

# 國立交通大學

電子工程學系 電子研究所碩士班

## 碩士論文

3D 合成視角基於深度或色彩失真的品質估測

**Quality Assessment of 3D Synthesized View with Depth  
or Color Distortion**

研究生：蔡長廷

指導教授：杭學鳴 博士

中華民國一〇二年七月

3D 合成視角基於深度或色彩失真的品質估測

Quality Assessment of 3D Synthesized View with Depth or Color Distortion

研究生:蔡長廷

Student: Chang-Ting Tsai

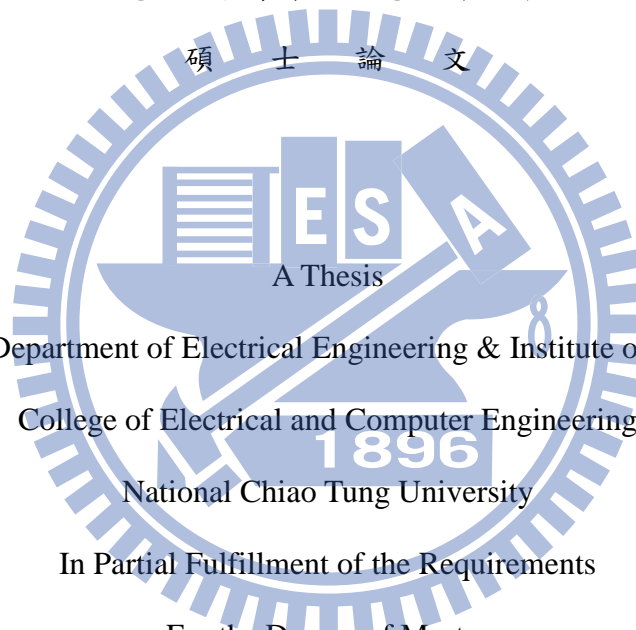
指導教授:杭學鳴 博士

Advisor: Dr. Hsueh-Ming Hang

國立交通大學

電子工程學系 電子研究所

碩 士 論 文



Submit to Department of Electrical Engineering & Institute of Electronics

College of Electrical and Computer Engineering

National Chiao Tung University

In Partial Fulfillment of the Requirements

For the Degree of Master

in

Electronics Engineering

July 2013

Hsinchu, Taiwan, Republic of China

中華民國一〇二年七月

# 3D 合成視角基於深度或色彩失真的品質估測

研究生：蔡長廷

指導教授：杭學鳴 博士

國立交通大學

電子工程學系 電子研究所碩士班

## 摘要



過去十幾年來，2D 影像品質估測的相關研究已經做了很多。隨著 3D 立體視訊的快速發展，能提供 3D 影像品質的數學模型就成了迫切的工具。MPEG 國際標準會議(ISO/IEC Moving Picture Expert Group)正在制定 3DVC(3D Video Coding)標準，這使解碼端能用解完彩色影像及相對的深度圖合成出自由視角的色彩圖。如何有效地衡量虛擬視角的品質為 3D 影像品質相關研究裡重要的課題之一。

大部分現存的 3D 品質估測模型是直接使用 2D 的方法來預測人類主觀的 3D 感受。但是對於 3D 自由視角的合成影像，解碼端產生的深度圖或者色彩圖若有失真，將在合成影像上產生如物體的位移、鬼影、前景及背景交接處的不自然現象……等等的現象。而這些現象和傳統的 2D 雜訊相當不一樣，所以使用 2D 品質估測模型來評量 3D 合成影像的品質是不適當的。

本篇論文中，提出了深度失真所合成影像的主觀品質資料庫，並且建立一個數學模型來評估這些合成影像。此模型分成兩個部分，一個是 IQS(Image Quality Score)，它負責衡量經過位移彌補的資訊的色彩品質；另一個 ESD(Edge Structural Distortion)使用 Hausdorff 距離來計算鬼影的程度。實驗結果顯示，我們的 3D 影像品質模型分數比傳統 2D 模型更能預測人類的主觀結果。

此論文中另外也建立了色彩失真所合成的自由視角影像的主觀品質資料庫，主要是要探討，雜訊及合成這兩個步驟的先後順序對於合成影像的影響。我們從實驗結果發現，對於高斯雜訊，先失真再合成會比先合成再失真有較高的 2D 品質以及主觀的感受；另外，彩色影像如果有模糊的失真，這樣和為失真的深度圖做合成會在合成影像前景及背景交接處的不自然現象。



# Quality Assessment of 3D Synthesized View with Depth or Color Distortion

Student: Chang-Ting Tsai

Advisor: Dr. Hsueh-Ming Hang

Department of Electrical Engineering &

Institute of Electronics

National Chiao Tung University

## Abstract

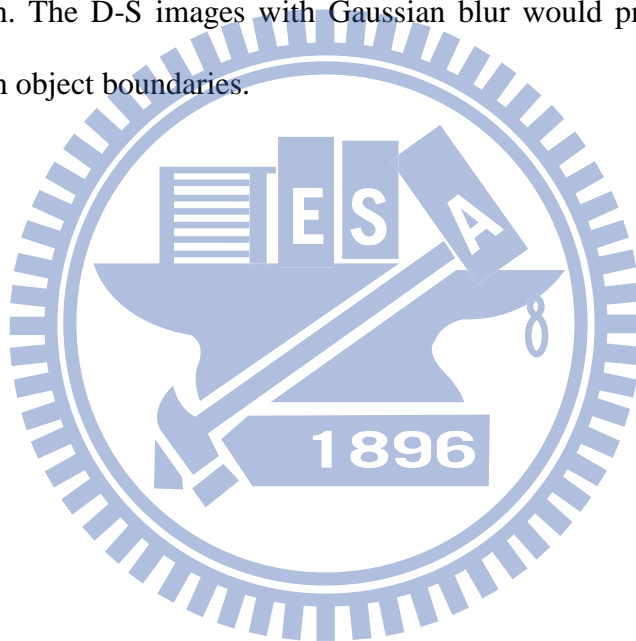
2D image/video quality assessment have been researched in the last decades. Since the popularity of 3D videos, quality assessment methods for the 3D contents become popular, too. The ISO/IEC Moving Picture Expert Group (MPEG) is in the process of specifying the 3D video coding (3DVC) standards based on the multiple-view plus depth (MVD) format. With the standard of 3D virtual view systems, how to predict the quality of the synthesized views becomes an important issue.

Most existing 3D image quality metrics use conventional 2D image quality assessment (IQA) models to predict the 3D subjective quality. But in a free viewpoint television (FTV) system, the depth map or color image errors often produce novel artifacts such as object shift, ghost artifact, sticker artifact etc. on the synthesized pictures due to the use of Depth Image Based Rendering (DIBR) technique. These artifacts are very different from the ordinary 2D distortions such as blur, Gaussian noise, and compression errors, and the pixel based 2D IQA metrics are sensitive to that. Thus, we describe a 3D databases with depth maps error.

We proposed an objective image QA model for depth map distortion. Proposed algorithm evaluates two scores, the Image Quality Score (IQS) and the Edge Structural Distortion (ESD). IQS computes 2D color quality of the synthesized image with object shift compensation. ESD estimates the degree of structural error by implying the Hausdorff distance. The final score of proposed model

is obtained by combining IQS and ESD together in the pooling stage. The experimental results show that the proposed method enhances the correlation of the objective quality score to the 3D subjective scores.

We also describe a 3D database with color distortion. There are two sets of view-synthesized images in the database, Distortion-Synthesis (D-S) images and Synthesis-Distortion (S-D) images. The most significant difference of these two kind of images is the distortion applied to images before rendering process or after it. In our collected data, the SSIM scores of the D-S images with Gaussian noise is much higher than those of the S-D images, that is, the view synthesis process can cover Gaussian noise distortion. The D-S images with Gaussian blur would produce the *sticker artifact* around the different depth object boundaries.



# 誌謝

在做碩士研究的旅途中我要感謝很多人，尤其是指導教授杭學鳴老師，我的題目算比較沒甚麼人做過的情況下，杭老師陪著我一起研究，一起學習；他同時也教導我許多做研究的態度及方法。

另外，要感謝 CommLab 的各位夥伴們在這過程中的陪伴，苦悶的時候，帶給我許多歡樂，尤其鈞凱、哲瑋和志堯以及杭 group 學弟們在研究上幫了我不少忙。而已經離開學校的峻利、朝雄、讀修、義文、維哲、凱翔、士傑學長們，謝謝你們在我剛進實驗時的那段時間對我的照顧。

最後要感謝我的家人以及在我身邊不斷鼓勵我的曼晴，你們在我的求學路上一直提供我精神上最大的慰藉。



蔡長廷

民國一〇二年七月 於新竹

# CONTENTS

Chapter 1	Introduction.....	1
1.1	Background.....	1
1.2	Motivation and Contribution.....	1
1.3	Organization of the Thesis .....	2
Chapter 2	Introduction to 3D Video Coding System.....	3
2.1	Overview.....	3
2.2	Depth Information.....	4
2.3	Depth Image Based Rendering (DIBR).....	5
Chapter 3	Quality Assessment (QA) .....	7
3.1	Subjective QA Methods .....	7
3.2	Objective QA Models .....	8
3.3	Evaluation of Objective Quality Assessment Models.....	10
3.4	Structural Similarity (SSIM) Metric .....	11
3.3	3D Quality Assessment Databases.....	12
Chapter 4	Proposed 3D Quality Assessment Model for Depth Map Distortion.....	16
4.1	3D Image Database with Depth Map Distortion.....	16
4.2.	Overview of Proposed 3D IQA Model .....	17
4.3.	Image Quality Score (IQS) .....	18
4.4.	The Hausdorff Distance .....	22
4.5.	Edge Structural Distortion (ESD).....	23
4.6.	Pooling .....	24
Chapter 5	Experimental Results of Depth Map Distortion Database.....	27
5.1.	Test Image Database Construction.....	27
5.2.	Subjective Test .....	29
5.3.	Parameters of the Model .....	30
5.3.1.	Pooling Proportion ( $p$ ).....	31
5.3.2.	Pooling weight ( $\alpha$ ).....	34
5.3.3.	Modified Hausdorff Distance ( $pKth$ ).....	35
5.3.4.	Gaussian filter ( $\sigma Gaussian$ ) .....	38
5.3.5.	Window size ( $N$ ).....	44
5.4.	Experimental Result and Analysis .....	45
Chapter 6	3D Quality Assessment for Color Distortion.....	56
6.1.	Synthesized View Database with Color Distortion.....	56



6.2.	Subjective Test .....	58
6.3.	Experimental Results and Analysis.....	60
Chapter 7	Conclusions and Future Works .....	66
7.1.	Conclusions.....	66
7.2.	Future Work .....	67
References	.....	68



# List of Figures

FIGURE 1. FRAMEWORK OF 3DVC SYSTEM.....	3
FIGURE 2. AN ILLUSTRATION OF TRANSLATION BETWEEN DISPARITY AND DEPTH DISTANCE .....	4
FIGURE 3. ILLUSTRATION OF SYNTHESIZING POINT P1 ON THE VIRTUAL IMAGE PLANE.....	6
FIGURE 4. STRUCTURE OF SS. ....	8
FIGURE 5. 3D DATABASES .....	13
FIGURE 6. ADDITIVE DISTORTIONS FOR THE VENUS MODEL [14]. ....	13
FIGURE 7. SCENES OF THE DATABASE [15]. ....	14
FIGURE 8. DIFFERENT DIBR ALGORITHMS FOR LOVEBIRD SEQUENCE [14]. ....	15
FIGURE 9. (A) ORIGINAL DEPTH MAP, (B) OFFSET, (C) QUANTIZATION, AND (D) GAUSSIAN_NOISE .....	17
FIGURE 10. FLOW CHART OF THE PROPOSED IQA MODEL.....	18
FIGURE 11. THE PROPOSED COMPARISON SYSTEM.....	18
FIGURE 12. (B) IS THE VIRTUAL VIEW SYNTHESIZED USING THE ORIGINAL DEPTH MAP (A).....	19
FIGURE 13. QUALITY INDEX MAP EVALUATED ON FIGURE 12(B) USING FIGURE 12(D). ....	19
FIGURE 14. FLOW CHART OF COMPUTING THE IMAGE QUALITY SCORE (IQS).....	20
FIGURE 15. SOME LOW IQS REGION OF FIGURE 13.....	21
FIGURE 16. (A) MAGNIFIED FROM FIGURE 12(B); (B) MAGNIFIED FROM FIGURE 12(D).....	21
FIGURE 17. MODIFIED FLOW CHART OF COMPUTING IQS. ....	21
FIGURE 18. THE IQS INDEX MAP OF FIGURE 13 USING THE PROCESS SHOWN IN FIGURE 17.....	22
FIGURE 19. (A) A SPECIFIED MODEL (B) A TEST IMAGE (C) THE MATCHED RESULT.....	23
FIGURE 20. AN ILLUSTRATION OF THE DIRECTED HAUSDORFF DISTANCE.....	23
FIGURE 21. FLOW CHART OF COMPUTING ESD. ....	24
FIGURE 22. EXAMPLE OF LOWER QUALITY PICTURE IN USING THE MEAN AS THE FINAL SCORE. ....	25
FIGURE 23. FLOW CHART OF THE POOLING STAGE.....	26
FIGURE 24. MOS OF THE SUBJECTIVE TEST. ....	29
FIGURE 25. VARIANCE OF THE SUBJECTIVE TEST.....	29
FIGURE 26. PARAMETERS OF THE PROPOSED MODEL.....	30
FIGURE 27. AN EXAMPLE OF SHIFT COMPENSATION WITH SEARCH_WINDOW=6 AND N=5. ....	30
FIGURE 28. AN EXAMPLE OF $pKth$ . ....	31
FIGURE 29. THE RELATION BETWEEN $\alpha$ AND THE PROPORTION OF IQS AND ESD.....	31
FIGURE 30. DIFFERENT VALUE OF THE POLLING PROPORTION P AFFECTS PLCC.....	32
FIGURE 31. ARTIFACT DUE TO GAUSSIAN_NOISE DEPTH MAP ERROR.....	32
FIGURE 32. POOLING PROPORTION P NEARBY 5% HAS BETTER PERFORMANCE UNDER DIFFERENT POOLING WEIGHT $\alpha$ . ( $N = 25, pKth = 70$ ) .....	33
FIGURE 33. ILLUSTRATION OF EYE DOMINATION. [31].....	34
FIGURE 34. ANOTHER ILLUSTRATION OF EYE DOMINATION. [31].....	34

FIGURE 35. A AGAINST PLCC. ( $N = 25, pKth = 70$ ).....	35
FIGURE 36. IQS DOMINATES THE PERFORMANCE IN <i>GAUSSIAN_NOISE</i> CASE. ( $N = 25, pKth = 70$ ) .....	35
FIGURE 37. THE EFFECT OF $pKth$ WHEN $\alpha = 0.7$ . (WINDOW = 25, POOLING P = 5%) .....	36
FIGURE 38. THE EFFECT OF $pKth$ WHEN $\alpha = 0.3$ . (WINDOW = 25, POOLING P = 5%) .....	36
FIGURE 39. $pKth = 35$ AND $pKth = 70$ HAVE CLOSE PERFORMANCE WHEN THE VALUE OF A IS BETWEEN 0.2 TO 1. (WINDOW = 25, POOLING P = 5%) .....	37
FIGURE 40. (A) SYNTHESIZED IMAGE OF <i>KENDO</i> WITH <i>OFFSET</i> DISTORTION ( <i>OFFSET_VALUE=60</i> ). .....	38
FIGURE 41. LEFT COLUMN SHOWS THE <i>QUANTIZATION</i> DISTORTED IMAGE OF SEQUENCE <i>POZNAN_STREET</i> WITH GAUSSIAN FILTER USING DIFFERENT $\sigma_{Gaussian}$ VALUES. RIGHT COLUMN IS THE CORRESPONDING IQS MAPS. (A) NO GAUSSIAN FILTER, (B) $\sigma_{Gaussian} = 1$ , (C) $\sigma_{Gaussian} = 2$ , AND (D) $\sigma_{Gaussian} = 3$ .....	39
FIGURE 42. LEFT COLUMN SHOWS THE <i>OFFSET</i> DISTORTED IMAGE OF SEQUENCE <i>BALLOONS</i> WITH GAUSSIAN FILTER USING DIFFERENT $\sigma_{Gaussian}$ VALUES. RIGHT COLUMN IS THE CORRESPONDING IQS MAPS.....	40
FIGURE 43. PLCC HAS MAXIMUM VALUE WHEN $\sigma_{Gaussian} = 2$ FOR THE <i>OFFSET</i> AND THE <i>QUANTIZATION</i> . ( $\alpha = 0.3$ , WINDOW = 25, POOLING P = 5%, $pKth = 70$ ) .....	41
FIGURE 44. $\sigma_{Gaussian}$ SHOULD BE CHOSEN SMALLER WHEN $\alpha$ IS SMALL.....	42
FIGURE 45. LARGE $\sigma_{Gaussian}$ IS BETTER FOR <i>GAUSSIAN_NOISE</i> DISTORTION.....	42
FIGURE 46. LEFT COLUMN SHOWS THE <i>GAUSSIAN_NOISE</i> DISTORTED IMAGE OF SEQUENCE <i>BALLOONS</i> WITH GAUSSIAN FILTER USING DIFFERENT VARIANCE PARAMETERS. RIGHT COLUMN IS THE CORRESPONDING IQS MAPS. (A) NO GAUSSIAN FILTER, (B) $\sigma_{Gaussian} = 1$ , (C) $\sigma_{Gaussian} = 2$ , AND (D) $\sigma_{Gaussian} = 3$ .....	43
FIGURE 47. WINDOW SIZE DOES NOT AFFECT THE PERFORMANCE MUCH.....	44
FIGURE 48. EXECUTION TIME PER IMAGE DECREASES WHEN THE WINDOW SIZE BECOMES LARGER. .	45
FIGURE 49. SCATTER PLOT OF THE OBJECTIVE QUALITY SCORES AGAINST THE DMOS. (A) PSNR (B) SSIM (C) MSSIM (D) VIF (E) VSNR (F) PROPOSED. ( <i>OFFSET</i> AND <i>QUANTIZATION</i> ).....	49
FIGURE 50. SCATTER PLOT OF THE OBJECTIVE QUALITY SCORES AGAINST THE DMOS. (A) PSNR (B) SSIM (C) MSSIM (D) VIF (E) VSNR (F) PROPOSED. ( <i>GAUSSIAN_NOISE</i> ).....	50
FIGURE 51. SCATTER PLOT OF THE OBJECTIVE QUALITY SCORES AGAINST THE DMOS. (A) PSNR (B) SSIM (C) MSSIM (D) VIF (E) VSNR (F) PROPOSED. (OVERALL DATABASE).....	53
FIGURE 52. STRUCTURE OF COMPUTING 2D QUALITY INDEX IN OUR MODEL.....	54
FIGURE 53. SCATTER PLOTS OF THE OBJECTIVE QUALITY SCORES AGAINST DMOS APPLIED TO THE LIVE DATABASE. (A) SSIM, AND (B) GAUSSIAN FILTER + SSIM .....	55
FIGURE 54. THE D-S IMAGE PRODUCING PROCESS AND ITS QUALITY EVALUATION.....	57
FIGURE 55. THE S-D IMAGE PRODUCING PROCESS AND ITS QUALITY EVALUATION.....	57
FIGURE 56. 3D SCENARIO AND 2D SCENARIO OF THE D-S IMAGES.....	59
FIGURE 57. 3D SCENARIO AND 2D SCENARIO OF THE S-D IMAGES.....	59
FIGURE 58. A SUMMARY OF THE DATABASE.....	61

FIGURE 59. SSIM COMPARISON BETWEEN D-S AND S-D IMAGES. (2D SCENARIO) ..... 62

FIGURE 60. THE COMPARISON OF *AGN* AND *SGN*. ..... 62

FIGURE 61. SUBJECTIVE EXPERIMENTAL RESULTS OF 2D SCENARIO. .... 63

FIGURE 62. GAUSSIAN BLUR DISTORTION. (A) D-S IMAGE. (B) S-D IMAGE. .... 64

FIGURE 63. AN ILLUSTRATION OF STICKER ARTIFACT. .... 64

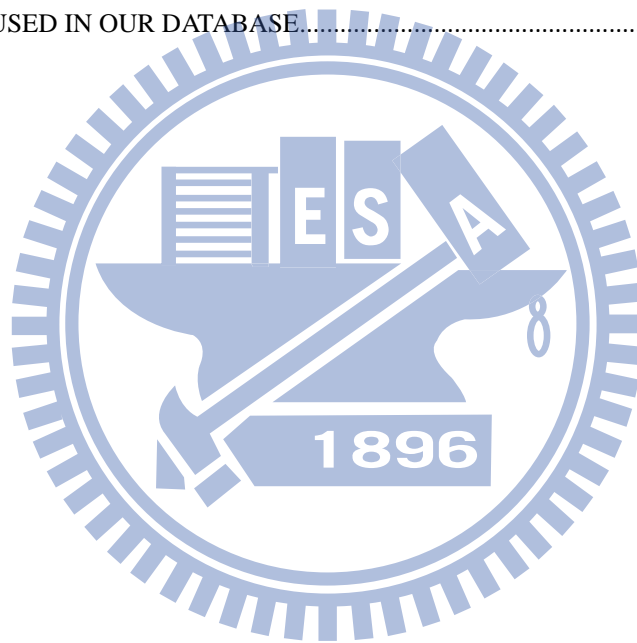
FIGURE 64. SSIM COMPARISON BETWEEN D-S AND S-D IMAGES (3D SCENARIO). .... 65

FIGURE 65. SUBJECTIVE COMPARISON BETWEEN D-S AND S-D IMAGES (3D SCENARIO). .... 65



# List of Tables

TABLE 1. 2D OBJECTIVE QA MODELS .....	10
TABLE 2. TEST SEQUENCES. ....	28
TABLE 3. THE INFORMATION OF THE TEST SEQUENCES.....	28
TABLE 4. PARAMETERS SET TO EVALUATE THE DATABASE. ....	45
TABLE 5. PERFORMANCE COMPARISON OF <i>OFFSET</i> AND <i>QUANTIZATION</i> . ....	47
TABLE 6. PERFORMANCE COMPARISON OF <i>GAUSSIAN_NOISE</i> . ....	48
TABLE 7. PARAMETERS SET TO EVALUATE THE OVERALL DATABASE.....	51
TABLE 8. PERFORMANCE COMPARISON OF OVERALL DATABASE.....	52
TABLE 9. PERFORMANCE COMPARISON.....	54
TABLE 10. PARAMETERS USED IN OUR DATABASE.....	58



# Chapter 1 Introduction

## 1.1 Background

3D contents are becoming more popular recently and widely used in our life. The ISO/IEC Moving Picture Expert Group (MPEG) standardizes 3D video coding (3DVC) based on the multi-view plus depth (MVD) format. With the progress of 3D video coding, how to predict the quality of the synthesized images becomes an important issue.

2D quality assessment (QA) is a popular research topic in last decade. Several well-known computation models have been proposed and some are still in used, for example, Peak Signal to Noise Ratio (PSNR), Structural Similarity (SSIM), and Visual Information Fidelity (VIF). However, unlike 2D QA, 3D QA is much more complex and has not been well-studied yet.

## 1.2 Motivation and Contribution

There are many challenges to evaluate a 3D images QA model [1]. Most of the proposed methods use the 2D image computation models to predict 3D subjective quality [2][3][4]. But for the free viewpoint television (FTV) system, errors in the depth map can cause object shift or produce ghost artifacts near object boundaries on the synthesized pictures after Depth Image Based Rendering (DIBR). These artifacts are different from those of the 2D distortions such as blur, Gaussian noise, and compression. Thus, the commonly used 2D quality computation metrics are not enough to assess the quality of synthesized images.

Because the pixel based 2D QA models are inadequate in measuring these artifacts, we

propose a new QA model to assess the quality of distorted image synthesized by the depth map distorted by different types of distortion. Our model uses block-matching algorithm to reduce the object shift effect. And the Hausdorff distance is used to determine the degree of ghost artifact. The experimental results show that the method enhances the correlation between the objective quality scores and the 3D subjective scores.

