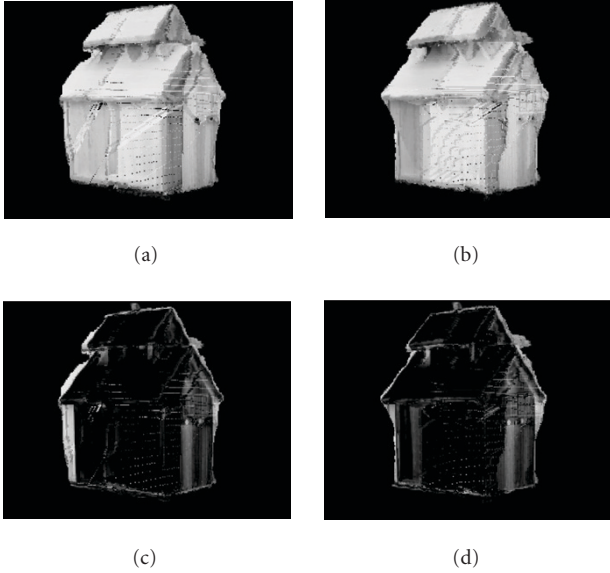
(a)          (b)

(c)          (d)

FIGURE 3: Reprojected images of a 3D reconstruction by space carving given 9 input images of (a) set $S_2$ and (b) set $S_4$. Difference images between the reprojections of 3D reconstruction and the input images at the same view given (c) set $S_2$ and (d) set $S_4$. The fattening effect is reduced as shown in (d).

TABLE 2: The effect of camera pose on the performance of space carving.

| Set | $S_1$ | $S_2$ | $S_3$ | $S_4$ |
|---|---|---|---|---|
| Mean (dB) | 7.0691 | 7.1498 | 7.6933 | 8.1108 |

Changing the camera pose can also help in reducing the fattening problem if we allocate more cameras/images to the part of the scene under concern, that is, we relocate the cameras/images while the number of input images remains fixed.

The set of 36 images are divided into 4 different subsets: $S_1$, $S_2$, $S_3$, and $S_4$, each of them contains 9 input images. The results of $S_1$ set have been shown in the previous experiment (Figures 2c and 2e). In sets $S_3$ and $S_4$, we allocate 5 cameras/images to the homogenous part of the house instead of 4 images allocated to the same part in sets $S_1$ and $S_2$. Table 2 shows the mean values of SNR measure of these sets. It is shown that allocating more images has enhanced the quality of the output reconstruction as it seems from SNR values for sets $S_3$ and $S_4$ compared to $S_1$ and $S_2$, even if the reconstruction of $S_2$ is slightly enhanced. Reprojected images at the same view as in Figure 2a for sets $S_2$ and $S_4$ are shown in Figures 3a and 3b, respectively. The corresponding difference images are also shown in Figures 3c and 3d, respectively.

### 4.2. Stereo versus space carving

A correlation-based stereo and the space carving algorithms are applied to the acquired images. Applying the reprojection methodology to the reconstructions of stereo and space carving, reprojected images are generated. Two input images



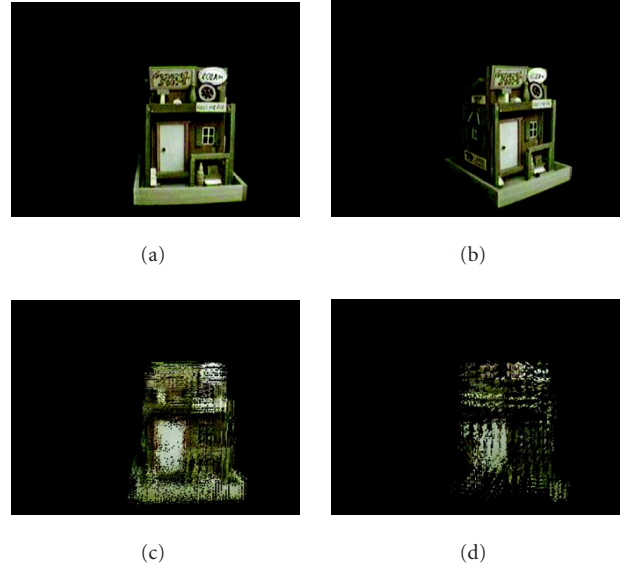(a)          (b)

(c)          (d)

FIGURE 4: (a) and (b) Two out of 12 images captured for a birdhouse object. A reprojected image to the same view as in (a) by (c) space carving and (d) stereo.