### **1.3 Organization of the Thesis**

In this thesis, we first introduce the standard 3D video coding system in chapter 2. And then in chapter 3 we briefly review the background of 2D quality assessment and introduce some public 3D QA databases. We design and create our 3D QA database with depth distortion in chapter 4 and also describe the proposed model. The experimental results, analysis and comparisons are shown and discussed in chapter 5. The second 3D QA database with color distortion and its discussion are put in chapter 6. Finally, conclusions and future work are briefly mentioned in chapter 7.

# Chapter 2 Introduction to 3D Video Coding System

## 2.1 Overview

Figure 1 shows the framework of 3D video coding (3DVC) system. It assumes that the input video either 2-view or 3-view. Each view has its color image and corresponding depth map which can be captured by specific device or generated by a depth estimation method. Then, these data are compressed by 3D video encoder. After transmission, 3D video decoder decodes the data. Finally, the virtual view images are generated by a view synthesis algorithm.

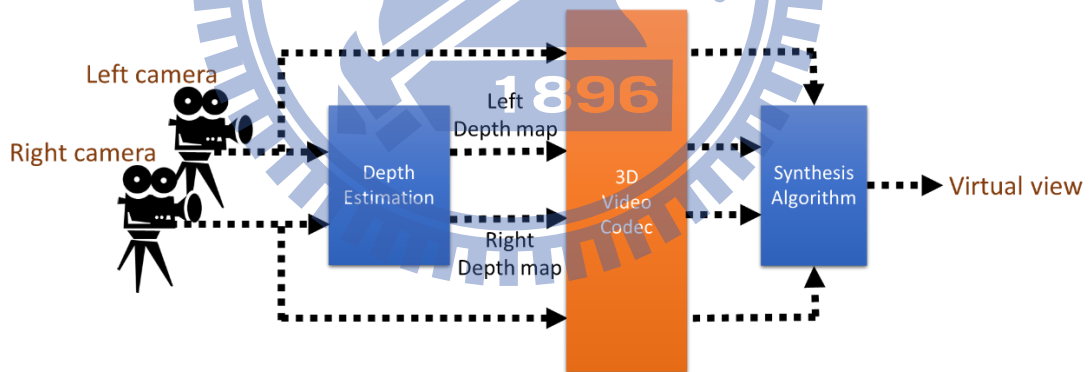


Figure 1. Framework of 3DVC system.

MPEG provides software to do depth estimation and view synthesis. They are Depth Estimation Reference Software (DERS) and View Synthesis Reference Software (VSRS). ATM (AVC Test Model) and HTM (HEVC Test Model) are AVC (Advance Video Coding) based and HEVC (High Efficiency Video Coding) based reference software of 3D video coding. The view synthesis algorithm used in ATM and HTM is called “1D-fast-VSRS.”



## 2.2 Depth Information

Depth value indicates the distance between the camera and the object. It plays an important role in 3D video coding. Disparity is the horizontal shift of two pixels on the respective stereo pictures that project into the same point in the scene. Disparity can be computed from the known depth distance, baseline between two cameras, and focal length of the camera as shown in Figure 2.

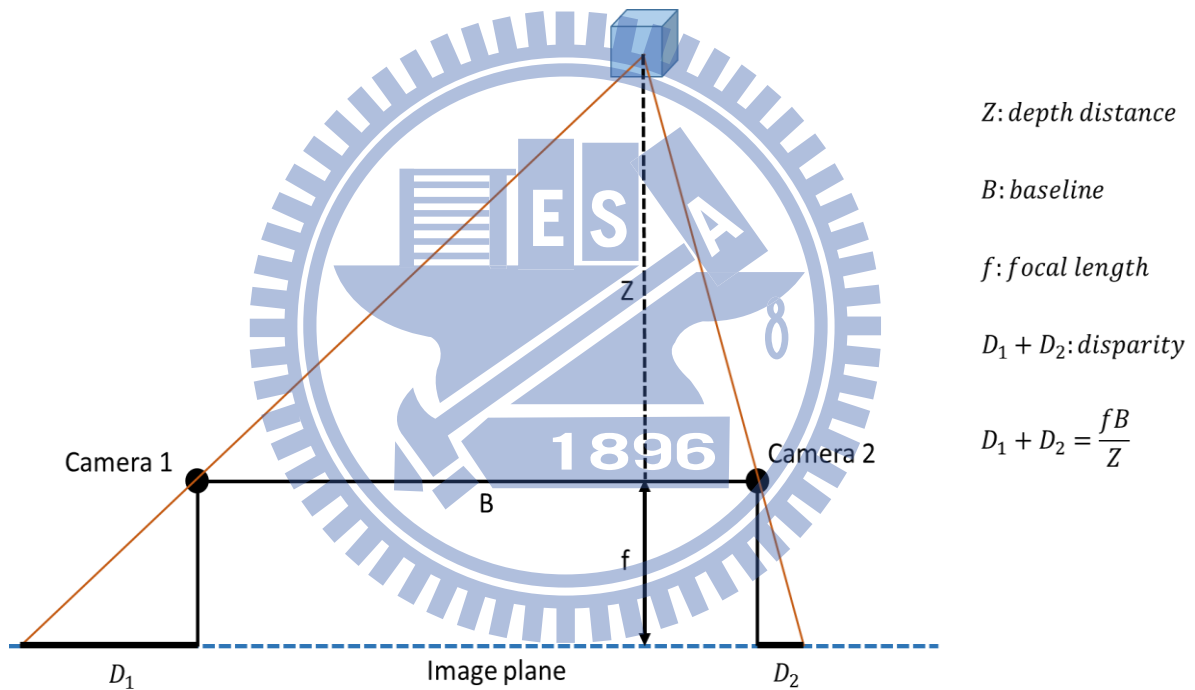


Figure 2. An illustration of translation between disparity and depth distance

Depth map can be truncated as a gray-level image. It quantizes depth distance into 256 levels. The Quantization levels of near field are denser than those of the far field. Because the human pay more attention on the foreground object, so its depth levels need to be represented with higher precision. The following equation shows the relation between the depth distance and the depth value.

$$z = \frac{1}{\frac{v}{255} \left( \frac{1}{Z_{near}} - \frac{1}{Z_{far}} \right) + \frac{1}{Z_{far}}} \quad (1)$$

Where  $z$  is the physical depth,  $Z_{near}$  and  $Z_{far}$  are the nearest and the farthest depth distance of the 3D scene, and  $v$  is depth value.

## 2.3 Depth Image Based Rendering (DIBR)

Depth Image Base Rendering (DIBR) is the process of generating virtual views of a scene [5]. It projects the original view image points into the 3D world via its associated depth information. Then, these 3D world points are reprojected into the image plan of a virtual camera. The process can be illustrated as Figure 3. The pixel positions  $P_1$  and  $P_2$  in both two image plane can be defined:

$$\lambda_1 P_1 = \mathbf{K}_1 \mathbf{R}_1 \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \mathbf{K}_1 \mathbf{R}_1 \mathbf{C}_1 \quad (2)$$

$$\lambda_2 P_2 = \mathbf{K}_2 \mathbf{R}_2 \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - \mathbf{K}_2 \mathbf{R}_2 \mathbf{C}_2 \quad (3)$$

where  $\mathbf{R}$  is the rotation matrix;  $\mathbf{C}$  describe the coordinates of the camera center;  $\mathbf{K}$  represents the 3x3 intrinsic matrix of the camera;  $\lambda$  is the scaling factor; and  $(X Y Z)^T$  is the corresponding 3D world point, and it can be written as

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = (\mathbf{K}_1 \mathbf{R}_1)^{-1} \cdot (\lambda_1 P_1 + \mathbf{K}_1 \mathbf{R}_1 \mathbf{C}_1) \quad (4)$$

Then we obtain the pixel position of the virtual image plane

$$\lambda_2 P_2 = \mathbf{K}_2 \mathbf{R}_2 (\mathbf{K}_1 \mathbf{R}_1)^{-1} \cdot (\lambda_1 P_1 + \mathbf{K}_1 \mathbf{R}_1 \mathbf{C}_1) - \mathbf{K}_2 \mathbf{R}_2 \mathbf{C}_2 \quad (5)$$

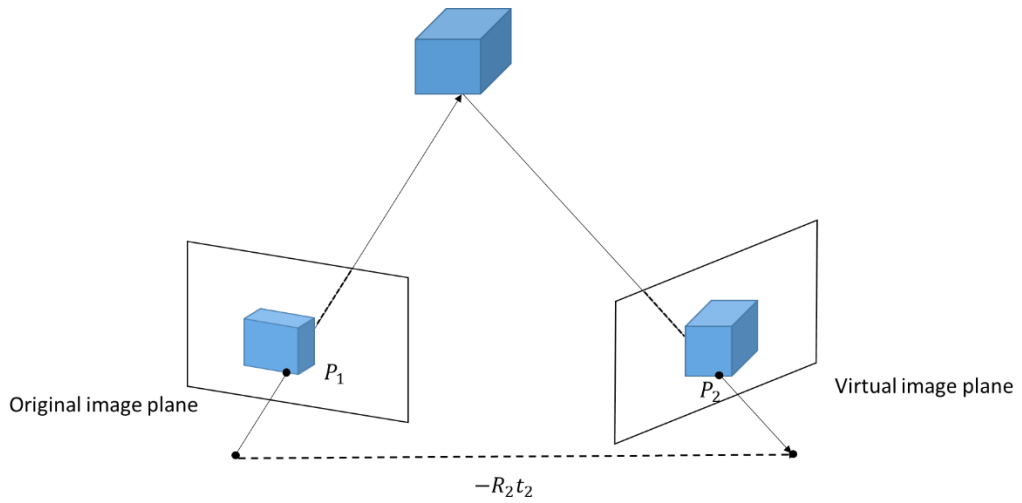


Figure 3. Illustration of synthesizing point  $P_1$  on the virtual image plane.



# Chapter 3 Quality Assessment (QA)

There are two quality assessment types: subjective and objective. Subjective quality assessment is done by human observers watching the test sequences and giving them quality scores. The subjective test is not only a costly and time consuming approach, but also cannot implement in a machine. Therefore, the objective measurements are very desirable. The goal of objective QA is to develop algorithms that can predict the visual quality of sequences does to that made by human.

## 3.1 Subjective QA Methods

Recommendation document ITU-R BT.500 [6] outlines a few subjective test processes to judge the quality of pictures. They are single stimulus (SS), double stimulus continuous quality (DSCQS), stimulus comparison (SC), and single stimulus continuous quality evaluation (SSCQE). We will describe the method used in this study.

The testing sequence of a typical SS is shown as Figure 4. In each test, the trail number should be shown no longer than 3 seconds. Then, the test image or video sequence should be designed for around 10 seconds. Finally, the word “*vote*” should last long enough time so that a viewer has time for giving a rating.

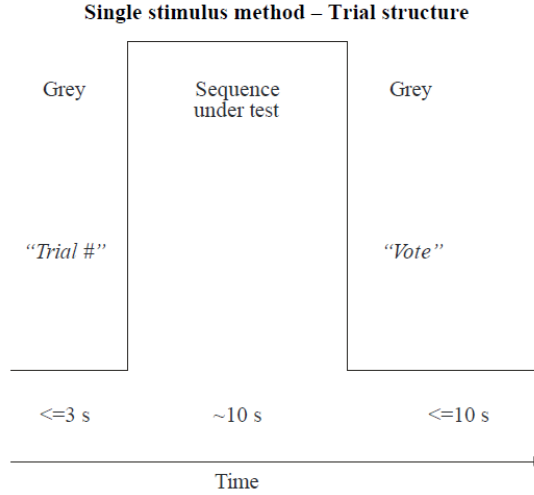


Figure 4. Structure of SS.

The rating for each sequence under experiment is termed “*opinion score*,” and the average of rating of all the observers is termed “*mean opinion score*” (MOS). The reference video sequences should be included in the test sequence. The arithmetic difference between the MOS of a processed picture and the corresponded reference image is called “*difference mean opinion score*” (DMOS).

$$MOS_i = \frac{1}{N} \sum_{o=1}^N rating_{i,o} \quad (6)$$

$$DMOS_i = MOS_{i,Reference} - MOS_i \quad (7)$$

where  $rating_{i,o}$  is the rating of the  $i$ th sequence given by the  $o$ th observer;  $N$  is the total number of observers in the experiment; and  $MOS_{i,Reference}$  is the MOS for the hidden reference, and the one relative to the  $i$ th sequence,  $MOS_i$ .

### 3.2 Objective QA Models

There are different categories or goals for objective QA models [7].

## **(1) Reference**

### **(a) Full Reference (FR)**

Most of the proposed QA models assume undistorted reference image exists and is available. And the distorted image is compare with the undistorted image.

### **(b) No Reference (NR)**

In many applications, QA system doesn't receive the reference image. Therefore, quality of the distorted image is computed by itself.

### **(c) Reduce Reference (RR)**

In this type QA model, the undistorted image is not fully available. Instead, some features are extracted from the undistorted image, and QA model evaluate quality from the features as side information.

## **(2) Purpose**

### **(a) General-Purpose**

General-purpose QA model do not assume any specific types, and is designed for varied application.

### **(b) Application-Specific**

There are a lot of QA models are designed for specific applications, such as image and video compression is one of the largest applications.

## **(3) Bottom-Up or Top-Down**

### **(a) Bottom-Up**

The goal of a bottom-up system is to build a system that has the same functions to the human visual system (HVS). This approach studies each component of the HVS, simulates all components as basic blocks, and then combines them together.

### **(b) Top-Down**

Top-down approach only concern about input-output relationship. So its system is not the same as the HVS. Instead, the HVS is treated as a black block.

2D QA methods have been proposed and work well in the last decades. Table 1 shows some popular 2D image QA algorithms. The simplest and most widely used 2D QA model is PSNR, but it has its limitation. Wang et al. [8] have proposed a full reference QA model, structural similarity index (SSIM), and it is more compatible to human perception than PSNR. We will describe later in this study.

Table 1. 2D Objective QA models.

Algorithm		Algorithm	
Peak Signal to Noise Ratio	PSNR	Universal Quality Index [10]	UQI
Structural Similarity index [8]	SSIM	Visual Information Fidelity [11]	VIF
Multi-scale SSIM [9]	MSSIM	Visual Signal to Noise Ratio [12]	VSNR

### 3.3 Evaluation of Objective Quality Assessment Models

Video Quality Experts Group (VQEG) Full Reference Television (FRTV) Phase II report [13] recommends the process that measures the performance of the objective QA models. First, use the following logistic function for fitting the DMOS and objective score.

$$DMOS_p = \frac{b_1}{1 + e^{(-b_2(score - b_3))}}, \quad (8)$$

where "score" is the quality score predicted by objective QA model;  $DMOS_p$  is the DMOS value that transferred from the "score" by the nonlinear regression function; and the  $b_1, b_2,$  and  $b_3$  are the parameters that are obtained through the regression step to minimize the error between  $DMOS$  and  $DMOS_p$ . Then, three common used index are defined as follow:

- (1) Pearson Linear Correlation Coefficient (PLCC)

$$PLCC = \frac{\sum_{i=1}^N (DMOS_i - \overline{DMOS})(DMOS_{pi} - \overline{DMOS_p})}{\sqrt{\sum_{i=1}^N (DMOS_i - \overline{DMOS})^2} \sqrt{\sum_{i=1}^N (DMOS_{pi} - \overline{DMOS_p})^2}} \quad (9)$$

(2) Spearman's Rank Order Correlation Coefficient (SROCC)

The raw scores  $DMOS_i$  and  $DMOS_{pi}$  are converted to ranks  $x_i$  and  $y_i$ , and

SROCC is computed from:

$$SROCC = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^N (y_i - \bar{y})^2}} \quad (10)$$

(3) Root Mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{1}{M} \sum_{i=1}^M (DMOS - DMOS_p)^2} \quad (11)$$

(4) Outlier Ratio (OR)

$$outlier\ ratio = \frac{total\ number\ of\ outlier}{total\ points} \quad (12)$$

where an outlier point is defined as follow:

$$|DMOS - DMOS_p| > 2\sigma \quad (13)$$

### 3.4 Structural Similarity (SSIM) Metric

SSIM is designed based on the observation that HVS is highly adapted for the structure information of a scene in which pixels have strong inter-dependency. The SSIM index is composed by three components: luminance, contrast, and structure.

$$SSIM(\mathbf{x}, \mathbf{y}) = [l(\mathbf{x}, \mathbf{y})]^\alpha \cdot [c(\mathbf{x}, \mathbf{y})]^\beta \cdot [s(\mathbf{x}, \mathbf{y})]^\gamma \quad (14)$$

Let  $\mathbf{x} = \{x_i | i = 1, 2, \dots, N\}$  and  $\mathbf{y} = \{y_i | i = 1, 2, \dots, N\}$  be the two windows extracted from the same location of the reference and the distorted images. Where  $x_i$  and  $y_i$  are the  $i$ th location in the window, and  $N$  is the number of pixels in a window. And  $\alpha$ ,  $\beta$ , and  $\gamma$  are three positive parameters used to adjust the relative importance of the three components.



The luminance, contrast, and structure comparison functions  $l(\mathbf{x}, \mathbf{y})$ ,  $c(\mathbf{x}, \mathbf{y})$ , and  $s(\mathbf{x}, \mathbf{y})$  are defined as follow.

$$l(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \quad (15)$$

$$c(\mathbf{x}, \mathbf{y}) = \frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2} \quad (16)$$

$$s(\mathbf{x}, \mathbf{y}) = \frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \quad (17)$$

where  $\mu_x$  and  $\mu_y$  are the means of  $\mathbf{x}$  and  $\mathbf{y}$ ;  $\sigma_x$  and  $\sigma_y$  are the standard deviations of  $\mathbf{x}$  and  $\mathbf{y}$ ; and  $\sigma_{xy}$  is the correlation coefficient between  $\mathbf{x}$  and  $\mathbf{y}$ . Positive parameters  $C_1$ ,  $C_2$ , and  $C_3$  avoid the denominators close to zero.

$$\mu_x = \frac{1}{N} \sum_{i=1}^N x_i \quad (18)$$

$$\sigma_x = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)^2} \quad (19)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y) \quad (20)$$

Finally, the overall image quality is evaluated from the mean SSIM (MSSIM).

$$MSSIM(\mathbf{X}, \mathbf{Y}) = \frac{1}{M} \sum_{j=1}^M SSIM(\mathbf{x}_j, \mathbf{y}_j) \quad (21)$$

where  $\mathbf{X}$  and  $\mathbf{Y}$  are the reference and the distorted images;  $\mathbf{x}_j$  and  $\mathbf{y}_j$  are the image contents at the  $j$ th local window; and  $M$  is the number of local windows of the image.

### 3.3 3D Quality Assessment Databases

There are several organizations that conduct the 3D quality research and publish their database on the website. For the purpose, we can classify these databases as shown in Figure 5.

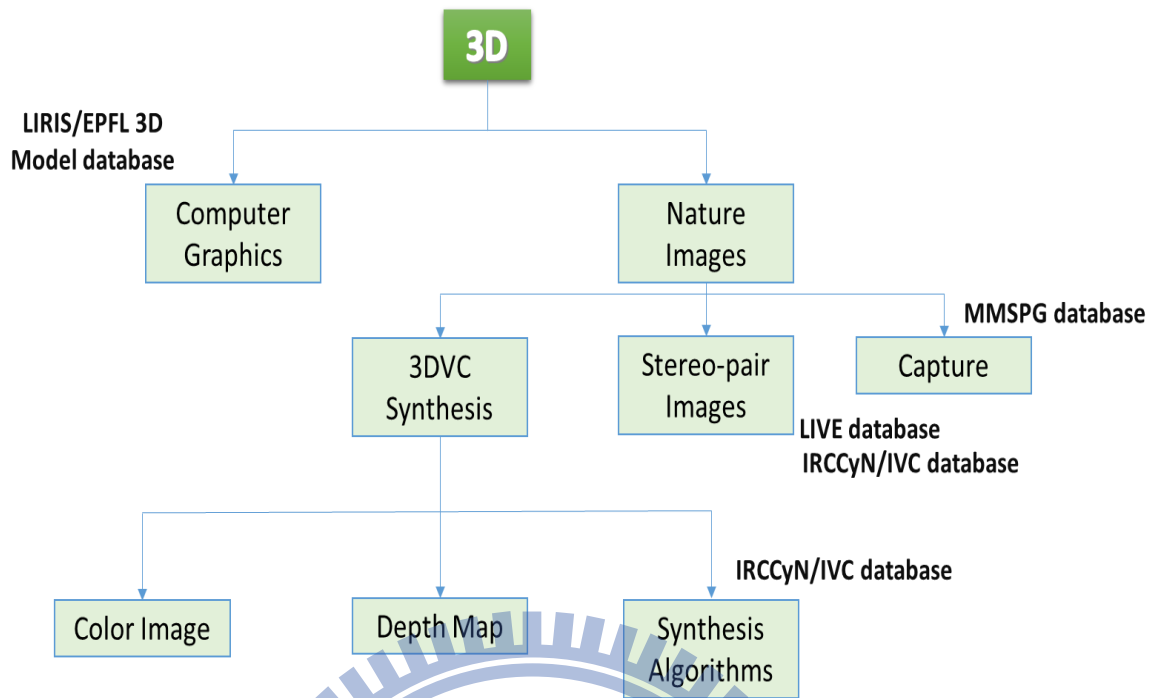


Figure 5. 3D databases

**(1) Computer Graphics [14]**

Lavoue et al. [14] describes a 3D model database. 88 models between 40K and 50K vertices were generated from 4 reference objects. There are two types of distortion: noise and smoothing. Subjective evaluations are using the Single Stimulus (SS) method with 12 observers.

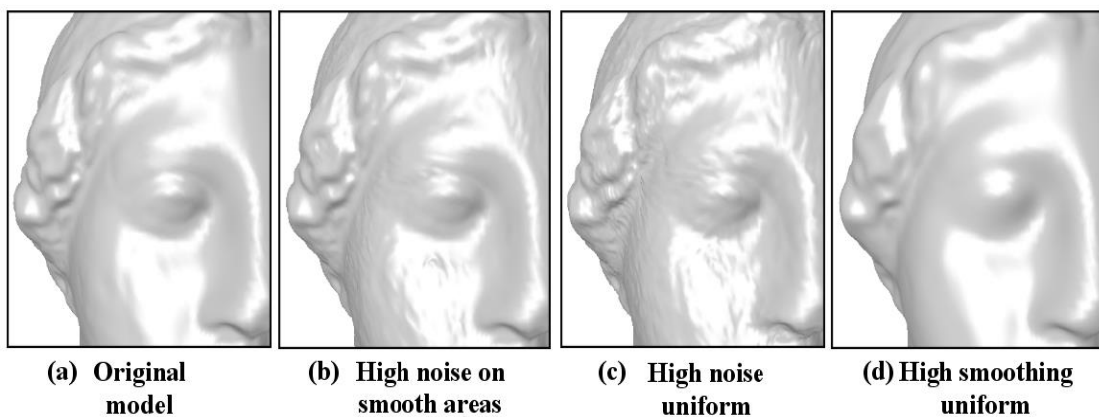


Figure 6. Additive distortions for the *Venus* model [14].

## (2) Captured 3D Images [15]

Goldmann et al. [15] presents a 3D image QA database which contains stereoscopic images with resolution of 1920x1088 pixels. 10 scenes with various textures and depth structures were captured. And each of the scenes has been captured with different baseline in the range 10-50 cm.



Figure 7. Scenes of the database [15].

## (3) Stereo-pair Images with Color Distortion [16]- [18]

Benoit et al. [16] describes a database applied three different distortions (JPEG, JPEG2000, and blur) symmetrically to the stereo pair images. Urvoy et al. [17] distort the stereo pair images based on H.264 and JPEG2000 coding and typical image processing steps such as down sampling and sharpening. The distortions that Moorthy et al. [18] use include JPEG and JPEG2000 coding, additive white Gaussian noise, Gaussian blur, and a fast-fading model based on the Rayleigh fading channel.

## (4) Synthesis Algorithms [19]

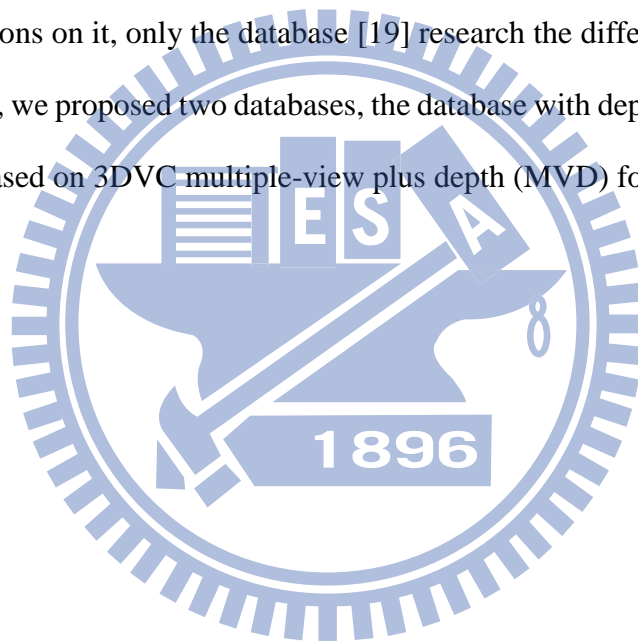
IRCCyN/IVC DIBR image database proposed by Bosc et al. [19] contains three different MVD sequences. And seven DIBR algorithms processed these three sequences to generate four

new viewpoints for each sequence.



Figure 8. Different DIBR algorithms for *lovebird* sequence [14].

The MPEG is in the process of specifying the 3D video coding (3DVC), but few databases discuss the distortions on it, only the database [19] research the different synthesis algorithms. Thus, in this thesis, we proposed two databases, the database with depth distortion and that with color distortion, based on 3DVC multiple-view plus depth (MVD) format.



# Chapter 4 Proposed 3D Quality Assessment Model for Depth Map Distortion

## 4.1 3D Image Database with Depth Map Distortion

We produce a database that contains three types of depth map distortions, *Offset*, *Quantization*, and *Gaussian\_Noise* as show in Figure 9. The distortions are added symmetrically to the left and right depth maps of input sequences. Then, the virtual view images are synthesized using the original images and distorted depth maps. Three types of distortion are described as follow:

### (1) *Offset*

The *Offset* operation adds a constant value to the entire depth map, that is,

$$d_{offset}(x, y) = d_{original}(x, y) + offset\_value \quad (22)$$

where  $d_{original}(x, y)$  and  $d_{offset}(x, y)$  denote the depth values of the original depth map and distorted depth map at  $(x, y)$ , respectively; and the *offset\_value* is a selected integer.

### (2) *Quantization*

The *Quantization* operation converts the depth value to a few specific levels. Usually, this is done by dividing the original depth value by a selected Quantization step.

$$d_{quantization}(x, y) = \left\lfloor \frac{d_{original}(x, y)}{quantization\_step} \right\rfloor \cdot quantization\_step \quad (23)$$

where  $d_{quantization}(x, y)$  is the depth value at  $(x, y)$  after the Quantization operation is applied to the depth map; floor operation  $\lfloor x \rfloor$  means the nearest integer smaller than or

equal to  $x$ ; and  $quantization\_step$  is a selected integer.

### (3) *Gaussian\_Noise*

It adds white Gaussian noise with zero mean and  $\sigma^2$  variance to the depth map.

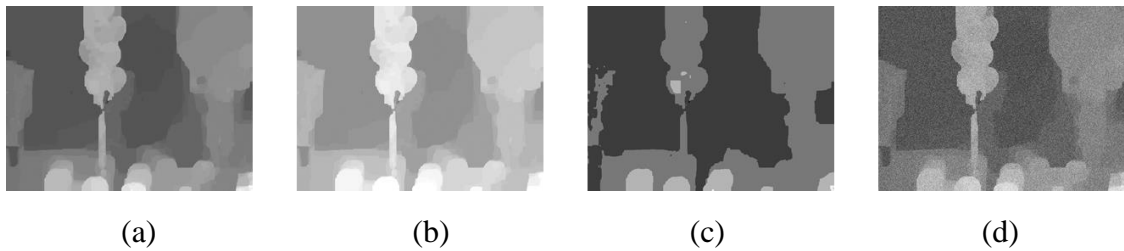


Figure 9. (a) Original depth map, (b) *Offset*, (c) *Quantization*, and (d) *Gaussian\_Noise*

## 4.2. Overview of Proposed 3D IQA Model

Our proposed 3D IQA model, shown in Figure 10, can be divided into two parts. One is computing the Image Quality Score (IQS) with shift compensation which is a block-matching method. And the other part is generating the Edge Structural Distortion (ESD) by computing the Hausdorff distance. IQS evaluates the visual quality of the distorted image, and ESD estimates the degree of structural artifact. The final score of the proposed model is obtained by combining these two scores together in the pooling stage. In the following section, we will elaborate the details of IQS and ESD operations. To evaluate our model, a comparison system is constructed as Figure 11. The distortion system uses three types of distortion described in the previous section.

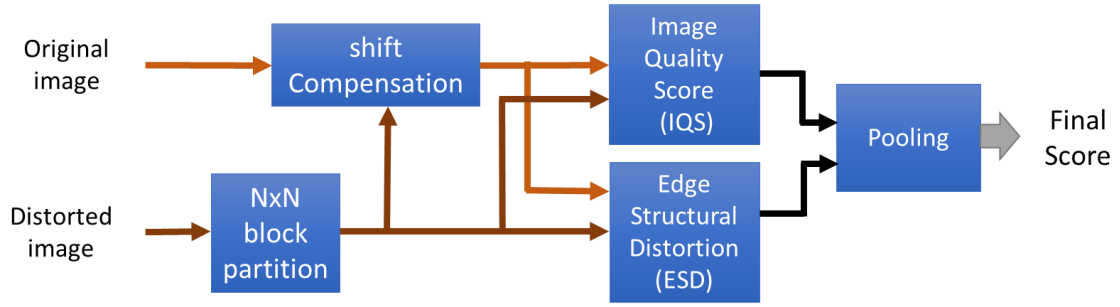


Figure 10. Flow chart of the proposed IQA model.

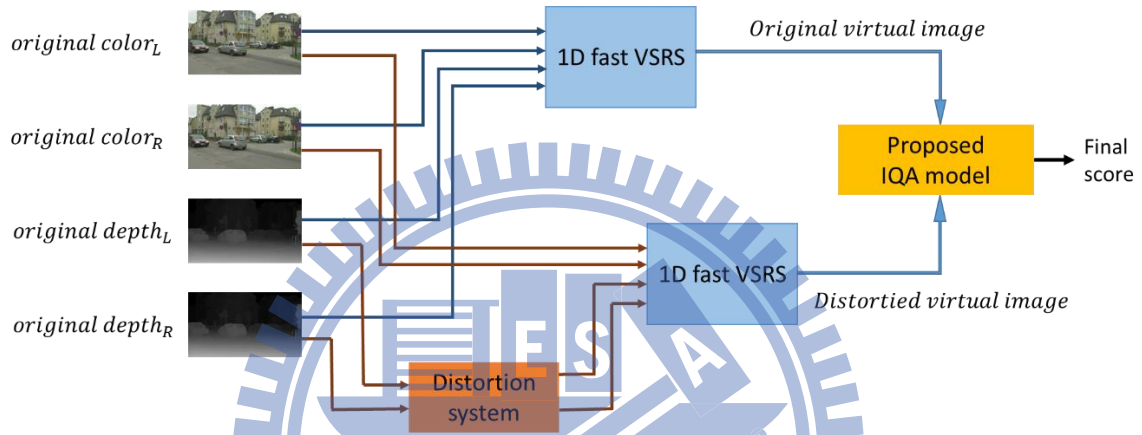


Figure 11. The proposed comparison system.

### 4.3. Image Quality Score (IQS)

In a synthesized image, the object shift is a phenomenon caused by an object rendered with an error depth value, which then produces a different disparity than its correct disparity. However, unlike 2D distortion, the object shift does not change subjective perception, but it may change the observer's 3D perception. Figure 12 is an example; the images (b) and (d) are synthesized by original depth maps like Figure 12(a) and depth maps which have *Quantization* distortion as shown in Figure 12(c) respectively. And Figure 12(d) leads to the object's horizontal shift relative to Figure 12(a). In fact, the subjective qualities of Figure 12(d) and Figure 12(b) are almost the same, but the SSIM index of Figure 12(d) is 0.701. This example indicates two facts. The first

is the object shift does not affect subjective image quality perception. The second is the conventional 2D QA models applied to pictures with the object shift like Figure 12(d) would result poor quality scores. The conventional pixel based 2D QA models penalize the object shifts as shown in Figure 13(a), where darker regions indicate lower quality index.

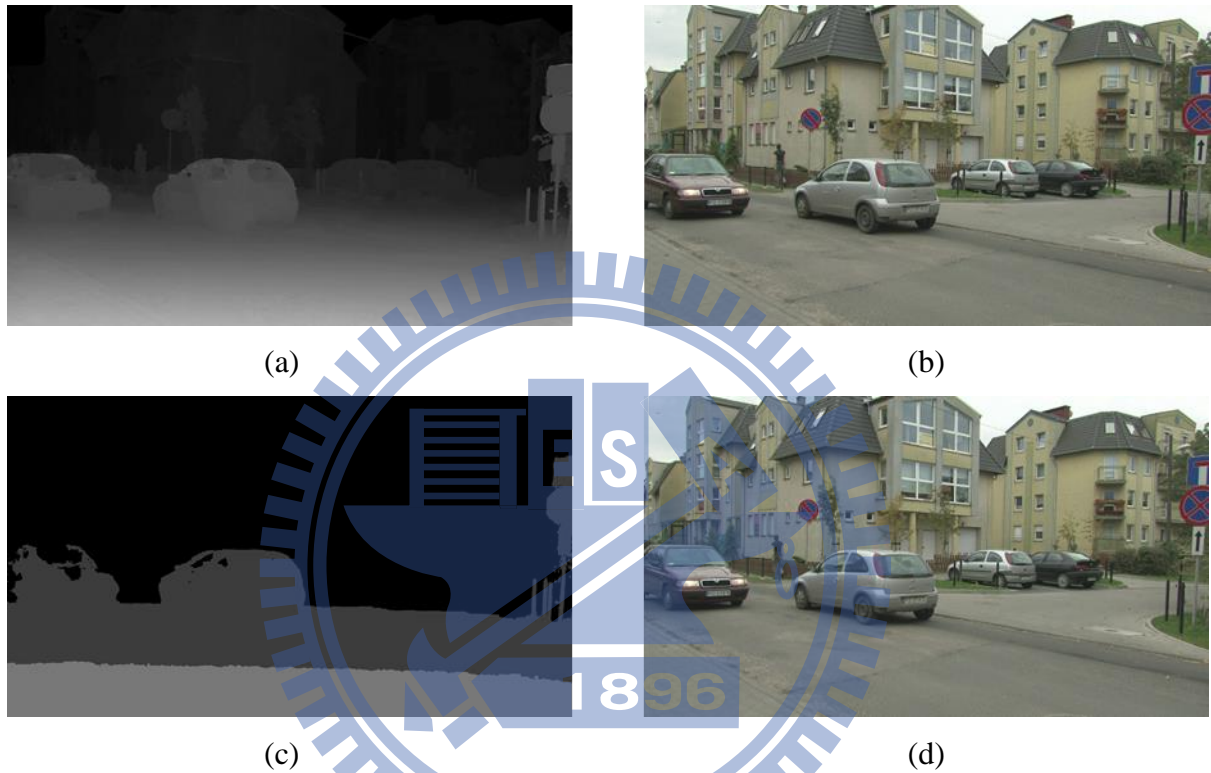


Figure 12. (b) is the virtual view synthesized using the original depth map (a). (d) is the virtual view synthesized using quantized depth map (c).

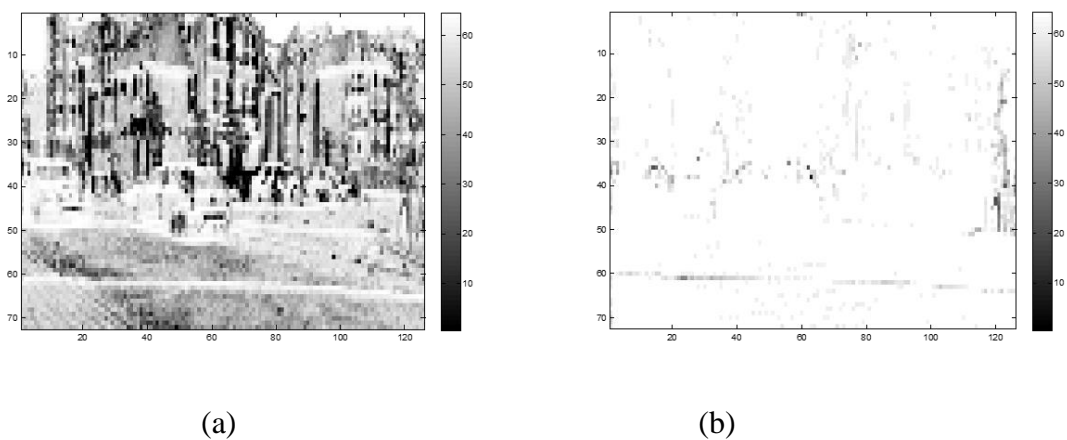


Figure 13. Quality index map evaluated on Figure 12(b) using Figure 12(d). (a) SSIM index map and (b) SSIM index map with object shift.



For this reason, we shall consider object shifts before performing 2D quality evaluation. We use a block matching algorithm to do this. The shift vectors are found by matching an  $N \times N$  block of distorted image ( $D_i$ ) with the corresponding block ( $R_i$ ) in the original image. Then, we compute the SSIM index between  $D_i$  and  $R_i$  to be the IQS of the  $i$ th block as equation (24). Therefore, the flow chart of computing IQS is shown in Figure 14.

$$IQS_i = SSIM(D_i, R_i) \quad (24)$$

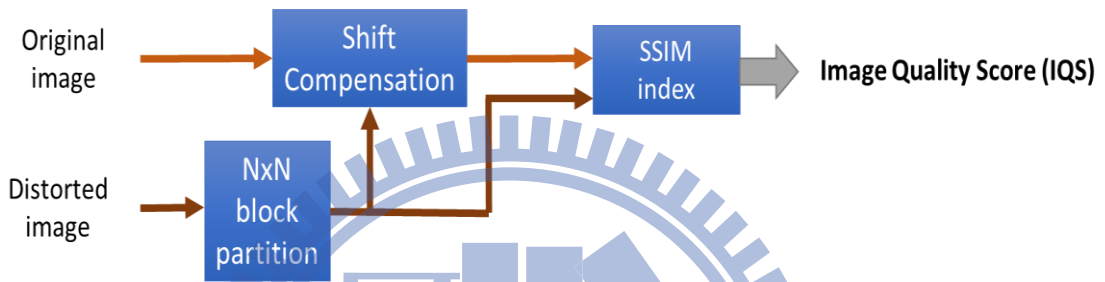


Figure 14. Flow chart of computing the image quality score (IQS).

In the example shown in Figure 12 and Figure 13, the subjective score of Figure 12(d) synthesized by the distorted depth map is higher than that of Figure 12(b), which is synthesized by the original depth map. But in Figure 13(b), there are some low quality regions, as mark in Figure 15, which decrease the overall IQS of Figure 12(d). And then, these regions impair the correlation of the objective scores to the subjective scores. One of these regions is enlarged as shown in Figure 16. The house window in Figure 16(a) is not identical to the window in Figure 16(b). The difference is introduced by the rendering process due to depth map errors. Incorrect depth values let synthesized virtual view pick the neighbor color information of the pixel that should be project. Therefore, subjectively, both windows have similar subjective quality. In contrast, the pixel-differences are penalized by the traditional 2D IQA model. Thus, we apply a Gaussian filter before computing SSIM index as shown in Figure 17, and it results an IQS index map, Figure 18.

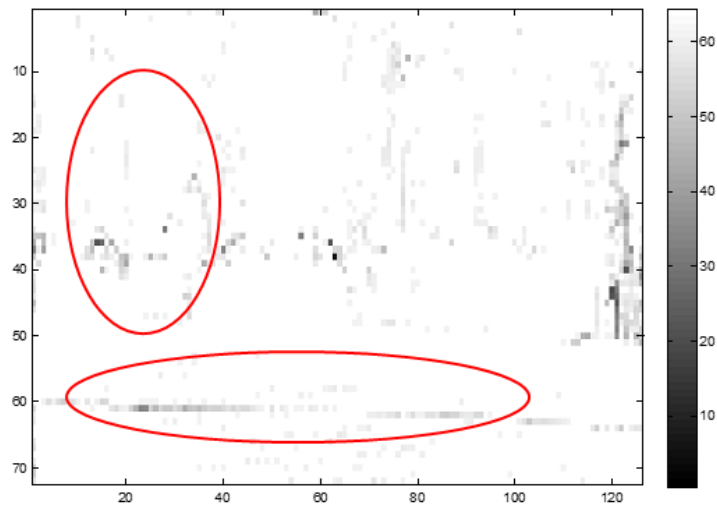


Figure 15. Some low IQS region of Figure 13.

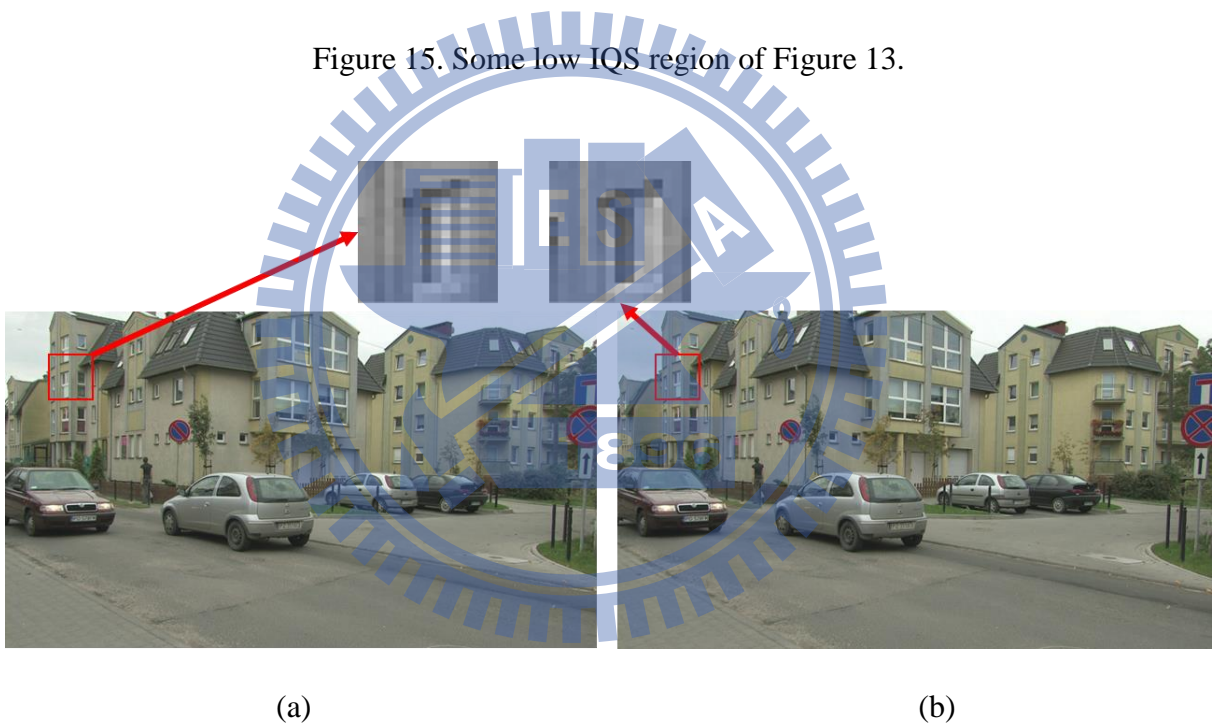


Figure 16. (a) Magnified from Figure 12(b); (b) Magnified from Figure 12(d).

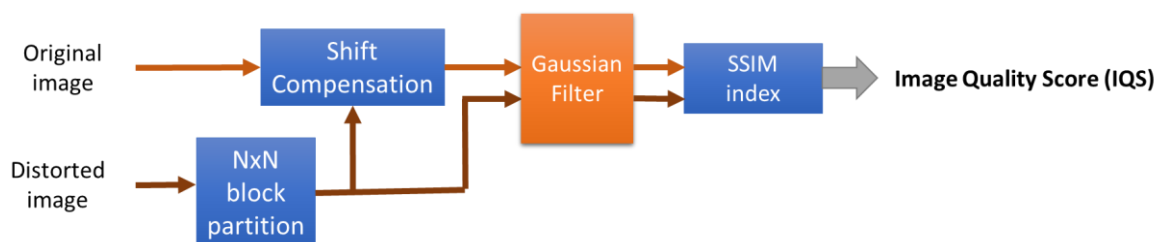


Figure 17. Modified flow chart of computing IQS.

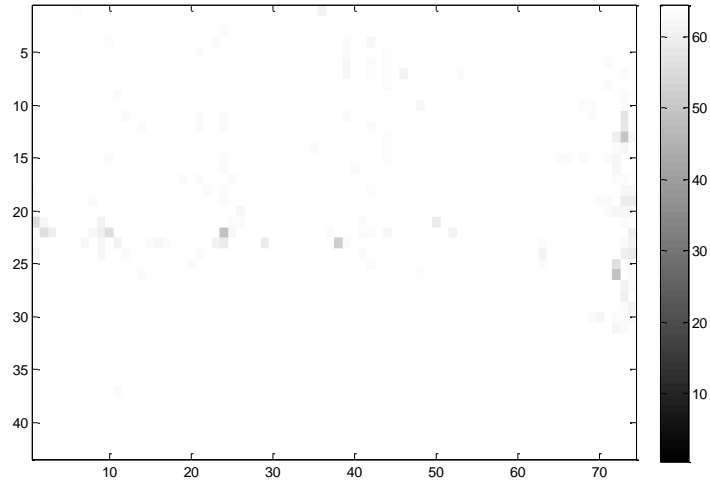


Figure 18. The IQS index map of Figure 13 using the process shown in Figure 17.

#### 4.4. The Hausdorff Distance

The Hausdorff distance is used to measure the degree of mismatching of two sets. In the computer vision, this distance can be used to find a given model in an image [20] as shown in Figure 19. The Hausdorff distance of two finite sets,  $A$  and  $B$ , is defined as follows,

$$H(A, B) = \max(h(A, B), h(B, A)), \text{ and} \quad (25)$$

$$h(A, B) = \max_{a \in A} d(a, B) \quad (26)$$

where  $h(A, B)$  is called the directed Hausdorff distance from  $A$  to  $B$ ; and  $d(a, B)$  is the shortest distance from a point  $a$  in set  $A$  to point set  $B$ , and the definition of distance is application-dependent. For example, in Figure 20, the directed Hausdorff distance  $h(A, B)$  is equal to  $d(a_1, B)$ .

We can generalize the definition of directed Hausdorff distance  $h(A, B)$ . The modified directed Hausdorff distance  $h_K(A, B)$  is given by considering the  $K^{th}$  ranked point of  $A$ . [21]

$$h_K(A, B) = \max_{a \in A}^{pK^{th}} d(a, B), \text{ and} \quad (27)$$

$$\frac{K}{N_A} = p\% \quad (28)$$

where  ${}^p K_{a \in A}^{th}$  denotes the  $K^{th}$  ranked distance in A, K is a selected parameter; and  $N_A$  is the number of points in set A. When  $p=50$ , in the example shown in Figure 20, the modified directed Hausdorff distance  $h_K(A, B)$  is equal to  $d(a_2, B)$ ; when  $p = 100$ , all the points are under considering, and the  $h_K(A, B)$  equals the directed Hausdorff distance  $h(A, B)$ .