(out of 12) acquired for a birdhouse object are shown in Figures 4a and 4b. Figures 4c and 4d show reprojected images of space carving, and stereo, respectively, at the same view of Figure 4a. As shown, the space carving has better reconstruction resolution than stereo. This is because the resolution of space carving can be controlled by the number of voxels in the initial volume. However, stereo resolution depends mainly on the success of the matching strategy in solving the correspondence problem. Dense reconstructions are possible by stereo techniques, however, finding correct correspondences is not guaranteed when homogenous, slanted, or occluded surfaces are reconstructed. Removing such false matchings causes the lower output resolution as shown in the lower part of the house object in Figure 4d. Fitting of such missed parts of the reconstruction could overestimate (if successful fitting happened) or underestimate (if fitting failed) the performance of the given 3D reconstruction technique. The quality of each reconstruction is also reflected by the SNR values in Figure 5a.

The IFIM, and $Q_m$ values of the 3D reconstructions of both space carving and stereo are shown in Figures 5b and 5c, respectively. These measures are consistent with the SNR measure in judging the quality of the given 3D reconstructions. The curve of IFIM is smoother than that of SNR, Figure 5a, and $Q_m$, Figure 5c, which agrees with the logic that the variations in adjacent views are not too much. The SNR measure has the advantage that the error and the signal are explicit, then error analysis is feasible. The $Q_m$ measure still has similarity with SNR measures, even though there is no explicit relation between the signal and the error. Actually, $Q_m$ can be considered as a similarity measure. In other words, it measures the deviations from the original signal.
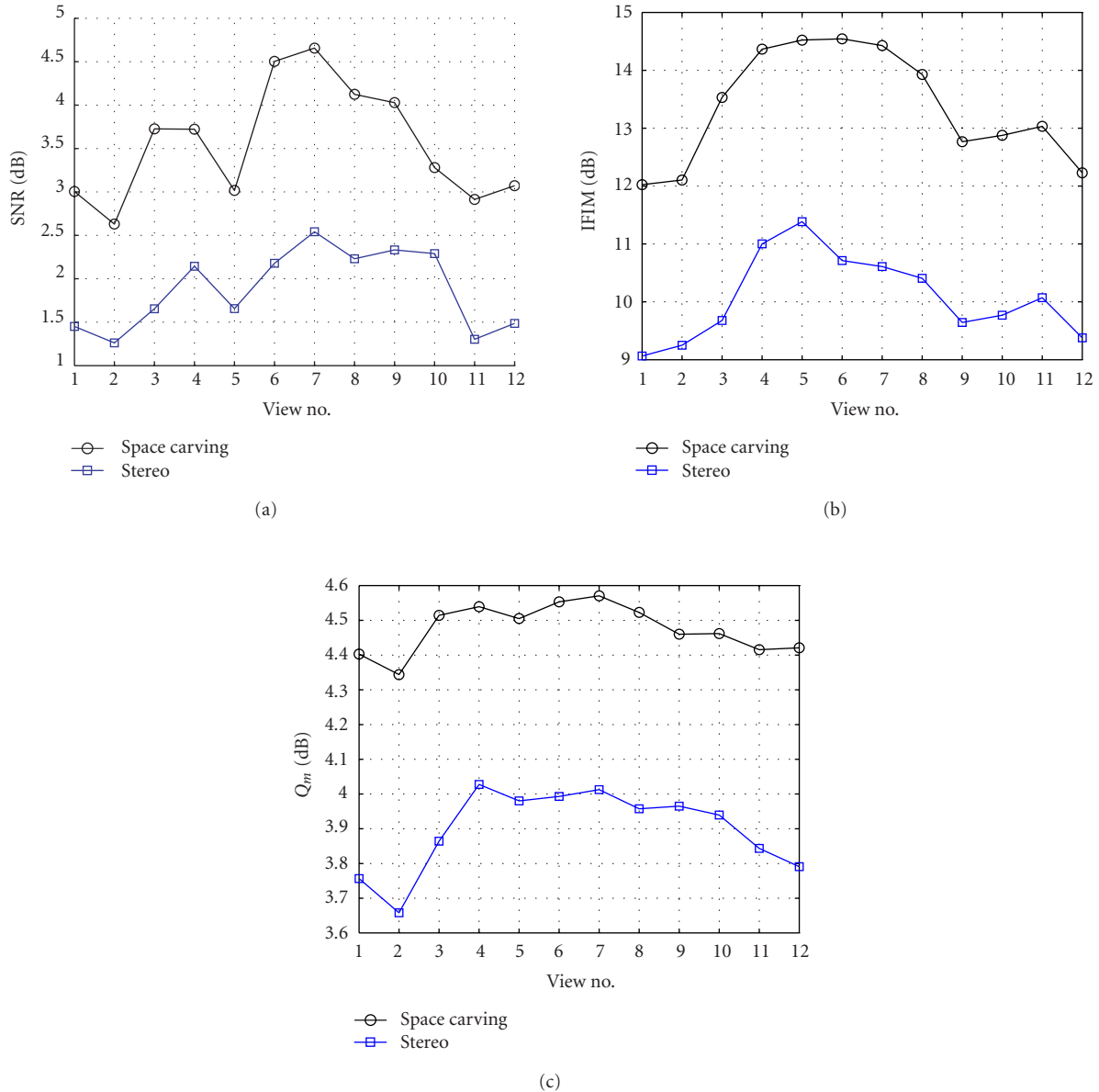
(a)