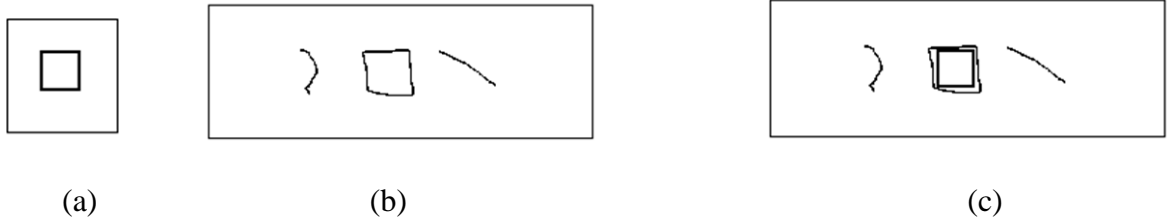


Figure 19. (a) A specified model (b) a test image (c) the matched result

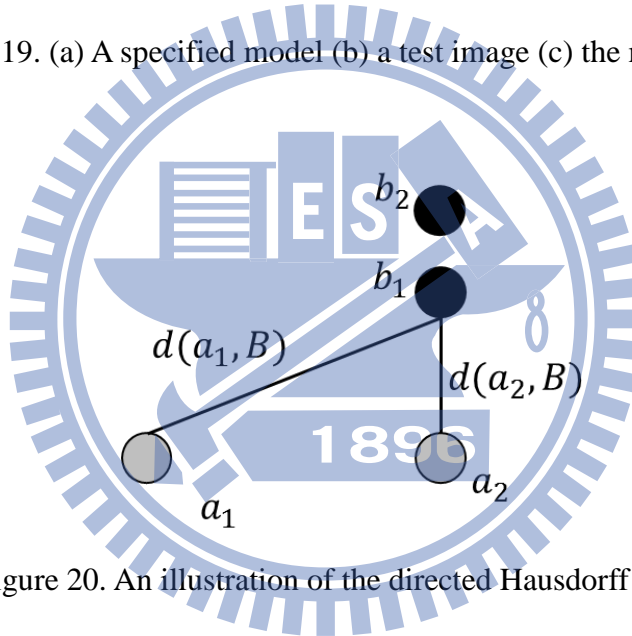


Figure 20. An illustration of the directed Hausdorff distance.

#### 4.5. Edge Structural Distortion (ESD)

Due to human eyes are sensitive to distortion around object edge, for example, discontinuous structure on object boundary or ghost artifact which foreground textures are rendered to background area. We thus proposed a special metric to detect object edge distortion. We first apply the Canny edge detector to block  $D_i$  and  $R_i$ . Then, we compute the Hausdorff distance of the two blocks to measure the edge structural distortion.

$$H(D_i, R_i) = \max(h(D_i, R_i), h(R_i, D_i)) \quad (29)$$

The distance function used in our model is defined as follows,

$$d(a, b) = |x_a - x_b| + |y_a - y_b| \quad (30)$$

where  $x_a$  and  $y_a$  are the x and y coordinate of the point a .

After obtaining the Hausdorff distance of each block, we normalize the distance between 0 and 1.

$$H_{normalize}(D_i, R_i) = \frac{H(D_i, R_i)}{2N} \quad (31)$$

where  $N$  is the width of block window.

At the end, smaller Hausdorff distance should indicate less structural distortion. Therefore, the normalized Hausdorff distance is subtracted from 1 to produce the correct score.

$$ESD_i = 1 - H_{normalize}(D_i, R_i) \quad (32)$$

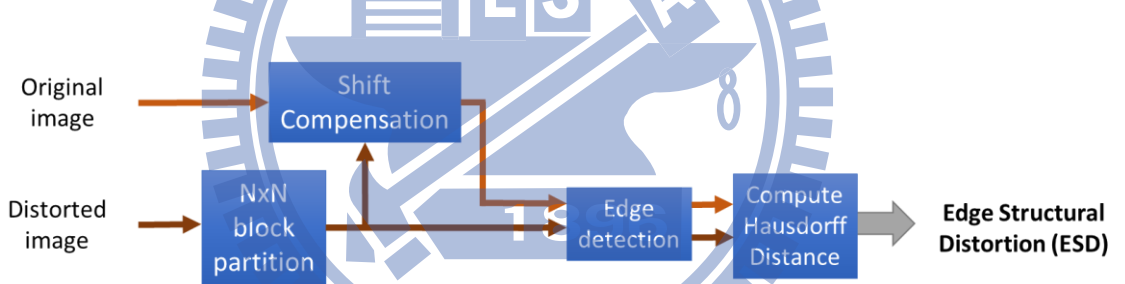


Figure 21. Flow chart of computing ESD.

## 4.6. Pooling

In the pooling stage, image quality score (IQS) and edge structural distortion (ESD) are combined together. There is no evidence shows the two scores, IQS and ESD, should operate in what operator such as multiplication and addition. In our proposed model, we use addition not multiplication. Because multiplication includes the concept of “weight”. Since IQS measures conventional 2D distortions such as blur, noise, compression coding et al. and ESD estimates the degree of structural artifact. IQS and ESD do not have strong relation that we can

treat one to be the weight of the other one. Thus the final quality score can be computed as follows,

$$\text{score} = \sum_{i=1}^B (\alpha \cdot \text{IQS}_i + (1 - \alpha) \cdot \text{ESD}_i) \quad (33)$$

where  $B$  is the total block number.

Moorthy and Bovik [22] suggest that the performance of using the lowest  $p\%$  quality scores in the pooling stage is better than using the average value of all scores. For example, there are many local low quality regions in Figure 22(a) and it has a lower subjective perception than Figure 22(b). But the mean SSIM score of Figure 22(a) is higher than the mean SSIM score of Figure 22(b).

In the virtual view images with distorted depth maps, object shifts and ghost artifacts often occur in specific region, especially the occlusion regions. Hence, using the lowest  $p\%$  quality scores can provide more precise estimation than the average score of the frame, and the flow chart of pooling stage is shown as Figure 23.

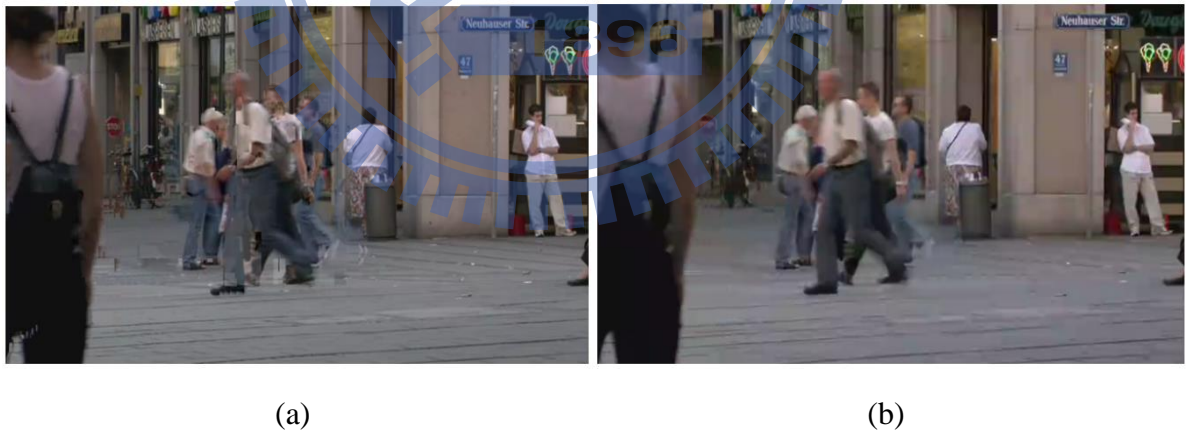


Figure 22. Example of lower quality picture in using the mean as the final score.

(a) mean SSIM = 0.9497 (b) mean SSIM = 0.9043 [23]

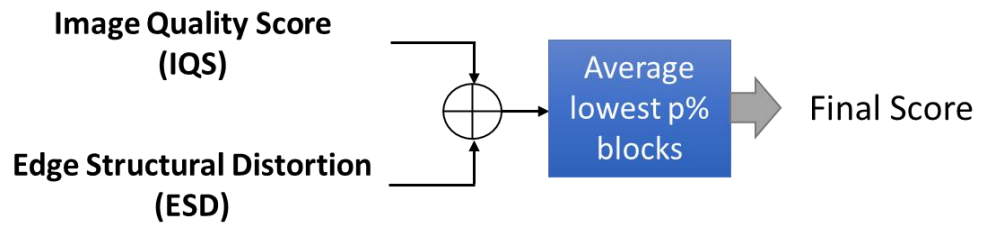


Figure 23. Flow chart of the pooling stage.



# Chapter 5 Experimental Results of Depth Map Distortion Database

## 5.1. Test Image Database Construction

All our test sequences are obtained from the ISO MPEG standard committee [24]. They are *Poznan\_Hall2* and *Poznan\_Street* provided by Poznan University [25]; *Kendo* and *Balloons* provided by Nagoya University [26]; *Lovebird1* provided by ETRI [27]; and *Newspaper* provided by GIST [28]. The sequences are used for experiment in the MPEG 3DVC contest. Their color images and associated depth maps are shown in Table 2. The processed frame numbers, input views, and output displayed stereo are shown in Table 3. They pretty much follow the specification.

The comparison system is illustrated in Figure 11. We choose one frame in each test sequence, then distort its left and right depth maps and render with undistorted color images to obtain *distorted images*. There are three types of distortion, *Offset*, *Quantization*, and *Gaussian Noise*, defined in section 4.1. The *Offset\_value* we used is 60 and 100; *Quantization\_step* is 60 and 80; and the variance of Gaussian white noise is 0.01 and 0.05. So there are 7 ( $2*3+1$ ) test images per sequence. Totally, there are 42 ( $7*6$ ) stereo-pairs in our database. All the virtual color images are synthesized by the “1D-fast-VSRS” algorithm in the ITU/MPEG JCT-3D reference software HTM 3.1 [29].



Table 2. Test sequences.





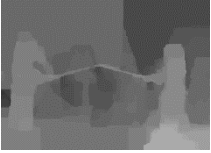





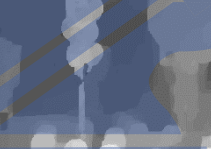

Poznan_Hall2 (1920x1088)	Kendo (1024x768)	Lovebird1 (1024x768)
		
		
Poznan_Street (1920x1088)	Balloons (1024x768)	Newspaper (1024x768)
		
		

Table 3. The information of the test sequences.

Sequence Name	Frame	Input Views	Output Stereo Pair
Poznan_Hall2	90	7 - 6	6.5 - 6
Poznan_Street	30	4 - 3	3.5 - 3
Kendo	32	3 - 5	4 - 5
Balloons	1	3 - 5	4 - 5
Lovebird1	80	6 - 8	7 - 8
Newspaper	100	4 - 6	5 - 6

## 5.2. Subjective Test

In our experiment, the Toshiba 47TL515U 47-inch 3D monitor is used to display the materials. Twenty-two observers (20 males and 2 female), whose average age is 23.8, participated in our experiment. The experiment method we adopted is single stimulus (SS) specified by ITU-R BT.500 [6]. For each test image, the observers are asked to give a quality score (5: Excellent; 4: Good; 3: Fair; 2: Poor; 1: Bad), which is called *opinion score*. The *mean opinion score* (MOS) and variance of all test images are shown in Figure 24 and Figure 25.

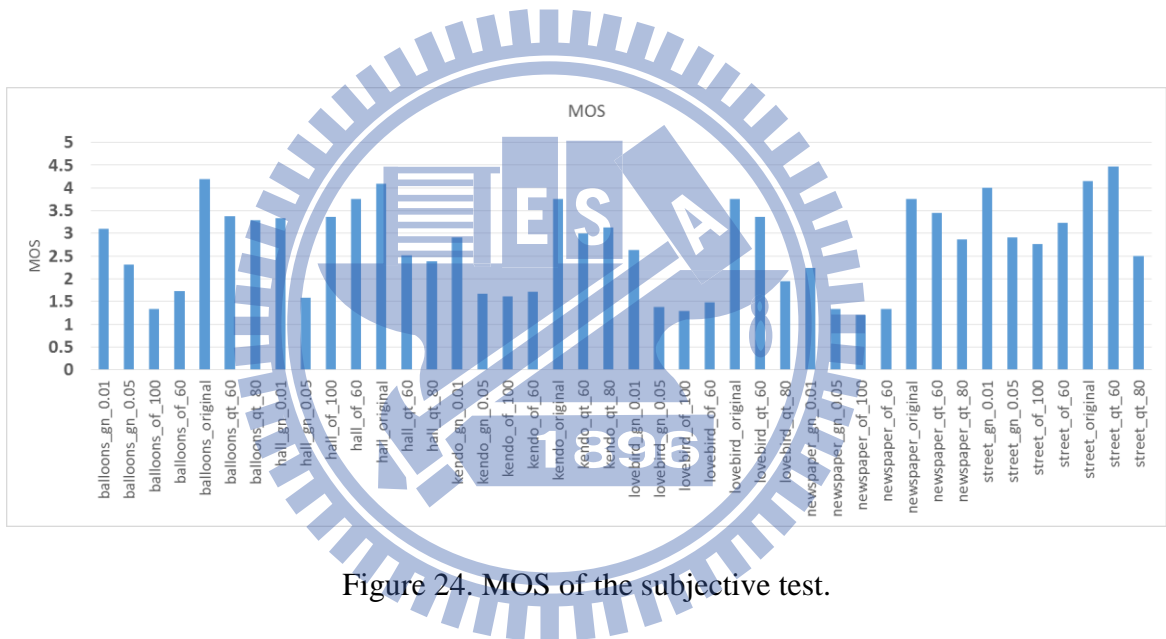


Figure 24. MOS of the subjective test.

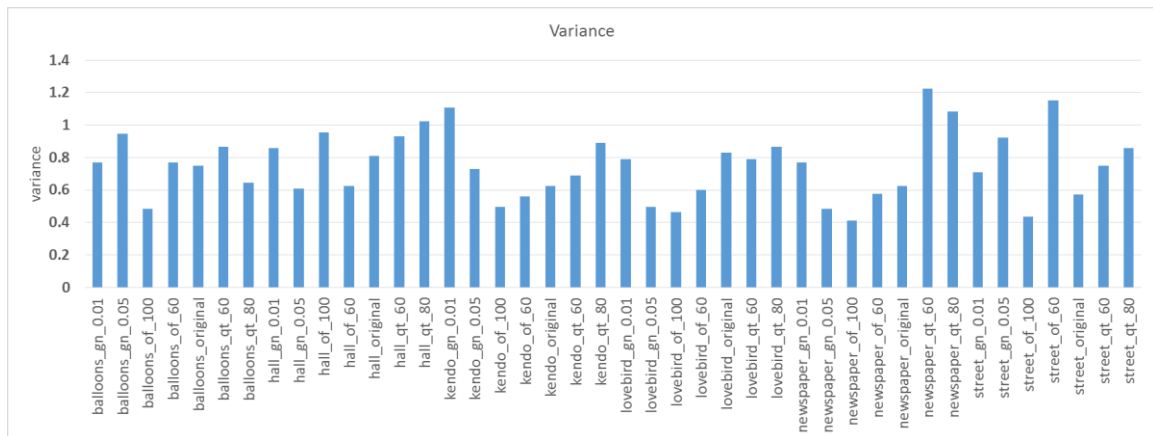


Figure 25. Variance of the subjective test.

### 5.3. Parameters of the Model

Figure 26 shows the operations and all the parameters in our proposed model.

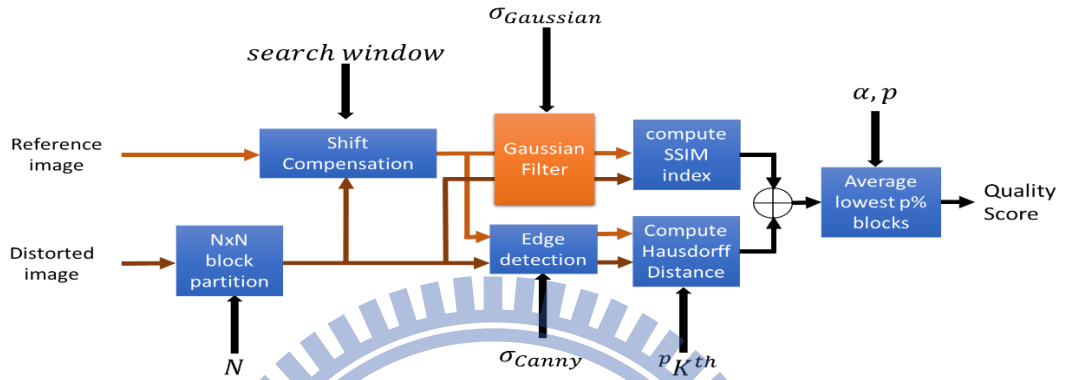


Figure 26. Parameters of the proposed model.

From the front stage to the last stage, parameters are described as follows:

- (1)  $N$  is the width of a block.
- (2)  $search\_window$  specifies the maximum value of shift vector in parallel block search. For example, in Figure 27, if the center of the processing block in distorted image is  $(x_{dis}, y_{dis})$  with  $N$  is 5 and  $search\_window$  is 6. Then the search area in reference image is an  $N$ -by- $(N+2*search\_window)$  block with center  $(x_{dis}, y_{dis})$ .

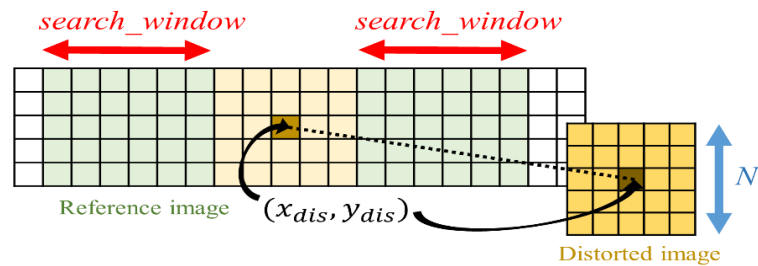


Figure 27. An example of shift compensation with  $search\_window=6$  and  $N=5$ .

- (3)  $\sigma_{Gaussian}$  is the standard deviation of the Gaussian filter with zero mean.

- (4) In the edge detection stage, the Canny method, using  $\sigma_{Canny}$  as the standard deviation of the Gaussian filter.
- (5)  $pK^{th}$  decides the number of points that are used to compute direct Hausdorff distance.

Figure 28 shows an example with  $pK^{th}$  equals to 100, 80, and 60.

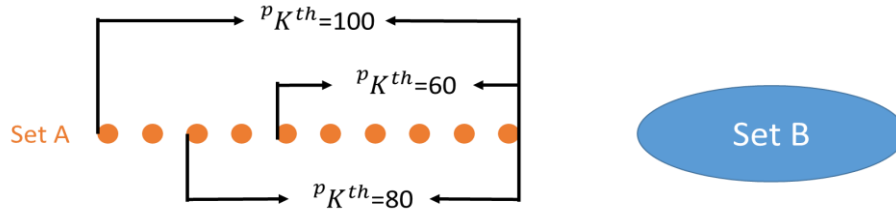


Figure 28. An example of  $pK^{th}$ .

- (6)  $\alpha$  decides the proportion of IQS and ESD, as shown in Figure 29.

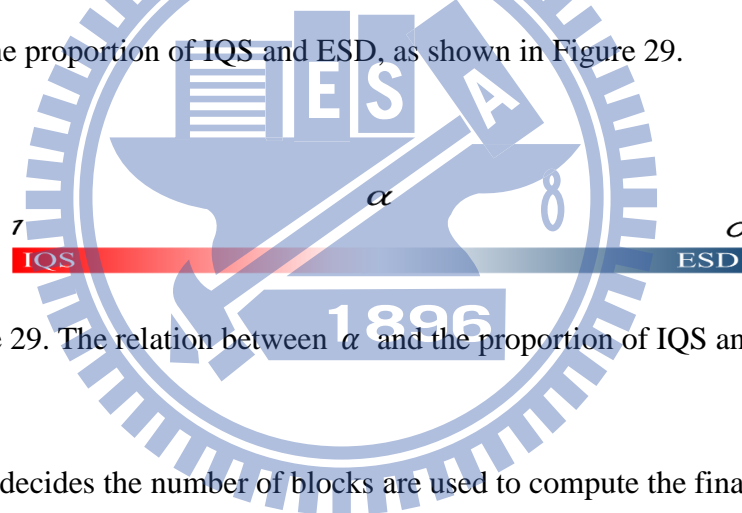


Figure 29. The relation between  $\alpha$  and the proportion of IQS and ESD.

- (7) Parameter  $p$  decides the number of blocks are used to compute the final quality score.

We will discuss some of these parameters later. Due to the shift vector is often short, the *search\_window* is set to 20 pixels, which is typically sufficient. And  $\sigma_{Canny}$  is relative to  $pK^{th}$  because larger  $pK^{th}$ , which indicates more edge points are considered, means smaller  $\sigma_{Canny}$ .

### 5.3.1. Pooling Proportion ( $p$ )

Parameter  $p$  in the pooling stage decides how many blocks are used to compute the final quality

score. The gray curve with circular patterns in Figure 30 shows that generally  $p=5\%$  leads to the highest PLCC for our database. That is, using 5% worst case blocks in an image best matches the subjective image quality. But when split our database into two parts: (1) *Gaussian\_Noise* and (2) *Offset + Quantization*, we notice that *Gaussian\_Noise* produces a rather blur-like artifacts on the entire synthesized image as shown in Figure 31. Therefore, as the orange curve with rectangular marks in Figure 30, we need to consider whole image ( $p=100\%$ ) to get a better performance. For *Offset* and *Quantization*, Figure 32 shows that if the depth map error produces structural artifacts such as object shift or ghost image, the pooling proportion  $p$  should be chosen near 5%, and the result is almost independent to  $\alpha$ .

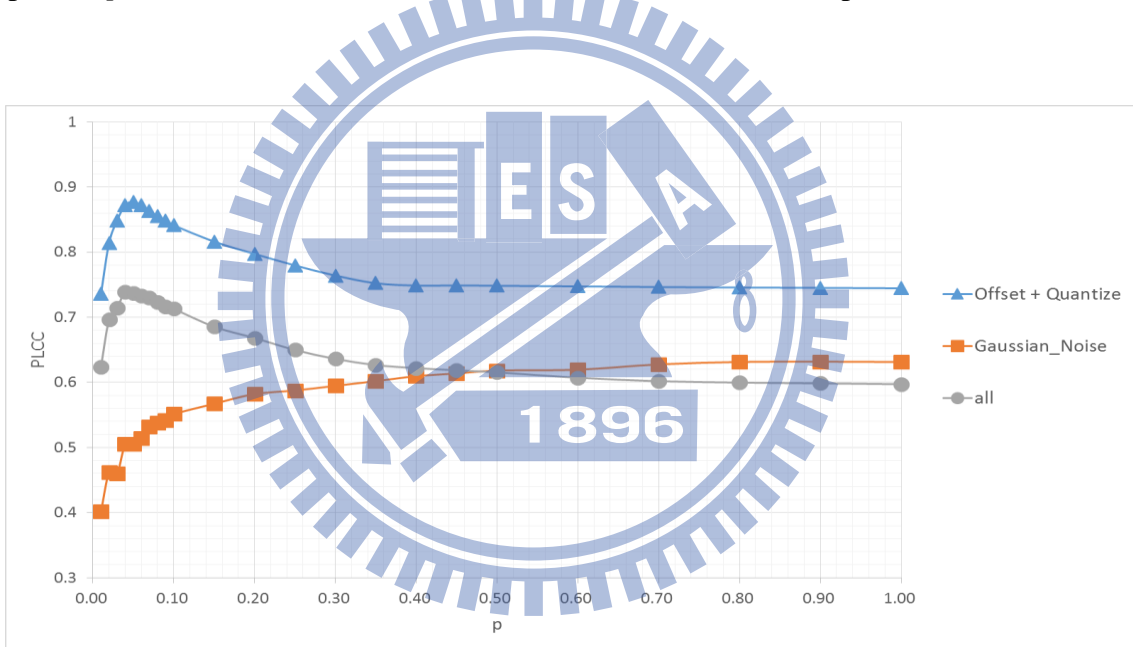


Figure 30. Different value of the pooling proportion  $p$  affects PLCC.

$$(\alpha = 0.5, N = 25, p_{K^{th}} = 70)$$

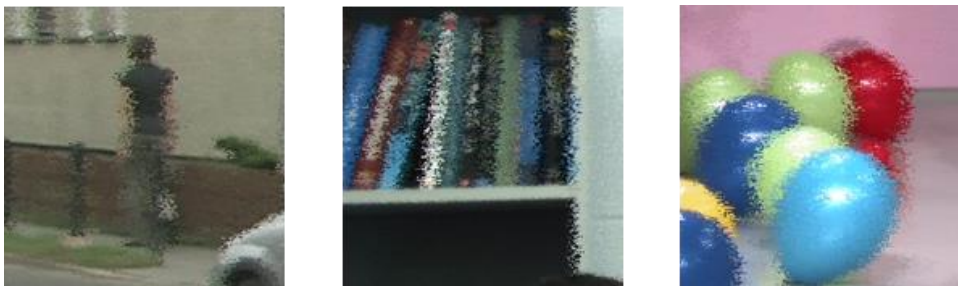


Figure 31. Artifact due to *Gaussian\_Noise* depth map error.