(b)



(c)

FIGURE 5: Quantifying the performance of both space carving and stereo using (a) the SNR measure, (b) the IFIM measure, and (c) the $Q_m$ measure.

Although, the IFIM is consistent with the other measures, it depends on a different philosophy. It is designed to resemble the human sensing of quality. Even if it is not clear how this is included in the measure, it is claimed that the FIM measure has some features of the subjective measures [15].

## 5. CONCLUSION AND FUTURE EXTENSIONS

In this paper, we have proposed a framework for the performance evaluation of 3D reconstruction techniques. Our goal is to set the terminology and the definitions of the process components as a starting step for standardizing and generalizing the evaluation process.

In our research laboratory, we developed a testing setup that helps bringing different 3D reconstruction techniques under the same testing methodology. In this paper, we test the performance of stereo and space carving based on the same data set and using an image reprojection technique. Different measuring strategies are used and they showed consistent results in estimating the performance of stereo and space carving techniques. It is shown by these measurements that the space carving performance can be enhanced if the number of input images is increased. In addition, we should distribute cameras (images) in such a way that considers the structure and the geometry of the given object under reconstruction.

A comparison example of space carving and stereo is included in this paper. The results show that space carving has better reconstruction than stereo. This is mainly because the 3D reconstructions by space carving have better resolution than that of stereo. This also exploits the fact that resolution of the 3D reconstructions using stereo is highly dependent on the matching strategy used to solve the correspondence problem.

The future work will focus on the design of other testing strategies and extending the evaluation domain to include different 3D reconstruction techniques. In addition, we will investigate if there is a link between the distribution of the errors that appear in the reprojected images and actual errors in the 3D space to validate the use of the images as ground truth in testing methodologies of 3D reconstruction techniques.

## REFERENCES

[1] R. Szeliski, "Prediction error as a quality metric for motion and stereo," in *Proc. 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 2, pp. 781–788, Kerkyra, Corfu, Greece, September 1999.

[2] R. Szeliski and R. Zabih, "An experimental comparison of stereo algorithms," in *Proc. International Workshop on Vision Algorithms*, pp. 1–19, Corfu, Greece, September 1999.

[3] J. Mulligan, V. Isler, and K. Daniilidis, "Performance evaluation of stereo for tele-presence," in *Proc. 8th IEEE International Conference on Computer Vision (ICCV '01)*, vol. 2, pp. 558–565, Vancouver, British Columbia, Canada, July 2001.