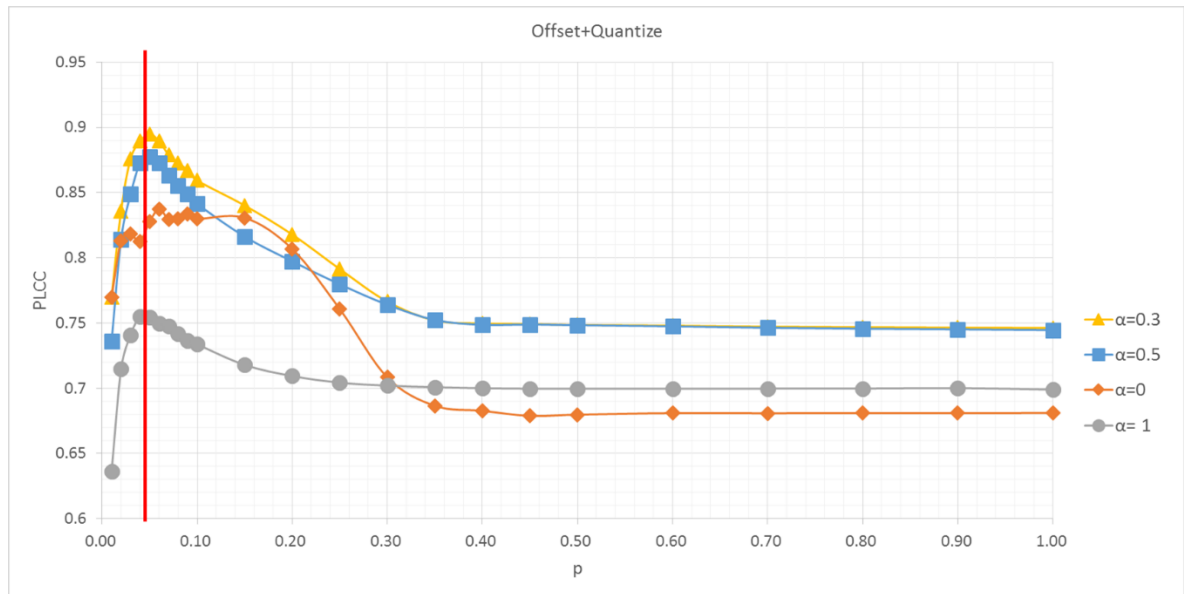


Figure 32. Pooling proportion  $p$  nearby 5% has better performance under different pooling weight  $\alpha$ . ( $N = 25$ ,  $pK^{th} = 70$ )

Although the *Gaussian\_Noise* distortion affects the whole color image, but from the subjective test, we found that the observers did not notice that the entire image was distorted in stereoscopic test. This phenomenon is discussed in [31], which states that the high frequency regions the subjective quality. For example, in Figure 33, if the left view of the stereo pair shows the left rectangle with horizontal high frequency as the original image. The experimental results show that the observers can still perceive the texture pattern of the left rectangle when the other view of the stereo pair shows the middle rectangle with low frequency as the distorted image. If the right view shows the right rectangle with vertical high frequency, the two patterns would interfere each other. Another illustration is shown in Figure 34. If the left view of the stereo pair shows the left rectangle with low frequency texture as the reference image and the right view shows the right rectangle with block effect, then observers can perceive the block artifacts in the stereoscopic test.



Figure 33. Illustration of eye domination. [31]

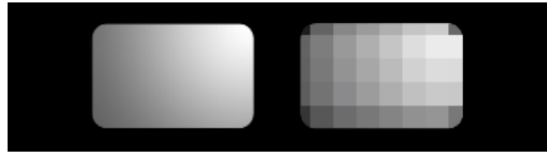


Figure 34. Another illustration of eye domination. [31]

### 5.3.2. Pooling weight ( $\alpha$ )

In the equation (33),  $\alpha$  decides the weights of IQS and ESD. If  $\alpha > 0.5$ , it means we consider IQS more than ESD, and vice versa. From the previous sections, we find out that to evaluate *Offset* and *Quantization*, we choose  $p=5\%$ ; and for *Gaussian\_Noise*, we pick  $p=100\%$ . The blue curve with triangular patterns in Figure 35 shows that ESD has a higher impact on the PLCC index than IQS in the case of the *Offset* and the *Quantization*. And the best parameter value is  $\alpha = 0.3$ . For the *Gaussian\_Noise*, there is no apparent difference between IQS and ESD as shown as the orange curve with circular patterns in Figure 35, and IQS is slightly more important than ESD. And generally  $\alpha = 0.5$  leads to the higher performance for our database. Figure 36 shows that, for all possible values of the pooling proportion  $p$ , IQS provides a higher PLCC. In the later experiment, we use  $\alpha = 1$  to test *Gaussian\_Noise*.

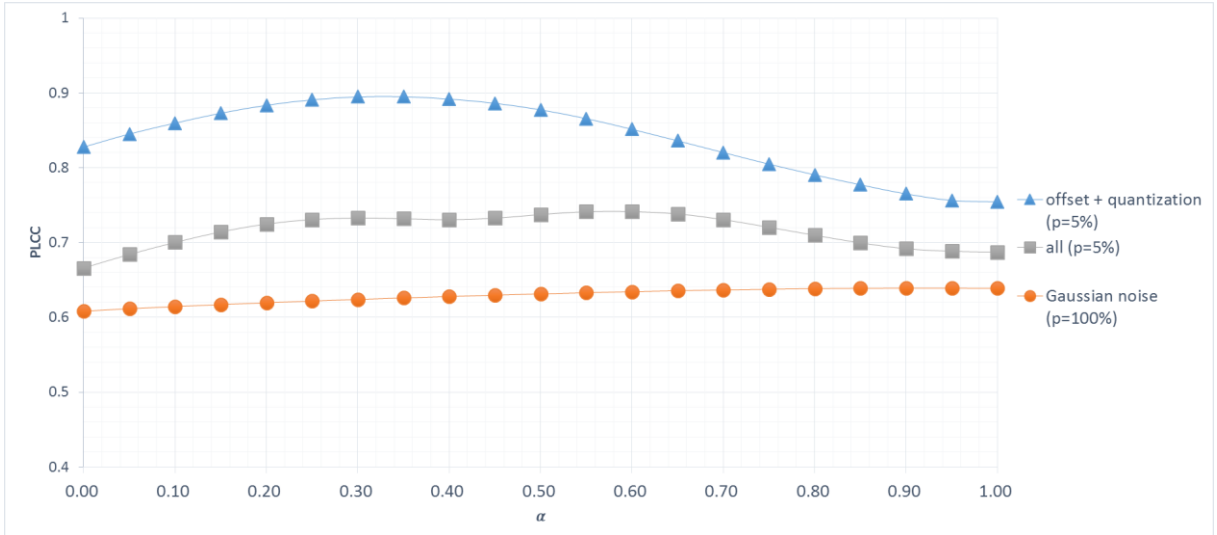


Figure 35.  $\alpha$  against PLCC. ( $N = 25$ ,  $pK^{th} = 70$ )

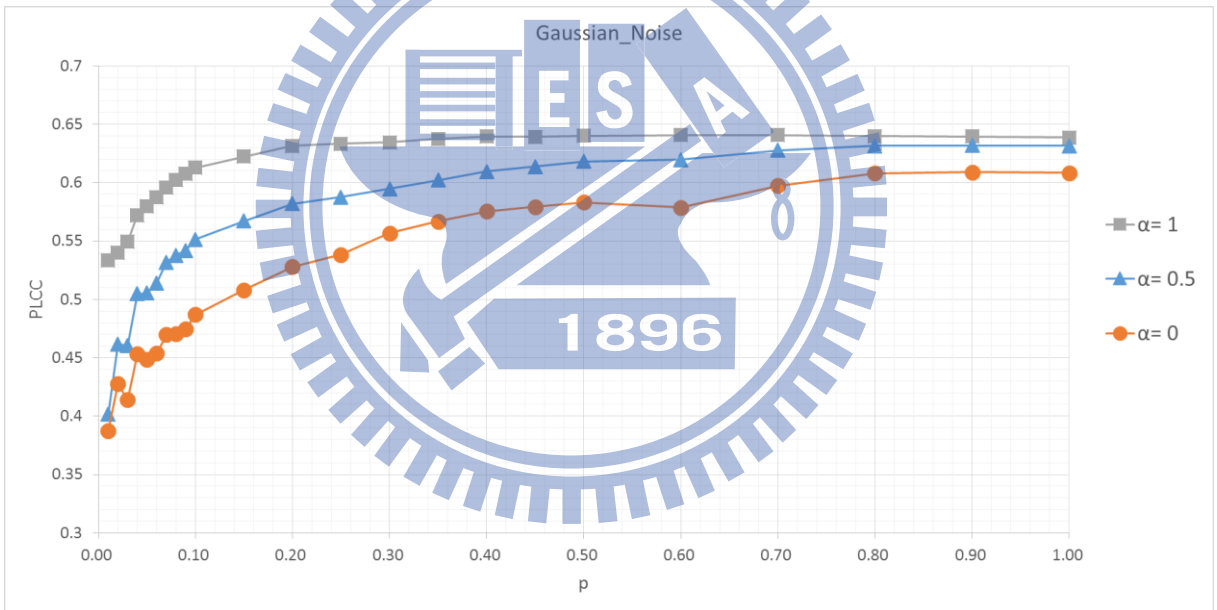


Figure 36. IQS dominates the performance in *Gaussian\_Noise* case. ( $N = 25$ ,  $pK^{th} = 70$ )

### 5.3.3. Modified Hausdorff Distance ( $pK^{th}$ )

Value of  $pK^{th}$  decides the number of ranked points that are used to compute ESD of each block.  $pK^{th} = 100$  means that all points are in used, and the result is the conventional Hausdorff Distance.



From the Figure 36, ESD has little influence on the final performance for the *Gaussian\_Noise* distortion type. Hence, in the case of *Gaussian\_Noise*, we neglect the comparison of  $pK^{th}$  in performance evaluation. For *Offset* and *Quantization*, Figure 37 and Figure 38 indicate that  $pK^{th}=70$  has nearly the best result in these two cases for  $\alpha = 0.5$  and  $\alpha = 0.3$ .

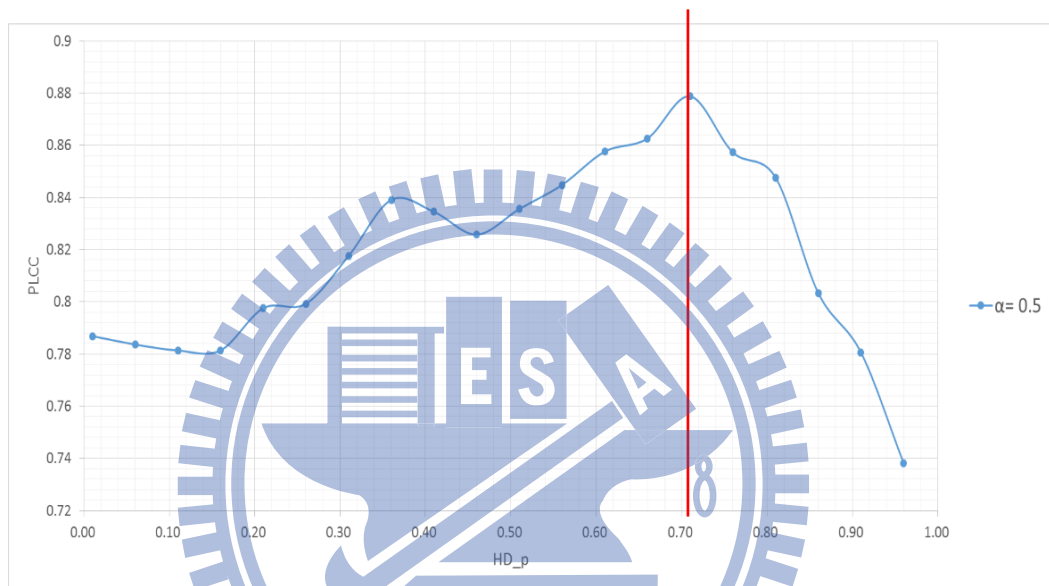


Figure 37. The effect of  $pK^{th}$  when  $\alpha = 0.7$ . (window = 25, pooling p = 5%)

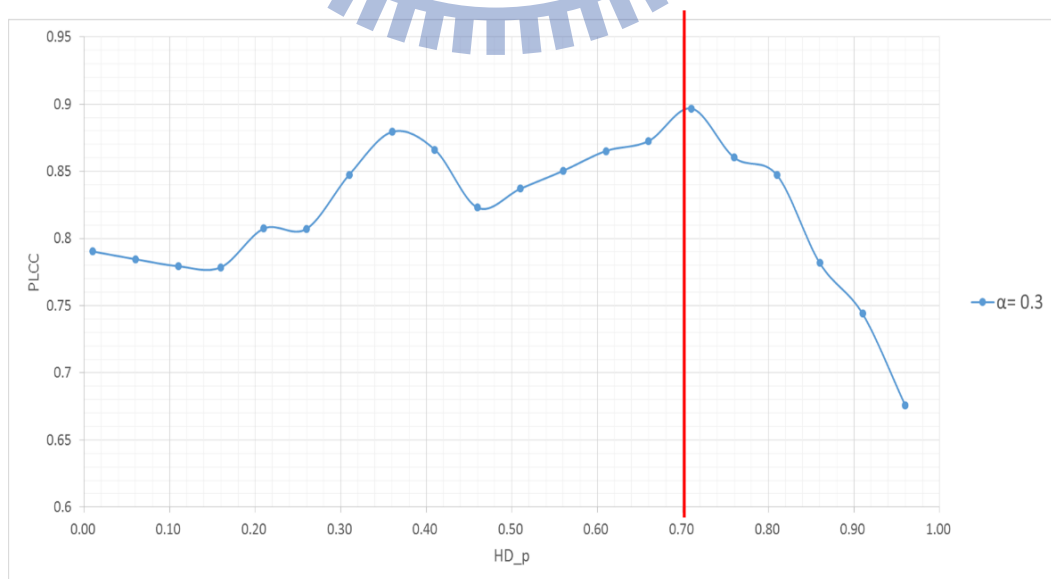


Figure 38. The effect of  $pK^{th}$  when  $\alpha = 0.3$ . (window = 25, pooling p = 5%)

One may notice that there is a peak near  $pK^{th} = 35$ . We discuss these two cases,  $pK^{th} = 35$  and  $pK^{th} = 70$ , against the value of  $\alpha$  as shown in Figure 39. Their performance is close between  $\alpha = 0.1$  and  $\alpha = 1$ . Figure 40 gives a good account of this phenomenon. Figure 40(a) is a virtual view image with ghost artifacts synthesized by a distorted depth map with *Offset* distortion (*offset\_value=60*). When  $pK^{th} = 70$ , the structural artifacts such as object edge distortion can be detected by our model as shown in Figure 40(b), but some non-structural distortion areas with rendering error are also observed. That is because  $pK^{th} = 70$  is a strict threshold condition. When  $pK^{th} = 35$ , as shown in Figure 40(c), structural distortions areas are not identified well as  $pK^{th} = 70$ , but it has a better performance for non-structural distortion areas.

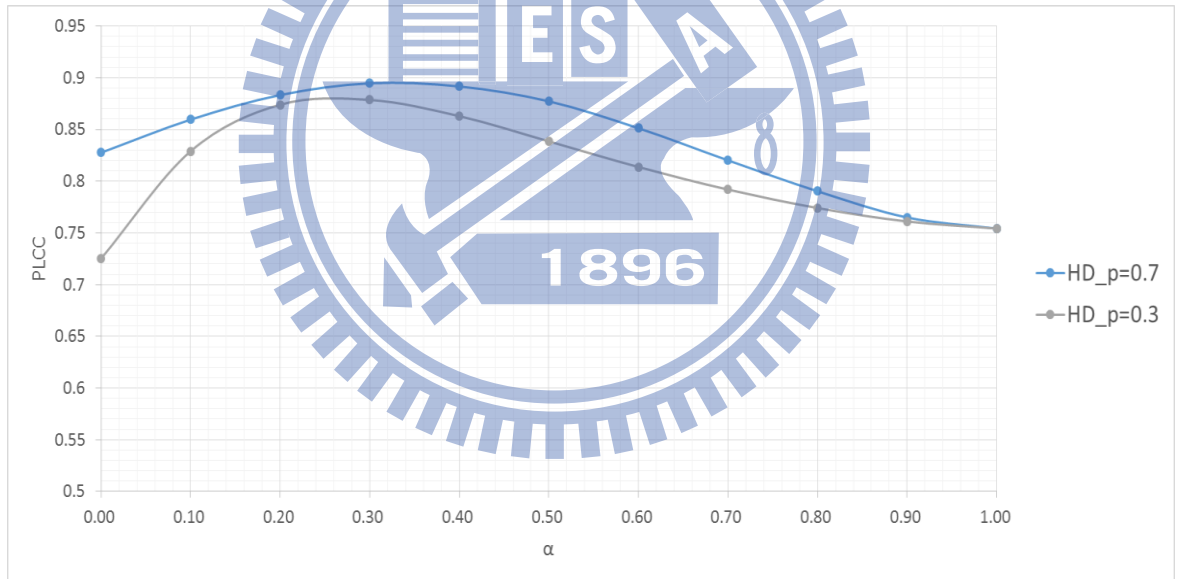


Figure 39.  $pK^{th} = 35$  and  $pK^{th} = 70$  have close performance when the value of  $\alpha$  is between 0.2 to 1. (window = 25, pooling p = 5%)



(a)

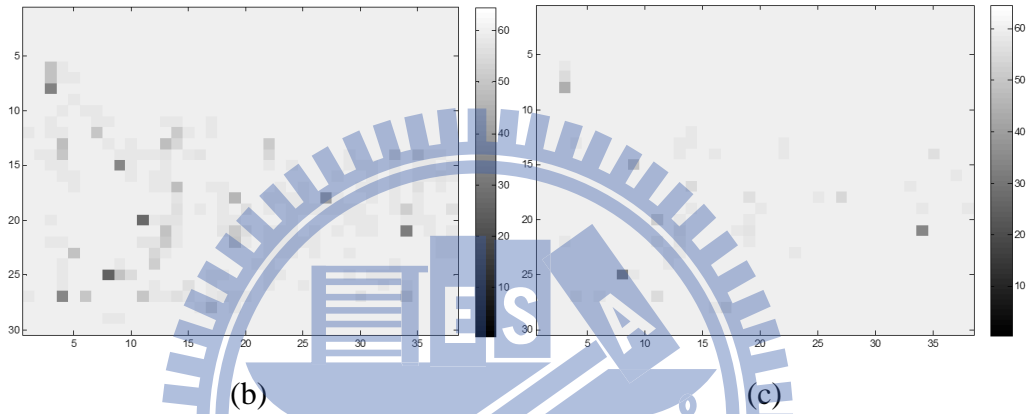


Figure 40. (a) Synthesized image of *kendo* with *Offset* distortion (*offset\_value*=60).  
 (b) and (c) are ESD maps where darker blocks indicate lower ESD index.

(b)  $p_K^{th} = 70$  (c)  $p_K^{th} = 35$ .

#### 5.3.4. Gaussian filter ( $\sigma_{Gaussian}$ )

The main purpose of Gaussian filter in the proposed model is to reduce incorrect judgment due to the image rendering process for view synthesis. The value of  $\sigma_{Gaussian}$  controls the blurring level. The  $\sigma_{Gaussian}$  value should be large enough to eliminate rendering error, but the details would be lost if it is too large. Figure 41 and Figure 42 show the different  $\sigma_{Gaussian}$  values and its corresponding IQS map.

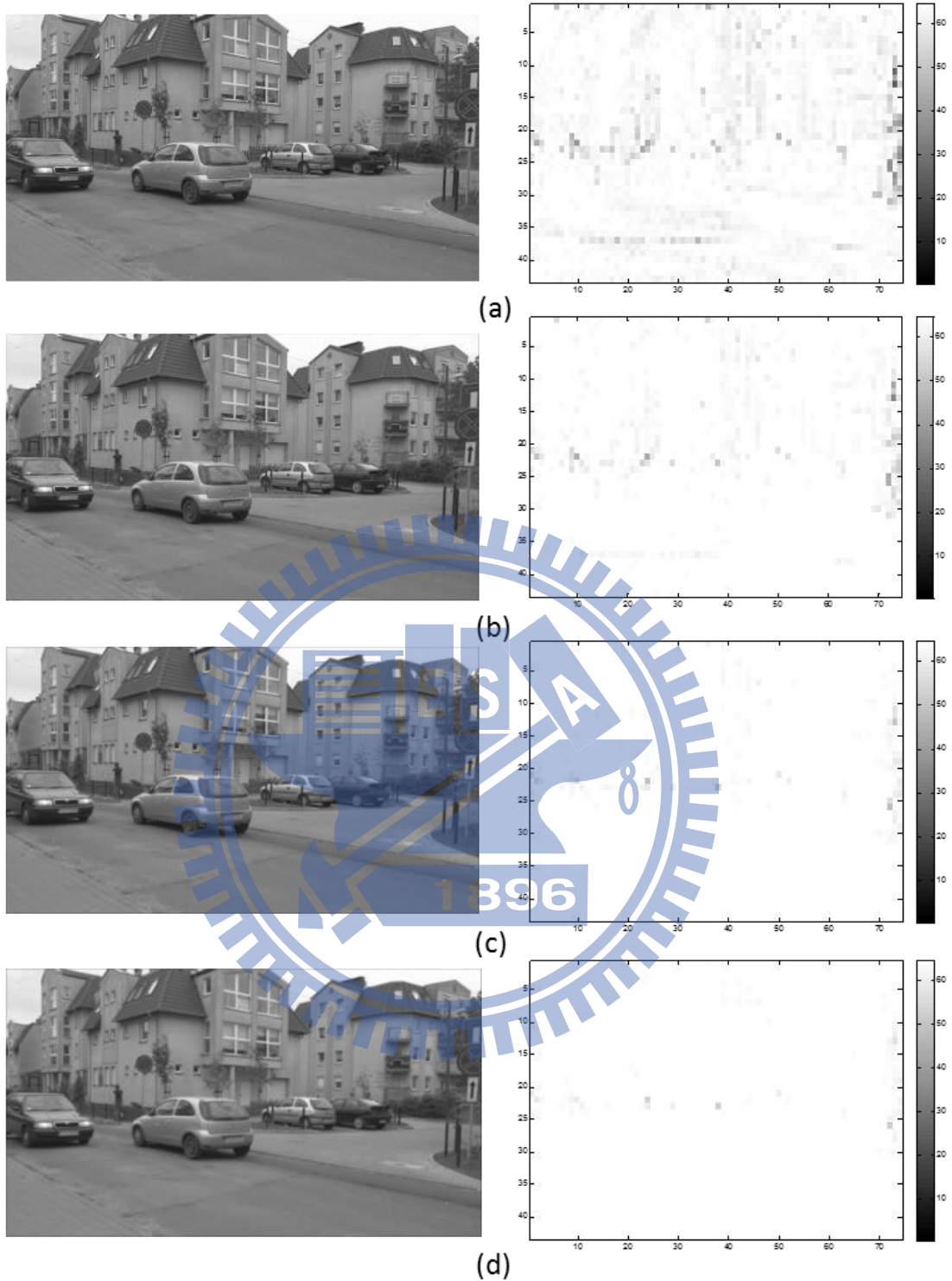


Figure 41. Left column shows the *Quantization* distorted image of sequence *Poznan\_Street* with Gaussian filter using different  $\sigma_{Gaussian}$  values. Right column is the corresponding IQS maps. (a) No Gaussian filter, (b)  $\sigma_{Gaussian} = 1$ , (c)  $\sigma_{Gaussian} = 2$ , and (d)  $\sigma_{Gaussian} = 3$ .