[4] K. N. Kutulakos and S. M. Seitz, "A theory of shape by space carving," in *Proc. 7th IEEE International Conference on Computer Vision (ICCV '99)*, vol. 1, pp. 307–314, Kerkyra, Corfu, Greece, September 1999.

[5] A. Eid and A. Farag, "On the performance characterization of stereo and space carving," in *Proc. Advanced Concepts for Intelligent Vision Systems (Acivs '03)*, pp. 291–296, Ghent, Belgium, September 2003.

[6] A. Farag and A. Eid, "Local quality assessment of 3-D reconstructions from sequence of images: a quantitative approach," in *Proc. Advanced Concepts for Intelligent Vision Systems (Acivs '04)*, Brussels, Belgium, August–September 2004.

[7] A. Eid and A. Farag, "On the fusion of 3-D reconstruction techniques," in *Proc. 7th International Conference on Information Fusion (IF '04)*, vol. 1, pp. 856–861, Stockholm, Sweden, June–July 2004.

[8] W. B. Culbertson, T. Malzbender, and G. Slabaugh, "Generalized voxel coloring," in *Proc. International Workshop on Vision Algorithms*, pp. 100–115, Corfu, Greece, September 1999.

[9] L. Robert, "Camera calibration without feature extraction," *Computer Vision and Image Understanding*, vol. 63, no. 2, pp. 314–325, 1996.

[10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2000.

[11] A. Eid and A. Farag, "Design of an experimental setup for performance evaluation of 3-D reconstruction techniques from sequence of images," in *Proc. Applications of Computer Vision Workshop in Conjunction with the European Conference on Computer Vision (ECCV '04)*, pp. 69–77, Prague, Czech, May 2004.

[12] C. Olson, "A general method for feature matching and model extraction," in *Proc. International Workshop on Vision Algorithms*, pp. 20–36, Corfu, Greece, September 1999.

[13] A. Eid and A. Farag, "A unified framework for performance evaluation of 3-D reconstruction techniques," in *Proc. Real-Time 3-D Sensors and Their Use in Conjunction with the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '04)*, pp. 33–33, Washington, DC, USA, June 2004.

[14] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 636–650, 2000.

[15] J. Li, G. Chen, and Z. Chi, "A fuzzy image metric with application to fractal coding," *IEEE Trans. Image Processing*, vol. 11, no. 6, pp. 636–643, 2002.

[16] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?" in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '02)*, vol. 4, pp. 3313–3316, Orlando, Fla, USA, May 2002.

**Ahmed Eid** received both the B.S. and M.S. degrees in electronics and communications engineering from Mansoura University, Egypt, in 1994 and 1999, respectively. He received the Ph.D. degree in electrical and computer engineering from the University of Louisville, USA, in 2004. He is currently a Research Associate at the Computer Vision and Image Processing (CVIP) Laboratory, University of Louisville. His current research interests include 3D model building from a sequence of images, 3D data registration, and data fusion. He is a Member of the IEEE and a Member of Eta Kappa Nu Honor Society. He was awarded the Silicon Graphics SGI Award "Excellence in Visualization and Computational Sciences" in 2003.

**Aly Farag** was educated at Cairo University (B.S. degree in electrical engineering), Ohio State University (M.S. degree in biomedical engineering), University of Michigan (M.S. degree in bioengineering), and Purdue University (Ph.D. degree in electrical engineering). He joined the University of Louisville in August 1990, where he is currently a Professor of electrical and computer engineering. His research interests are concentrated on the fields of computer vision and medical imaging. He is the Founder and Director of the Computer Vision and Image Processing Laboratory (CVIP Lab), the University of Louisville, which supports a group of over 20 graduate students and postdocs. His contribution has been mainly in the areas of active vision system design, and volume registration, segmentation, and visualization, where he has authored or coauthored over 80 technical articles in leading journals and international meetings in the fields of computer vision and medical imaging. He is an Associate Editor of the IEEE Transactions on Image Processing. He is a regular reviewer for a number of technical journals and to national agencies including the NSF and the NIH. He is a Senior Member of the IEEE and SME, and a Member of Sigma Xi and Phi Kappa Phi. He has recently been awarded a "University Scholar."