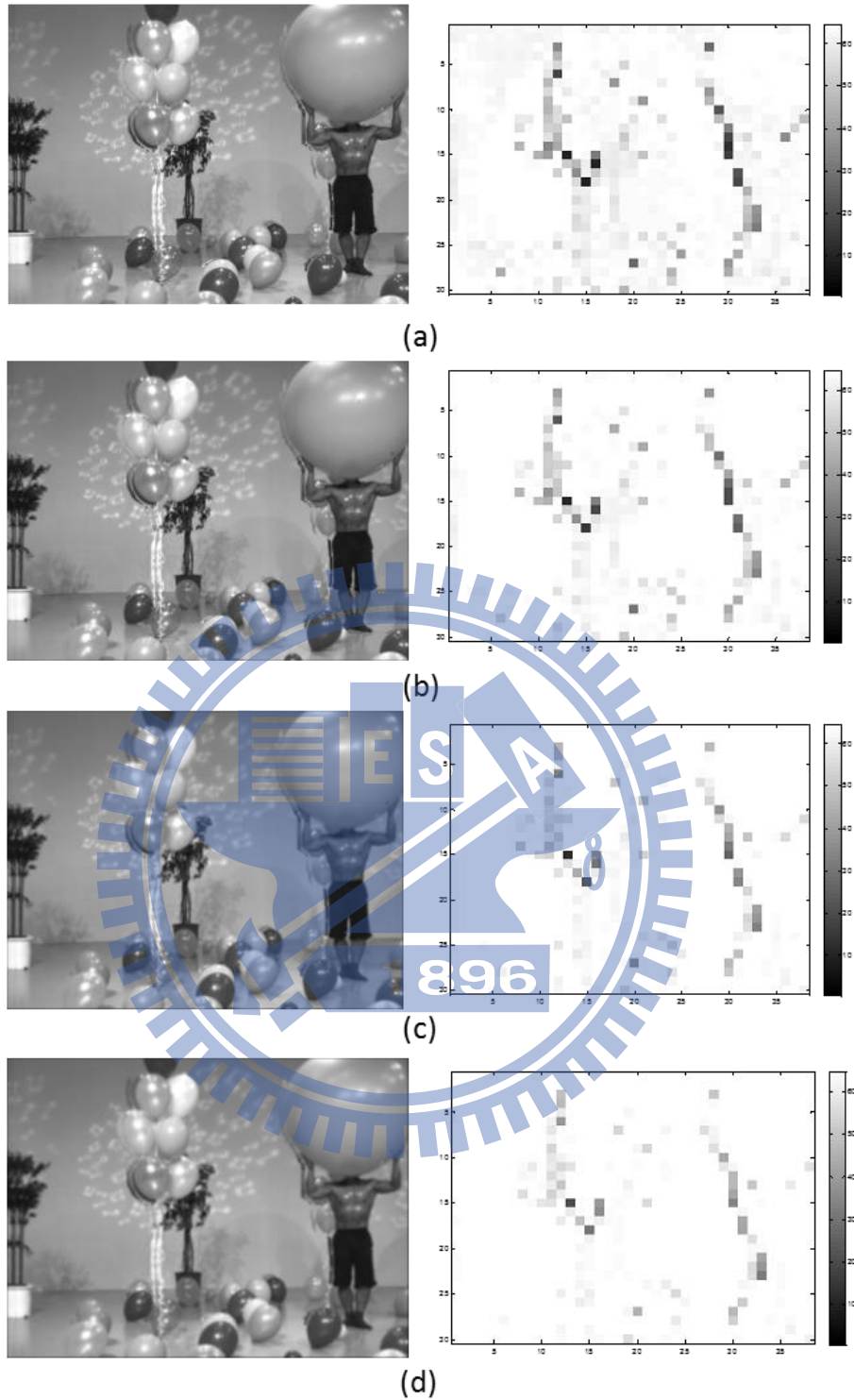


Figure 42. Left column shows the *Offset* distorted image of sequence *Balloons* with Gaussian filter using different  $\sigma_{Gaussian}$  values. Right column is the corresponding IQS maps.

(a) Without Gaussian filter, (b)  $\sigma_{Gaussian} = 1$ , (c)  $\sigma_{Gaussian} = 2$ , and (d)  $\sigma_{Gaussian} = 3$ .

For the *Offset* and the *Quantization* distortion types,  $\sigma_{Gaussian} = 2$  has the best performance when  $\alpha = 0.3$  as shown in Figure 43. And Figure 44 shows a result that when we select a smaller  $\alpha$  value, that is, weight IQS index less, the influence of  $\sigma_{Gaussian}$  is getting smaller too, and the most appropriate  $\sigma_{Gaussian}$  value is that also smaller.

For the *Gaussian\_Noise* distortion, unlike the *Offset* and the *Quantization* types, a larger  $\sigma_{Gaussian}$  value shows better performance than the smaller ones as shown in Figure 46. The PLCC is increase with  $\sigma_{Gaussian}$  comes larger. Because *Gaussian\_Noise* introduces blur like artifact that we mention before. And from the subjective test, we discover that human have high acceptability to this artifact, but SSIM penalize it. So Gaussian filter blurs this artifact, and makes these regions more similar to blurring effect.

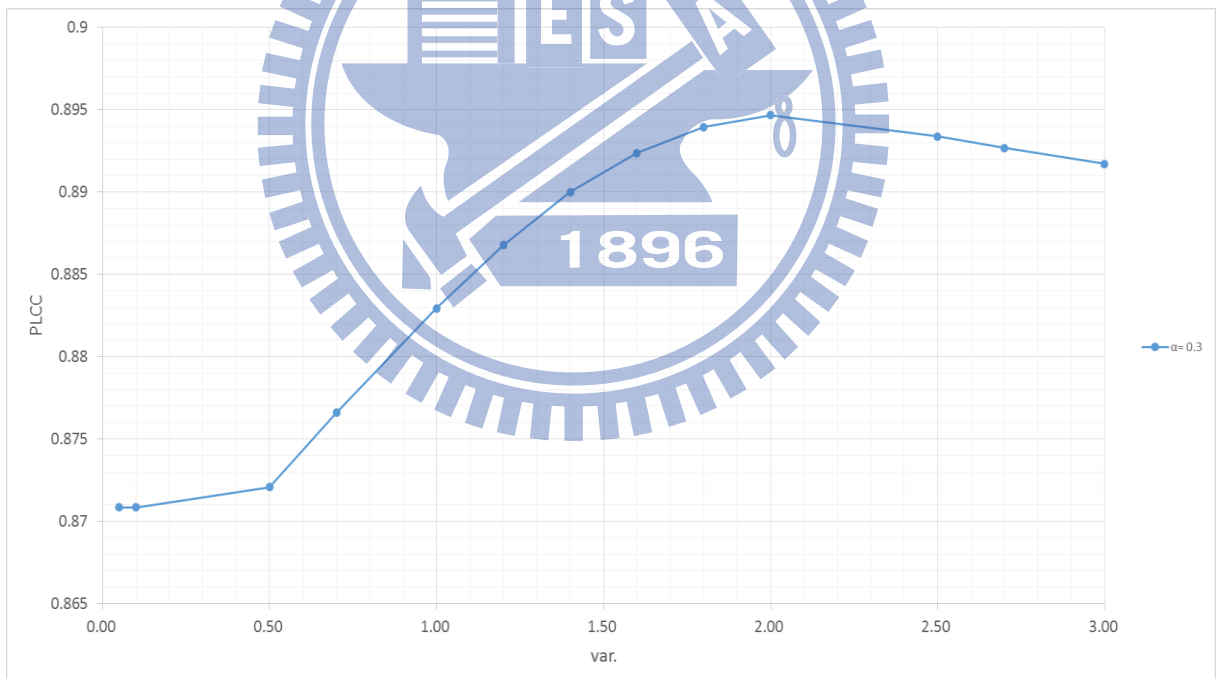


Figure 43. PLCC has maximum value when  $\sigma_{Gaussian} = 2$  for the *Offset* and the *Quantization*. ( $\alpha = 0.3$ , window = 25, pooling p = 5%,  $pK^{th} = 70$ )

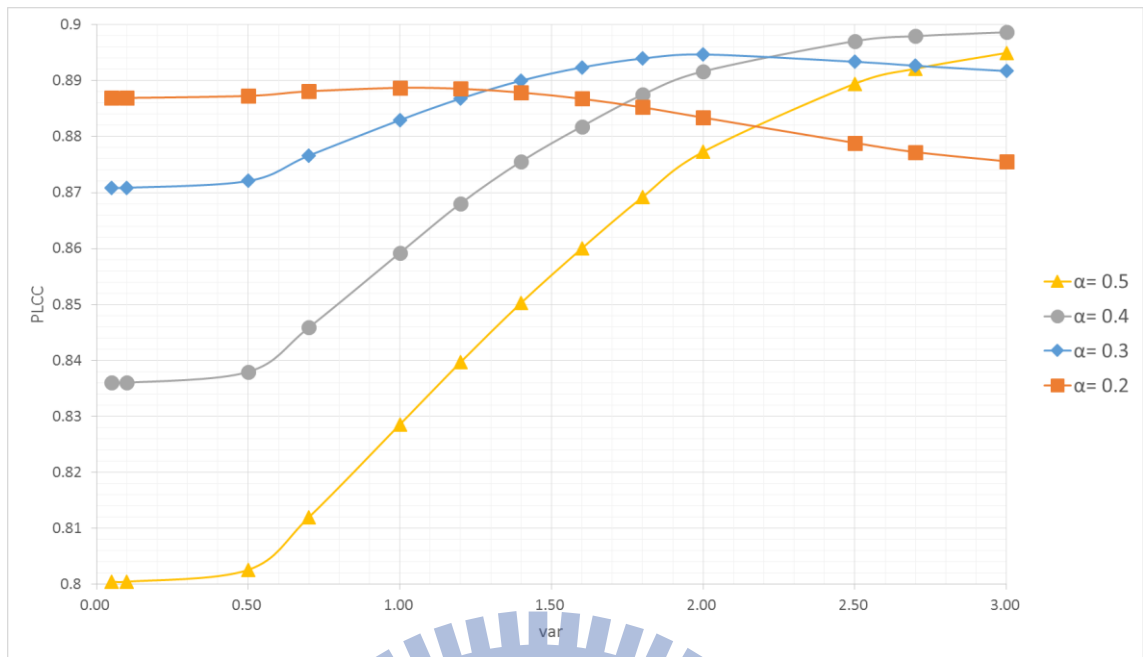


Figure 44.  $\sigma_{Gaussian}$  should be chosen smaller when  $\alpha$  is small.

(window = 25, pooling p = 5%,  $p_{K^{th}} = 70$ )

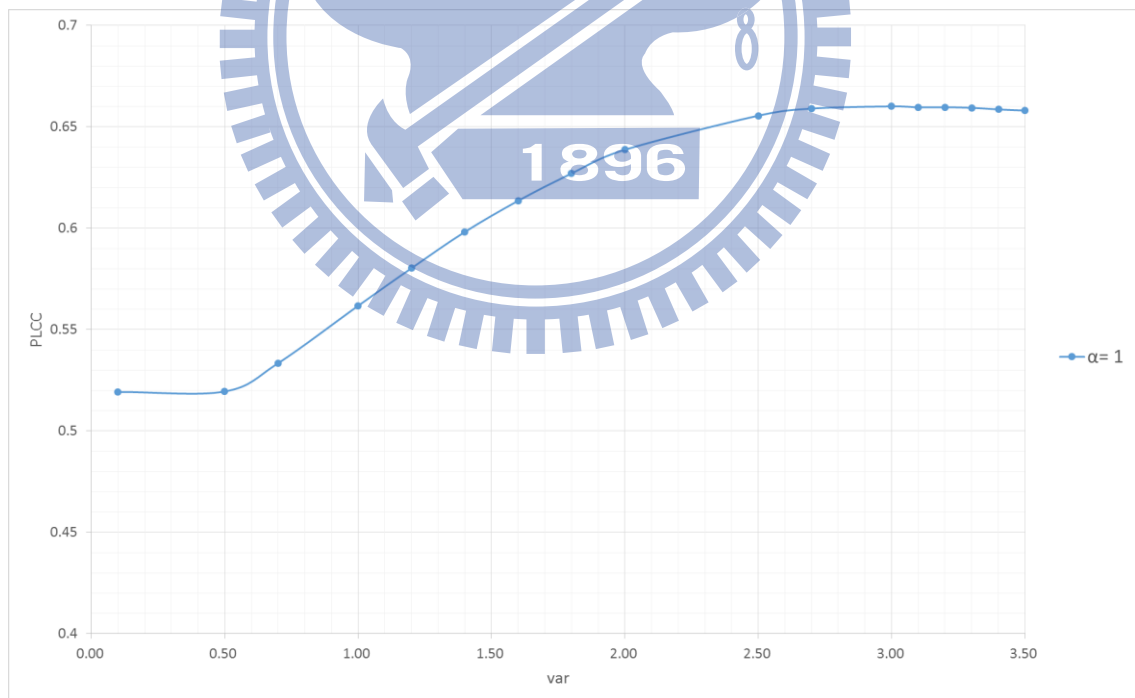


Figure 45. Large  $\sigma_{Gaussian}$  is better for *Gaussian\_Noise* distortion.

(window = 25, pooling p = 100%)

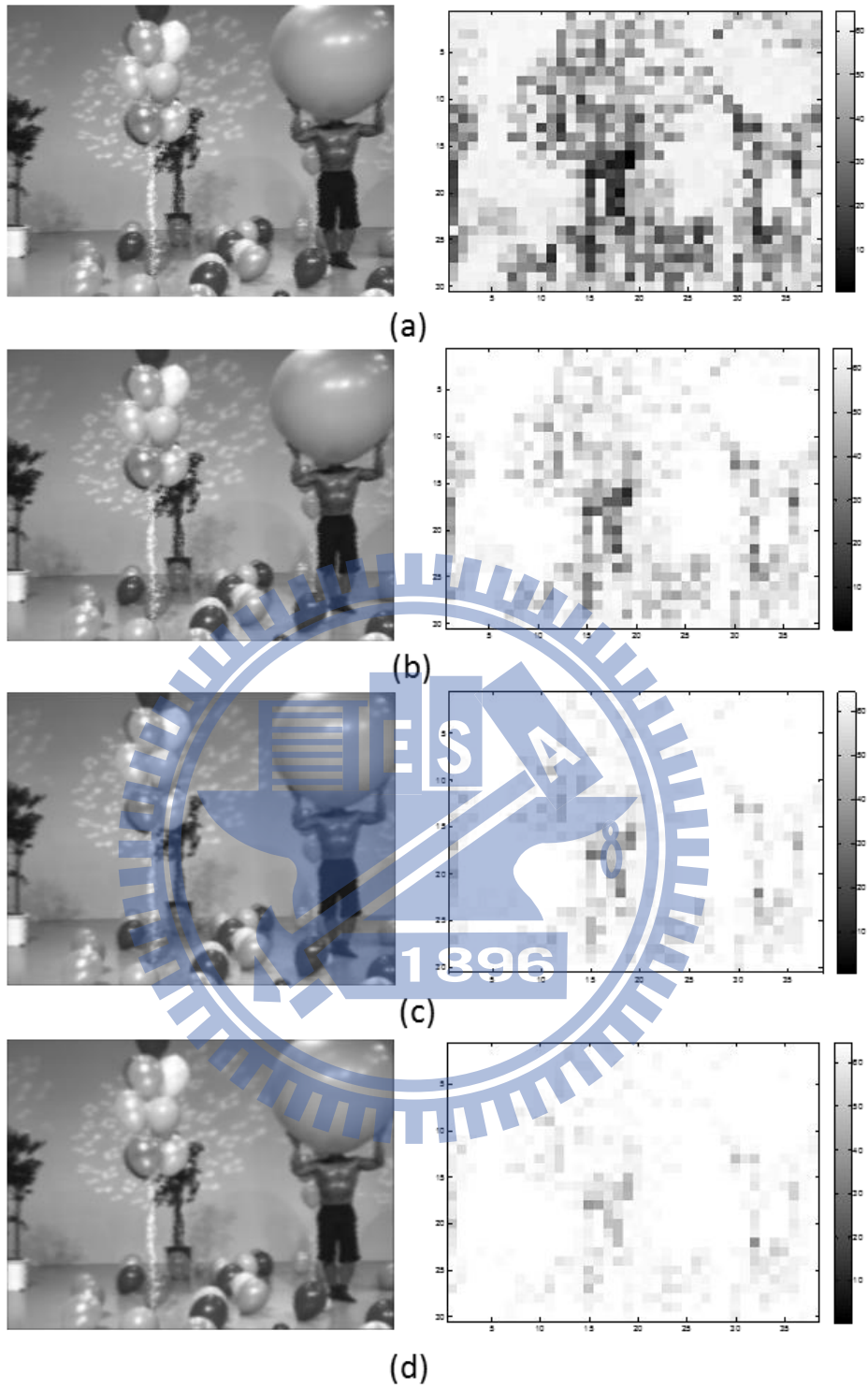


Figure 46. Left column shows the *Gaussian\_Noise* distorted image of sequence *Balloons* with Gaussian filter using different variance parameters. Right column is the corresponding IQS maps. (a) No Gaussian filter, (b)  $\sigma_{Gaussian} = 1$ , (c)  $\sigma_{Gaussian} = 2$ , and (d)  $\sigma_{Gaussian} = 3$ .



### 5.3.5. Window size ( $N$ )

Figure 47 shows that there is no special rule to decide the window size. When  $N=25$ , the performance is relatively higher than the other window sizes for the *Offset* and the *Quantization* types as the orange curve with triangular patterns in Figure 47. That is why we pick  $N=25$  in the previous experiment. And from Figure 48, we can find out that the average execution time per image of  $N=25$  is decreased by about 2 times than that of  $N=15$ . So  $N=25$  not only has relative high performance but it also runs faster.

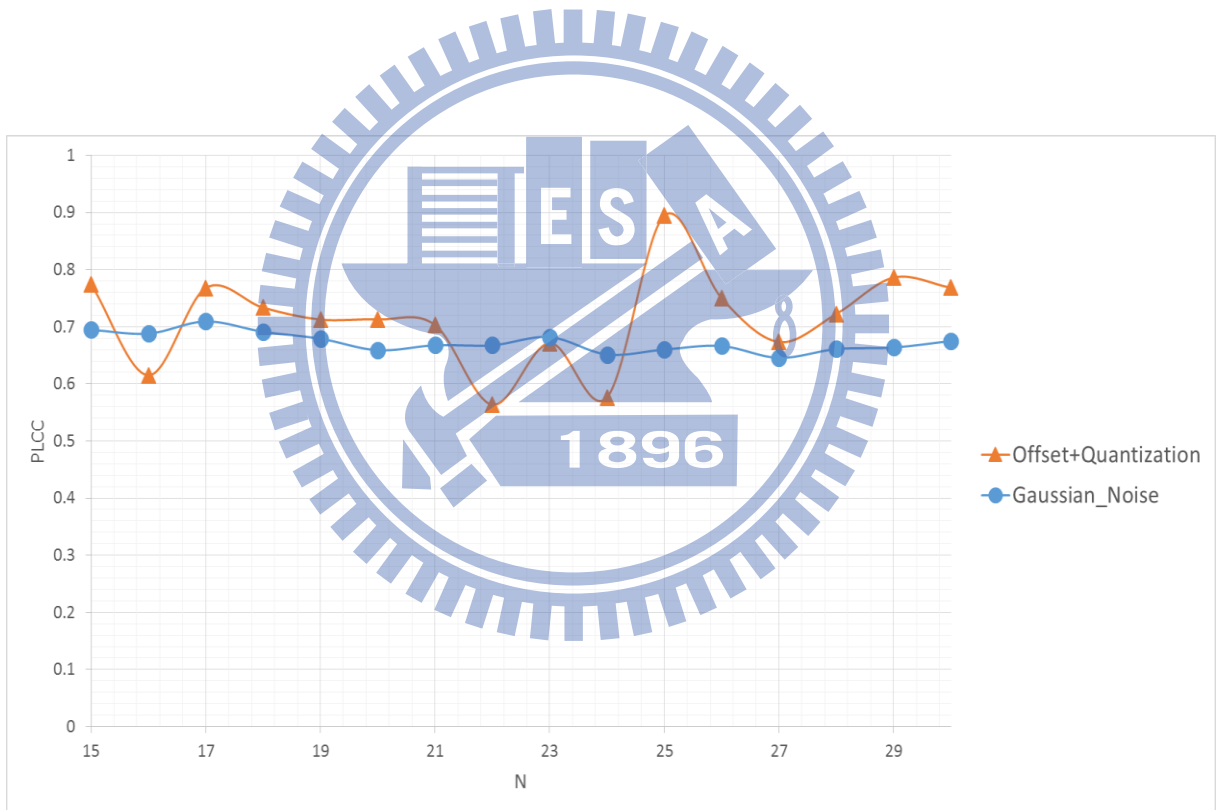


Figure 47. Window size does not affect the performance much.

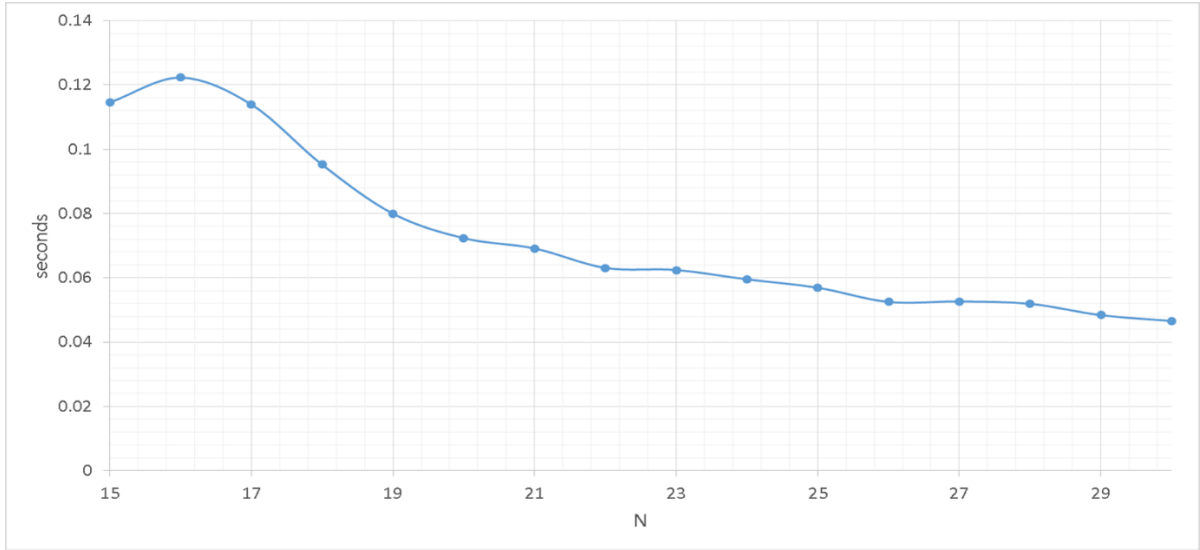


Figure 48. Execution time per image decreases when the window size becomes larger.

## 5.4. Experimental Result and Analysis

Based on the collected data in section 5.3, the parameters were used to test our database are shown in Table 4. We compare the proposed model with the commonly used 2D quality assessment metrics such as PSNR, SSIM [8], MSSIM [9], UQI [10], VIF [11], and VSNR [12]. They are implemented by the MeTriX MuX Visual Quality Assessment Package [30].

Table 4. Parameters set to evaluate the database.

	<i>Offset and Quantize</i>	<i>Gaussian_Noise</i>
Window Size (N)	<b>25</b>	<b>25</b>
$\sigma_{Gaussian}$	<b>2</b>	<b>3</b>
$p_{K^{th}}$	<b>70</b>	-
$\alpha$	<b>0.3</b>	<b>1</b>
Pooling proportion (p)	<b>5%</b>	<b>100%</b>

Table 5 and Table 6 show the performance comparison in terms of Pearson Linear Correlation Coefficient (PLCC), Spearman's Rank Order Correlation Coefficient (SROCC), Root Mean Square Error (RMSE), and Outlier Ratio (RO). The Tables show that our model increases about 30%-55% in PLCC and 30%-60% in SROCC, decreases about 0.3-0.4 in RMSE, and less outlier points for the *Offset* and *Quantization* types. For the *Gaussian\_Noise* distortion, the enhancement is not obvious as the *Offset* and the *Quantization* types, our model increases about 5%-35% in PLCC and 35%-0% in SROCC, decreases about 0.01-0.2 in RMSE, and less outlier points. Especially compare with VIF metric, the performances are almost the same. That's because the artifacts of *Gaussian\_Noise* is similar to blur which is a kind of conventional 2D distortion. Thus it doesn't have clear advantage. Figure 49 and Figure 50 show the scatter plots which the horizontal axes are indexes of the objective quality assessment metrics and the vertical axes are the DMOS of our collected data. The regression formula of Figure 49(f) is

$$y = \frac{2.528}{1 + \exp(50.5829(x - 0.894))}$$

and that of Figure 50(f) is

$$y = \frac{2.5609}{1 + \exp(173.0846(x - 0.9919))}$$

There is a defect for our model, from Figure 49(f) and Figure 50(f), we can see the ranges of final scores computed by proposed model are narrow. The scores is between 0.76 to 0.96 for the *Offset* and *Quantization* and 0.92 to 1 for the *Gaussian\_Noise*. That is, the scores of the best quality and the worst images have no significant difference.

Table 5. Performance comparison of *Offset* and *Quantization*.

(The red and bold numbers are comparisons between each 2D models and proposed model.)

Metrics	PLCC	SROCC	RMSE	OR
PSNR	0.5835	0.5400	0.6997	0.0417
	<b>+0.3131</b>	<b>+0.3617</b>	<b>-0.4202</b>	<b>-0.0417</b>
SSIM	0.3736	0.3930	0.7992	0.0833
	<b>+0.5248</b>	<b>+0.5087</b>	<b>-0.4202</b>	<b>-0.0833</b>
MSSIM	0.5120	0.5104	0.7400	0.0833
	<b>+0.3864</b>	<b>+0.3913</b>	<b>-0.361</b>	<b>-0.0833</b>
UQI	0.3955	0.2617	0.7925	0.1667
	<b>+0.5029</b>	<b>+0.6400</b>	<b>-0.4135</b>	<b>-0.1667</b>
VIF	0.5291	0.4409	0.7314	0.0833
	<b>+0.3693</b>	<b>+0.4608</b>	<b>-0.3524</b>	<b>-0.0833</b>
VSNR	0.3462	0.3478	0.8083	0.1667
	<b>+0.5522</b>	<b>+0.5539</b>	<b>-0.4293</b>	<b>-0.1667</b>
<b>Proposed</b>	<b>0.8984</b>	<b>0.9017</b>	<b>0.3790</b>	<b>0.</b>

Table 6. Performance comparison of *Gaussian\_Noise*.

(The red and bold numbers are comparisons between each 2D models and proposed model.)

Metrics	PLCC	SROCC	RMSE	OR
PSNR	0.7247	0.7455	0.5207	0
	<b>+0.0548</b>	<b>0</b>	<b>-0.0478</b>	<b>0</b>
SSIM	0.5655	0.5182	0.6224	0.0909
	<b>+0.214</b>	<b>+0.2273</b>	<b>-0.1495</b>	<b>-0.0909</b>
MSSIM	0.6798	0.5455	0.5535	0
	<b>+0.0997</b>	<b>+0.2000</b>	<b>-0.0806</b>	<b>0</b>
UQI	0.4631	0.5000	0.6688	0
	<b>+0.3164</b>	<b>+0.2455</b>	<b>-0.1959</b>	<b>0</b>
VIF	0.7691	0.7273	0.4831	0
	<b>+0.0104</b>	<b>+0.0182</b>	<b>-0.0102</b>	<b>0</b>
VSNR	0.4314	0.3909	0.6809	0.0909
	<b>+0.3481</b>	<b>+0.3546</b>	<b>-0.208</b>	<b>-0.0909</b>
<b>Proposed</b>	<b>0.7795</b>	<b>0.7455</b>	<b>0.4729</b>	<b>0</b>

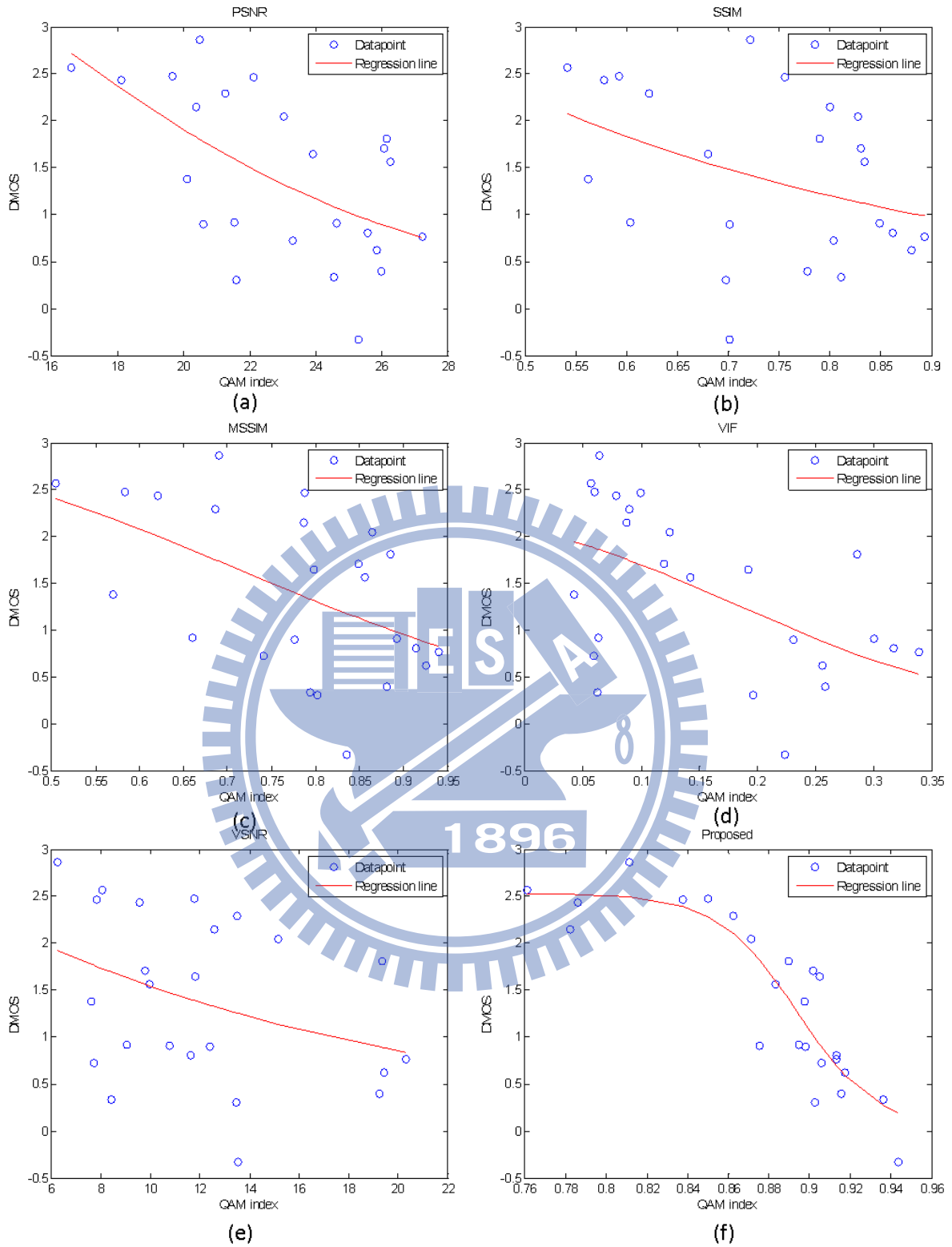


Figure 49. Scatter plot of the objective quality scores against the DMOS. (a) PSNR (b) SSIM (c) MSSIM (d) VIF (e) VSNR (f) Proposed. (*Offset and Quantization*)

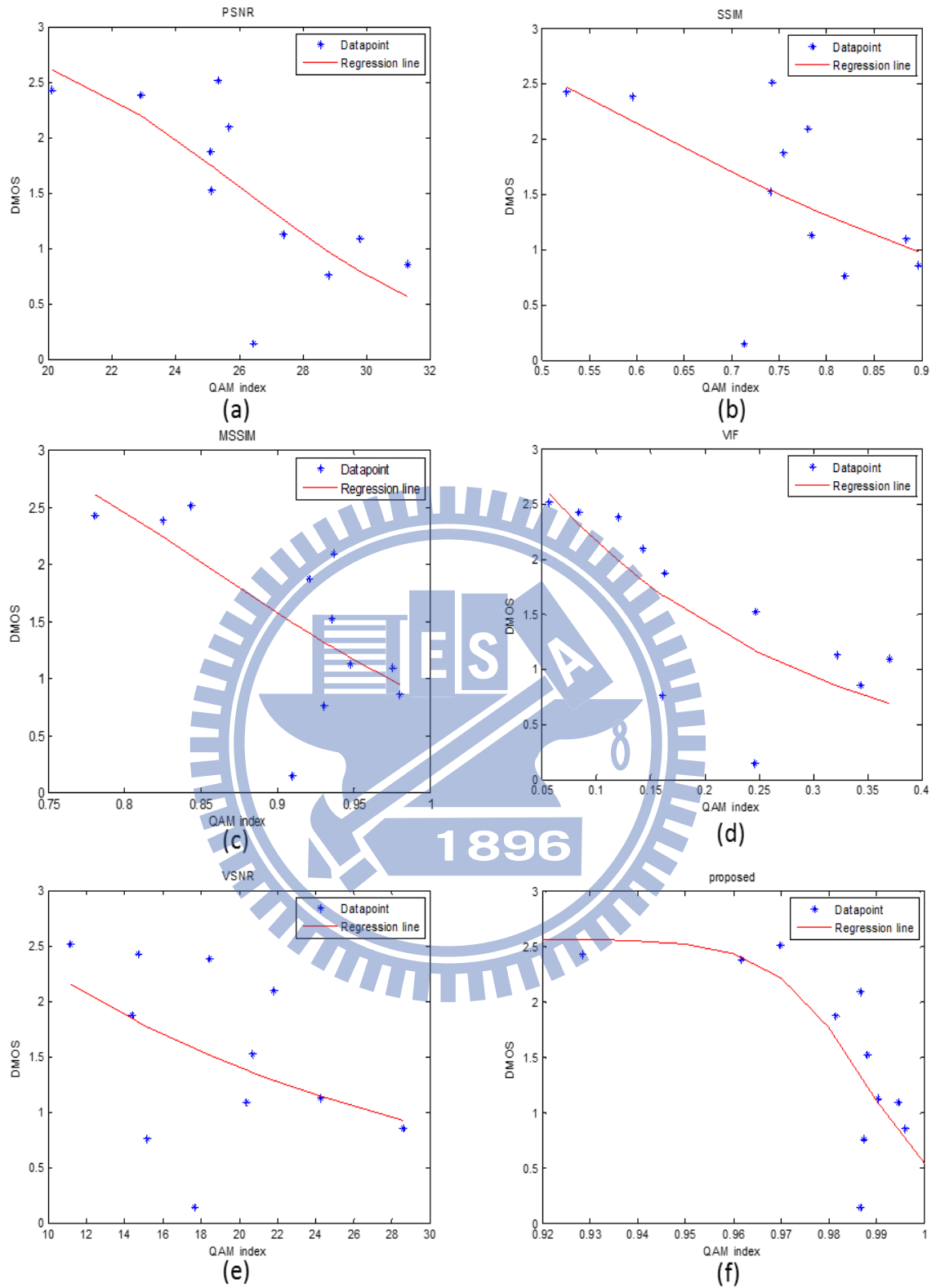


Figure 50. Scatter plot of the objective quality scores against the DMOS. (a) PSNR (b) SSIM

(c) MSSIM (d) VIF (e) VSNR (f) Proposed. (*Gaussian\_Noise*)

For overall database, the parameters were used to test are shown in Table 7. It is similar to the parameter values of test *Offset* and *Quantization*. The only difference is the parameter  $\alpha$ . From Figure 35, the best performance is near  $\alpha = 0.5$ . Table 8 shows the performance comparison, and it shows that our model increases about 21%-52% in PLCC and 25%-50% in SROCC, decreases about 0.16-0.27 in RMSE, and less or equal to other 2D metrics in OR. Figure 51 shows the scatter plots which the blue circular points are *Offset* and *Quantization* distortions and the green star marks are *Gaussian\_Noise* data. The regression formula of Figure 51(f) is

$$y = \frac{2.5849}{1 + \exp(24.7015(x - 0.8948))}$$

Table 7. Parameters set to evaluate the overall database.

	Overall	
Window Size (N)	1896	25
$\sigma_{Gaussian}$		2
$p_{K^{th}}$		70
$\alpha$		0.5
Pooling proportion (p)		5%



Table 8. Performance comparison of overall database.

(The red and bold numbers are comparisons between each 2D models and proposed model.)

Metrics	PLCC	SROCC	RMSE	OR
PSNR	0.5233	0.5114	0.7084	0.0857
	<b>+0.2509</b>	<b>+0.2597</b>	<b>-0.1819</b>	<b>-0.0571</b>
SSIM	0.4235	0.4091	0.7531	0.0571
	<b>+0.3507</b>	<b>+0.3620</b>	<b>-0.2266</b>	<b>-0.0285</b>
MSSIM	0.4213	0.4090	0.7539	0.0571
	<b>+0.3529</b>	<b>+0.3621</b>	<b>-0.2274</b>	<b>-0.0285</b>
UQI	0.3648	0.3088	0.7744	0.1143
	<b>+0.4094</b>	<b>+0.4623</b>	<b>-0.2479</b>	<b>-0.0857</b>
VIF	0.5575	0.5173	0.6903	0.0286
	<b>+0.2167</b>	<b>+0.2538</b>	<b>-0.1638</b>	<b>0</b>
VSNR	0.2535	0.2625	0.8042	0.2000
	<b>+0.5207</b>	<b>+0.5086</b>	<b>-0.2777</b>	<b>-0.1714</b>
<b>Proposed</b>	<b>0.7742</b>	<b>0.7711</b>	<b>0.5265</b>	<b>0.0286</b>

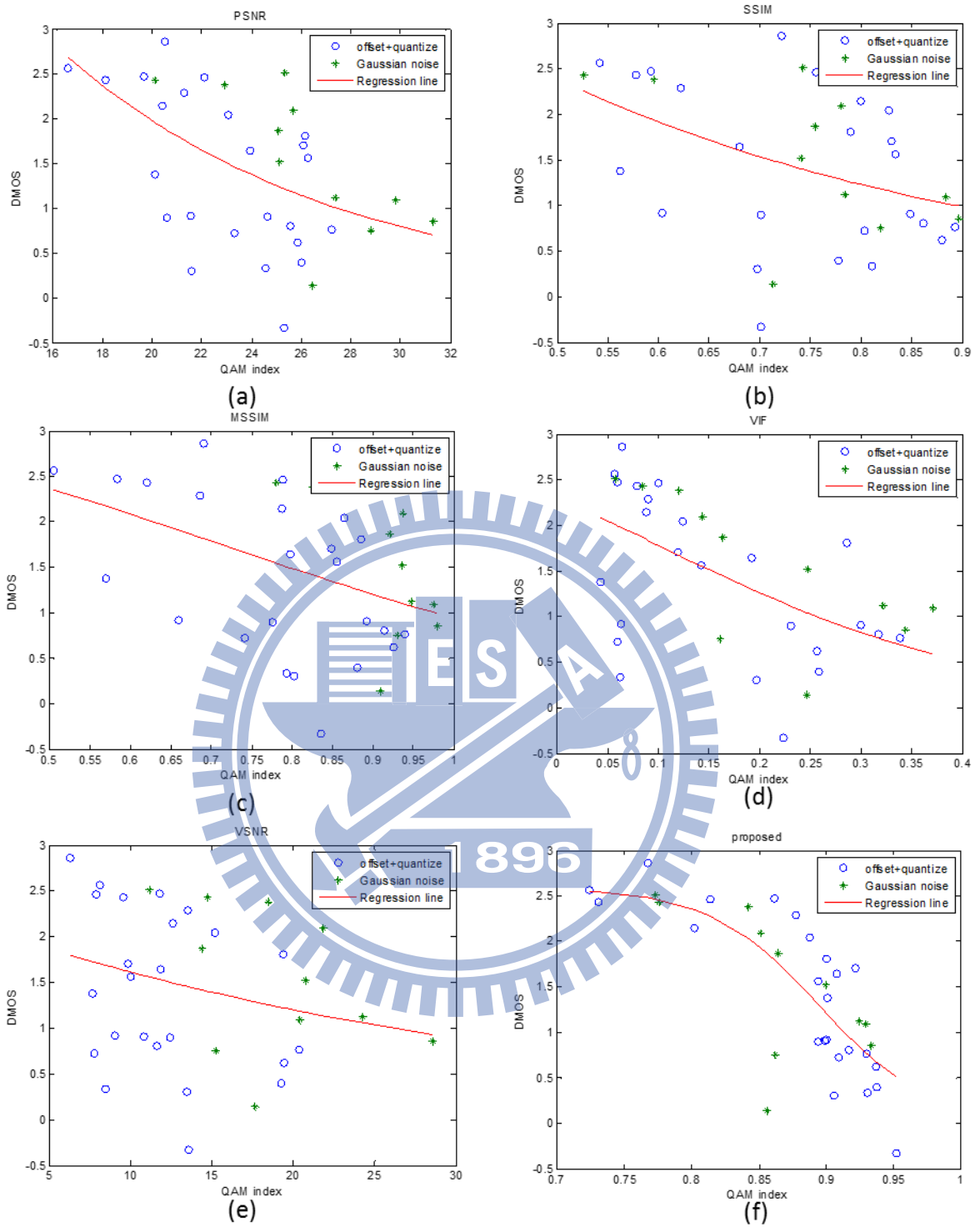


Figure 51. Scatter plot of the objective quality scores against the DMOS. (a) PSNR (b) SSIM (c) MSSIM (d) VIF (e) VSNR (f) Proposed. (Overall database)

We also note that adding a Gaussian filter before calculating SSIM, as shown as Figure 52, has no apparent benefit in evaluating the 2D image quality, but it improves in our 3D image testing. We applied the model shown in Figure 52 with a zero mean and  $\sigma_{Gaussian} = 2$  Gaussian filter to LIVE Image Quality Assessment Database available from the Laboratory for Image & Video Engineering (LIVE) from the University of Texas at Austin [32]-[34]. The results in Table 9 show that PLCC index and RMSE decrease about 0.5% and 0.2 respectively though the model in Figure 52. But it increases about 2.8% in SROCC. And the scatter plots are shown in Figure 53. The figure shows that the model in Figure 52 lets the objective scores converge on the high-rank region because high frequency distortion such as block effect and noise would be filtered by the Gaussian filter.

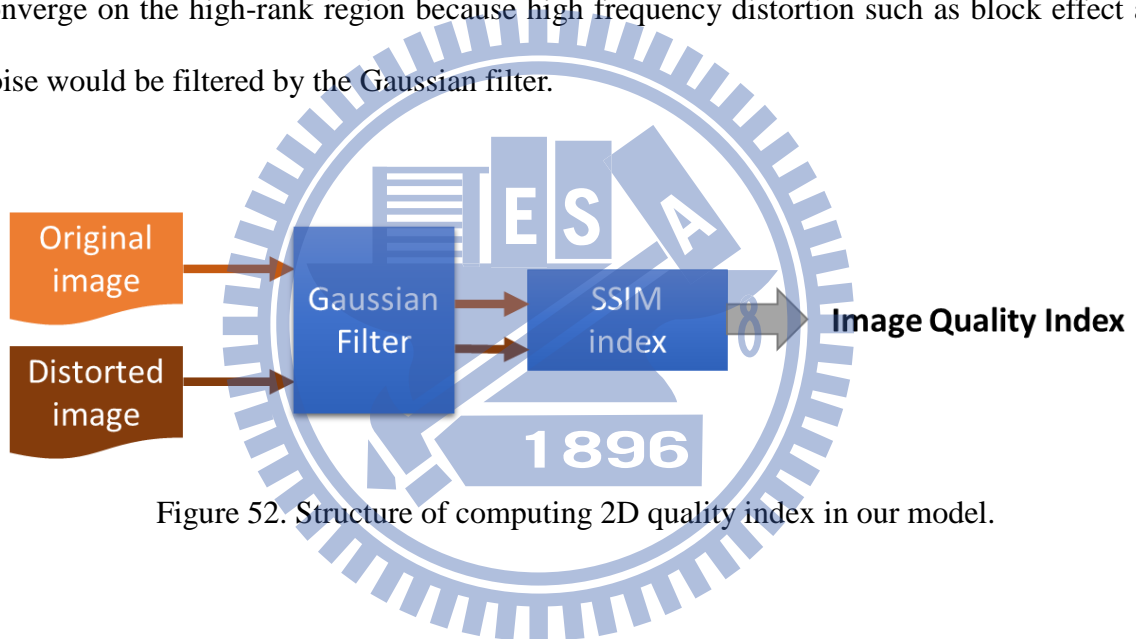
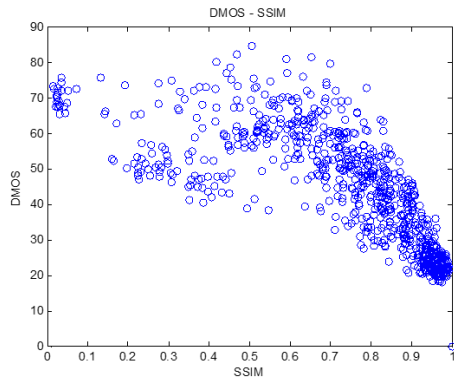


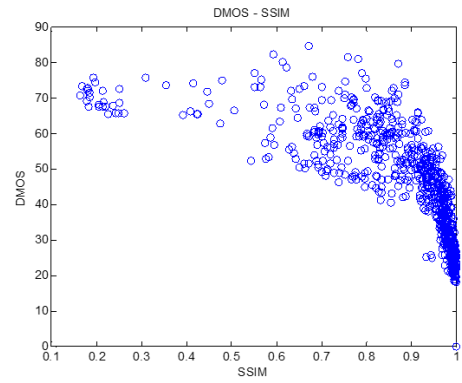
Figure 52. Structure of computing 2D quality index in our model.

Table 9. Performance comparison.

	<i>PLCC (%)</i>	<i>SROCC (%)</i>	<i>RMSE</i>
SSIM	<b>91.71</b>	92.50	<b>9.2845</b>
Gaussian filter + SSIM	91.25	<b>95.37</b>	9.5033

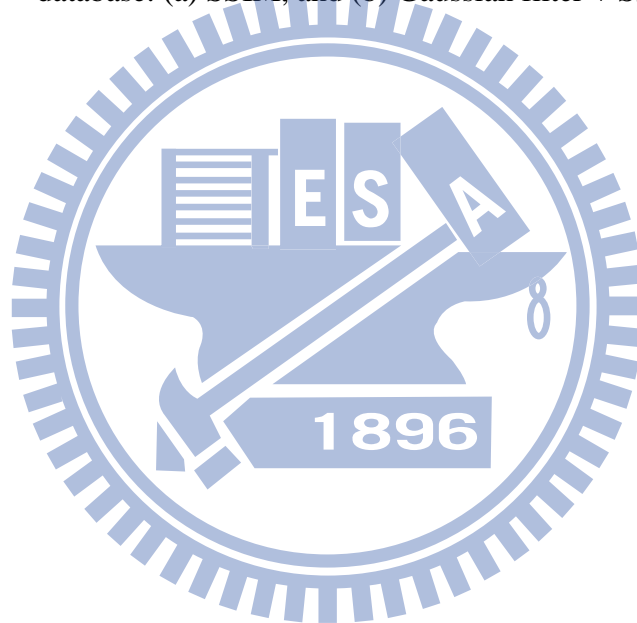


(a)



(b)

Figure 53. Scatter plots of the objective quality scores against DMOS applied to the LIVE database. (a) SSIM, and (b) Gaussian filter + SSIM



# Chapter 6 3D Quality Assessment for Color Distortion

In section 3.3, some stereoscopic color distortion databases were introduced. The databases [16]-[18] include the popular types of color distortion, such as Gaussian blur, JPEG coding, JPEG2000 coding, Gaussian noise etc., symmetrically on the stereo-pair images. Differ to these databases, we produce a database containing color distortion for view synthesis system. We consider Gaussian blur, JPEG coding, and Gaussian noise, in the 3DVC system based on multiple-view plus depth (MVD) format.

## 6.1. Synthesized View Database with Color Distortion

There are two sets of view-synthesized images with color distortion in our database, Distortion-Synthesis (D-S) images and Synthesis-Distortion (S-D) images. Their most significant difference is how the distorted color image is used to produce the synthesized images. We describe these two image synthesis processes below.

### (1) Distortion-Synthesis (D-S) images

Figure 54 shows the structure of D-S synthesis process. Both the left and right color images are distorted by the color distortion process first. The color distortion process distorts both images by the same type of distortion and parameter independently. Then, the distorted virtual view image is synthesized based on the distorted color images and corresponding non-distorted depth maps using the 1D-fast-VSRS in reference software HTM version 3.1 [29].

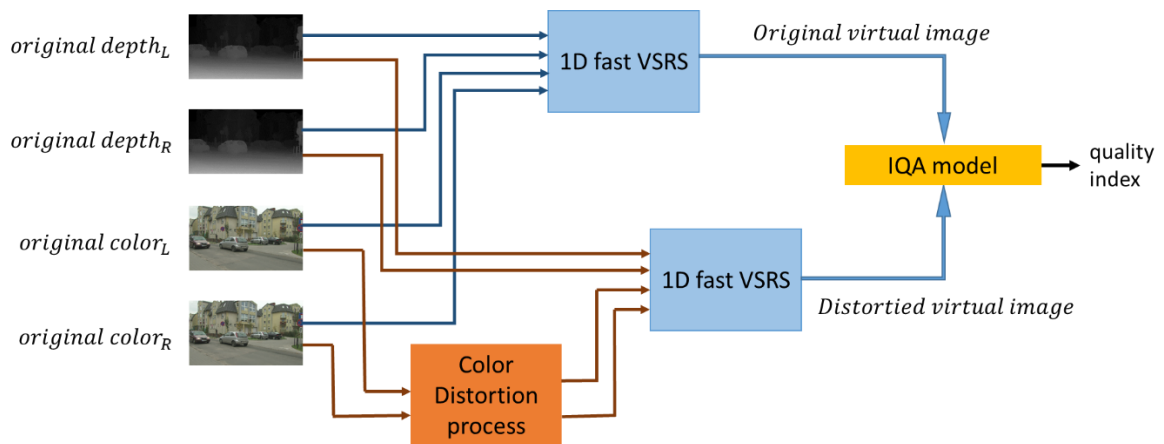


Figure 54. The D-S image producing process and its quality evaluation.

## (2) Synthesis-Distortion (S-D) images

The D-S image synthesis procedure is shown in Figure 55. The original left and right color images and depth maps are used to produce the virtual view image via the 1D-fast-VSRS. Then, the color distortion process distorts the synthesized virtual view images. The S-D image generation is very similar to that of the conventional 2D image quality assessment set-up. Both systems evaluate the image quality index by an original image and its distorted image. The only difference is the original images in the S-D images are the synthesized virtual view images not the original images.

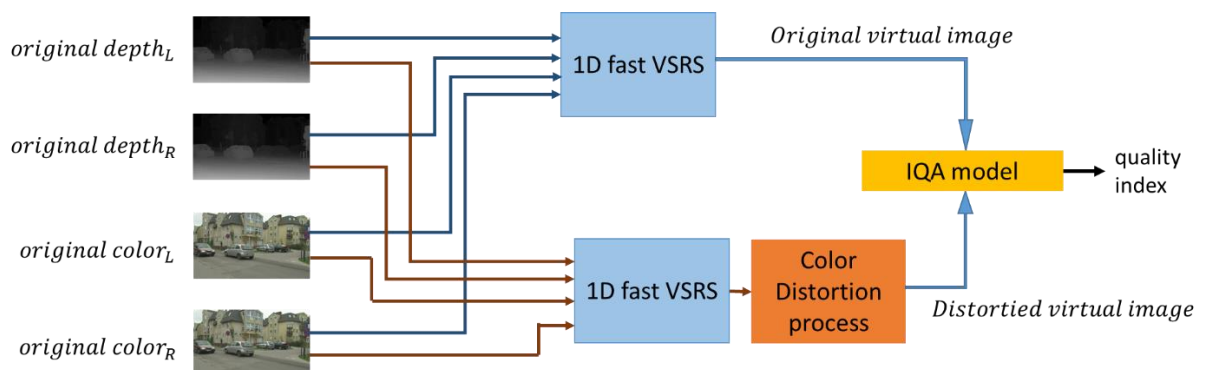


Figure 55. The S-D image producing process and its quality evaluation.

Three commonly used types of distortion we use in the color distortion system are Gaussian blur, JPEG compression coding, and Gaussian noise. Table 10 shows the parameters of these distortions in our database. Two parameter values are selected for Gaussian blur and JPEG coding distortion respectively; and three selected parameter values for Gaussian noise distortion. The difference on the numbers of selected parameter is based on the subjective experience. The Gaussian noise produces more apparent difference between the D-S images and the S-D images. Thus, there are in total 7 distorted images for each test sequence. Note that the Gaussian noise distortions for left and right views in the D-S images of our database are not identical, but with the equal variance.

Table 10. Parameters used in our database.

Type of distortion	Parameter
Gaussian blur	Zero mean and $\sigma = 2$ and 4
JPEG	Quality = 15 and 20
Gaussian noise	Zero mean and $\sigma = 0.005, 0.01$ and 0.05

## 6.2. Subjective Test

We are interested in the difference between 3D stereoscopic viewing and 2D viewing with color distortion. Thus, there are two test scenarios in our subjective experiment, 2D scenario and 3D scenario. As illustrated in Figure 56, for D-S images, the 3D scenario shows the left input original view with distortion to the left eye and the synthesized virtual view synthesized by distorted left and right views to the right eye; and the 2D scenario shows the synthesized virtual view to both eyes. For S-D images, the 3D scenario shows the left input original view with distortion to the left eye and the synthesized virtual view which synthesized by the original left

and right views with distortion to the right eye; and the 2D scenario shows the distorted synthesized virtual view to both eyes.

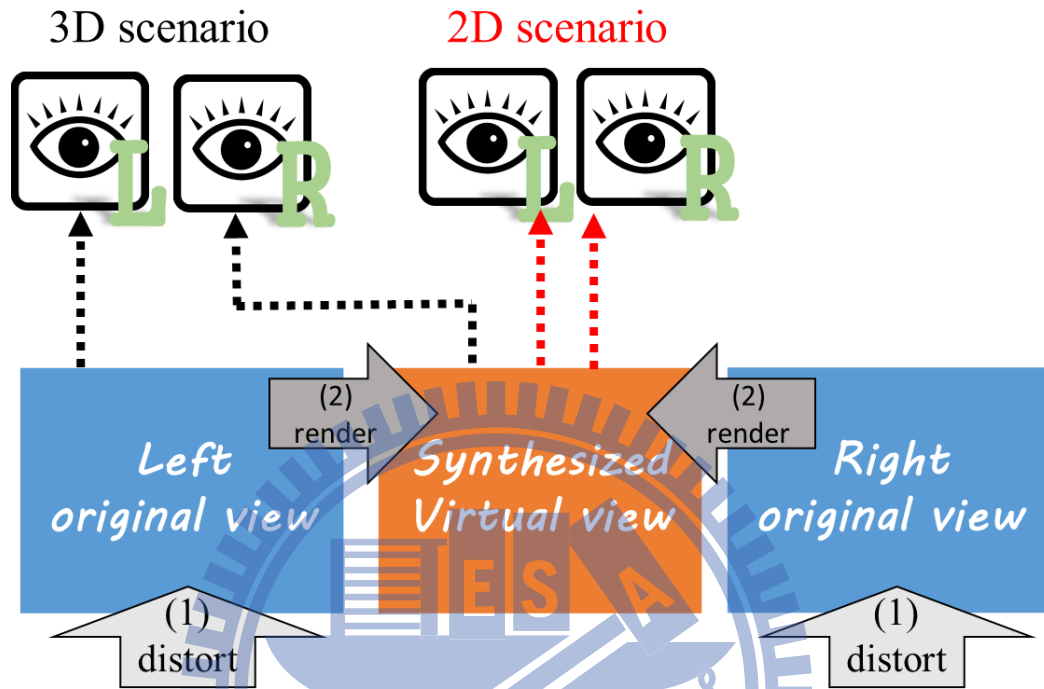


Figure 56. 3D scenario and 2D scenario of the D-S images.

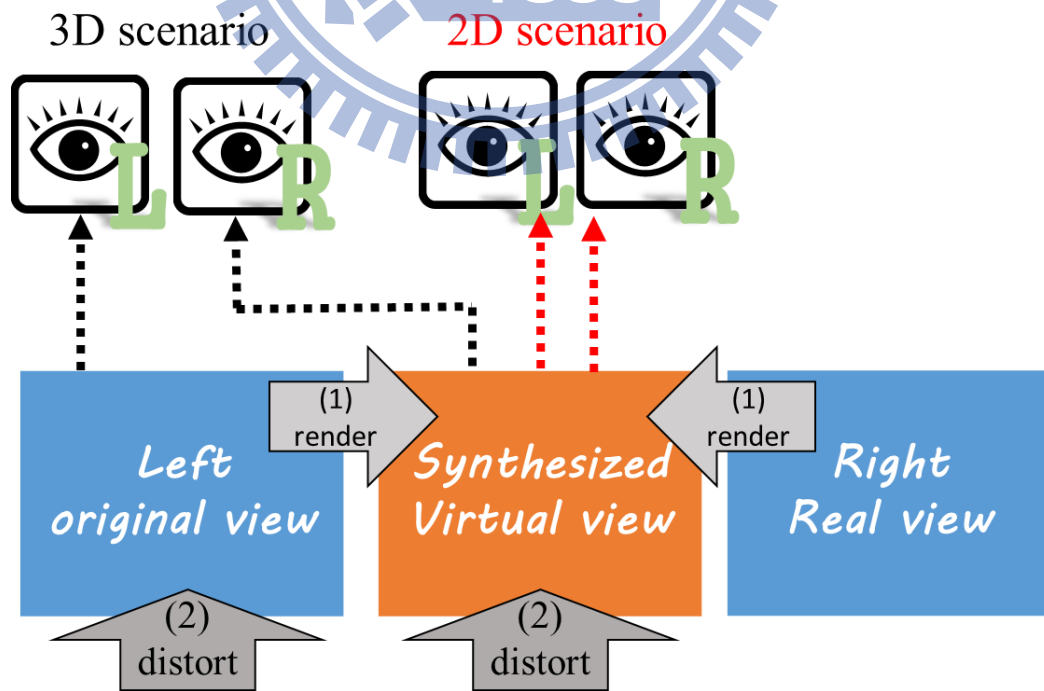


Figure 57. 3D scenario and 2D scenario of the S-D images.



The test sequences used in our database are *Poznan\_CarPark*, *Poznan\_Hall1*, *Poznan\_Hall2* and *Poznan\_Street* provided by Poznan University [25], *Kendo* and *Balloons* provided by Nagoya University [26], *Lovebird1* provided by ETRI [27], and *Newspaper* by GIST [28]. *Poznan\_CarPark*, *Poznan\_Hall1*, *Poznan\_Hall2* and *Poznan\_Street* are used in 2D test scenario due to their equivalent resolution (1920x1088), and *Kendo*, *Balloons*, *Lovebird1*, and *Newspaper* are used in 3D test scenario with resolution 1024x768.

There are 7 distorted images for a sequence with three types of color distortion and 28 (7\*4) distorted images for each test scenario for either D-S or S-D images. Thus, there are 56 (28\*2) distorted images for each test scenario. Totally, 60 (56 distorted images + 4 reference images) pictures in either 3D or 2D test scenarios.

For the subjective test, 20 observers (16 males and 4 females) participated in the experiment. The display we used is 23" LG D237IPS 3D display. And Single Stimulus (SS) method with five discrete scores was used according to the document ITU-R BT.500 [6].

In the 3D subjective test, the "main eye" of observers was record. Because all the stereo images in the database have the synthesized virtual view images for right eye and left original view images for left image, different main eye may lead to different quality perception. For example, the subjective quality perception of an observer, whose main eye is right eye will be dominated by quality of the synthesized virtual images. In our subjective test, there are 10 observers with the right main eye, and 10 with the left.

### **6.3. Experimental Results and Analysis**

According to the types of distorted images and viewing scenarios, our database can be divided into 4 classes as the blue blocks in Figure 58. There are two aspects can be discussed: (1) the difference between the D-S and the S-D distorted images (orange rectangle in Figure 58) and

(2) the difference between 3D and 2D viewing scenarios (green rectangle in Figure 58).

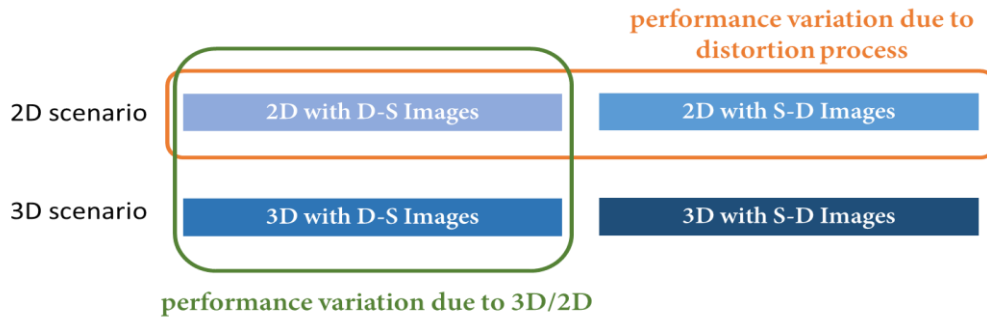


Figure 58. A summary of the database.

It is an interesting issue to see the difference in applying the distortion to images before rendering process or after it. Before doing subjective test, we wonder the effects of SSIM indexes on different types of color distortion, the D-S and S-D images may produce different result on SSIM. Figure 59 shows the SSIM score variation between the D-S and S-D images in the 2D viewing scenario. Each blue point indicates a type of distortion (Gaussian blur/JPEG coding/Gaussian noise) with a selected parameter value. The horizontal axis is the SSIM index applied to the D-S images, and the vertical axis is the SSIM index applied to the S-D images. Every samples locate on the red lines have the same SSIM indexes from the D-S and S-D images, and we called the red line *equalization line*.

In Figure 59(a), we find out that blurring color images before the synthesis step or after it does not have obvious difference using SSIM. But for the Gaussian noise and JPEG coding, Figure 59(b) and Figure 59(c), the D-S images lead to higher SSIM index than the S-D images. After all, SSIM is a 2D QA model, it may not be able to predict the subjective quality of our 3D color distortion database.

As mentioned before, the Gaussian noise distortions for left and right views in the D-S images of our database are not identical, but with the equal variance, called *asymmetric Gaussian Noise (AGN)*. Then, what would happen for the *symmetric Gaussian Noise (SGN)*,

which adds the same Gaussian noise pattern to the both views? Figure 60 shows the points of *AGN* and *SGN* both locate on the lower right side of chart. This means that SSIM indexes of the D-S images are higher than those of the S-D's. In addition, *SGN* produces better quality of the D-S images in SSIM than *ASN*.

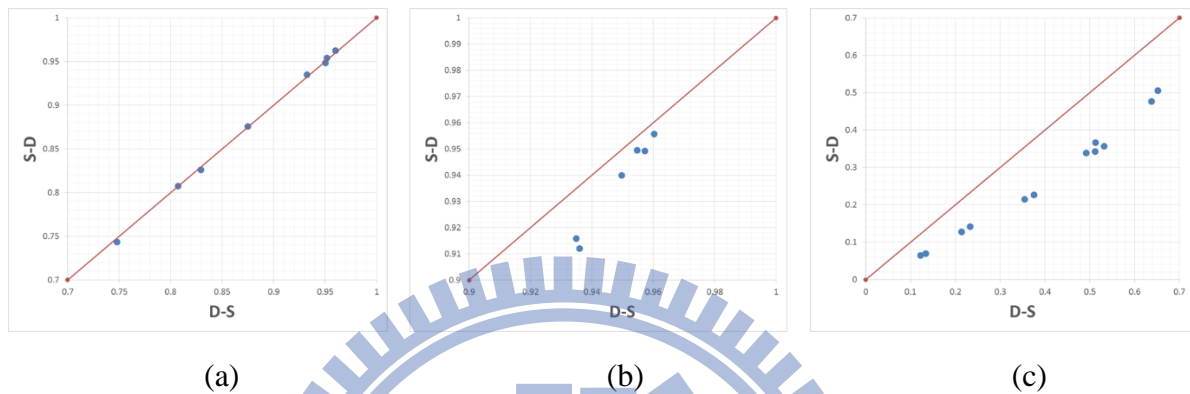


Figure 59. SSIM comparison between D-S and S-D images. (2D scenario)  
 (a) Gaussian blur (b) JPEG coding (c) Gaussian noise.

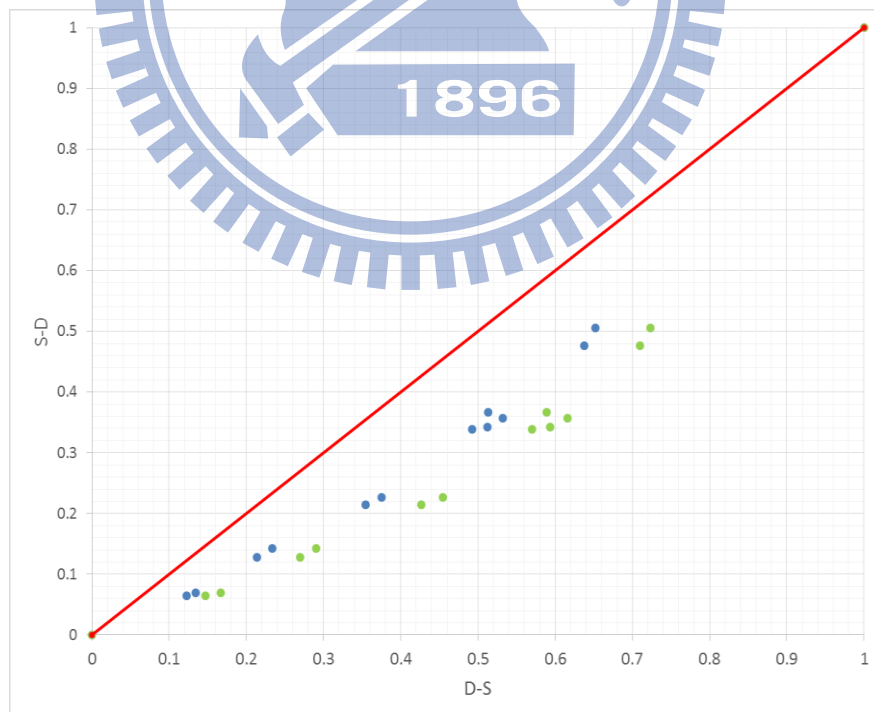


Figure 60. The comparison of *AGN* and *SGN*.

BLUE: *AGN*; GREEN: *SGN*

In the subjective test, we find out that for the JPEG coding and Gaussian noise distortions, as shown in Figure 61(b) and (c), the results are similar to SSIM does. The axes of Figure 61 are the mean opinion score (MOS) between 0 and 5. The D-S images have higher subjective quality than the S-D images. This phenomenon is caused by the late step of synthesis process, *blending*. Blending step merges two rendered color images from left and right views into one image. Thus, the distortions of the synthesized images would be “averaged.” On the other hand, for the Gaussian blur as shown in Figure 61(a), some points marked in red locate on the left side of the *equalization line*, that is, the D-S images produce lower subjective quality than the S-D images. One of these images are shown in Figure 62. There exist unnatural contour around the boundary of different depth plane in Figure 62(a), and we call this phenomenon “*sticker artifact*.” The cause of *sticker artifact* is in the synthesis process. A blurred color image synthesized with a perfect depth map with sharp boundaries produce significant distortions along object boundaries. Figure 63 illustrates this type of distortion due to the view synthesis procedure.

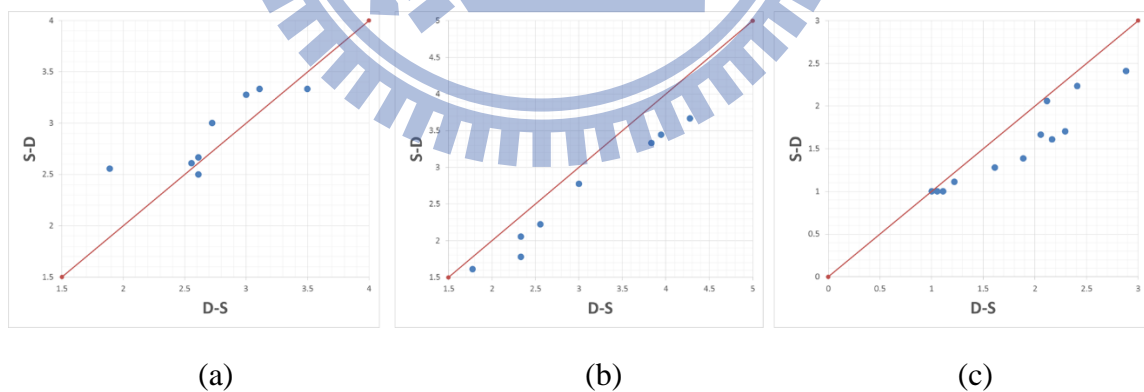


Figure 61. Subjective experimental results of 2D scenario.

(a) Gaussian blur (b) JPEG coding (c) Gaussian noise.

The *sticker artifact* can be perceived in both 2D and 3D viewing scenarios. Figure 64 and Figure 65 show the comparison of D-S and S-D images. It is similar to Figure 59 and Figure 61

in that most points locate on the right side of the *equalization line*. And some images with *sticker artifact* are perceivable by observers.

Conclude the experimental results above, the SSIM scores of the D-S images with Gaussian noise is higher than those of the S-D images, that is, the view synthesis process can cover Gaussian noise distortion. The D-S images with Gaussian blur would produce the *sticker artifact* around the different depth object boundaries. And the *sticker artifact* can be perceived even if the image of the other eye do not contain this artifact in 3D viewing.



Figure 62. Gaussian blur distortion. (a) D-S image. (b) S-D image.



Figure 63. An illustration of sticker artifact.

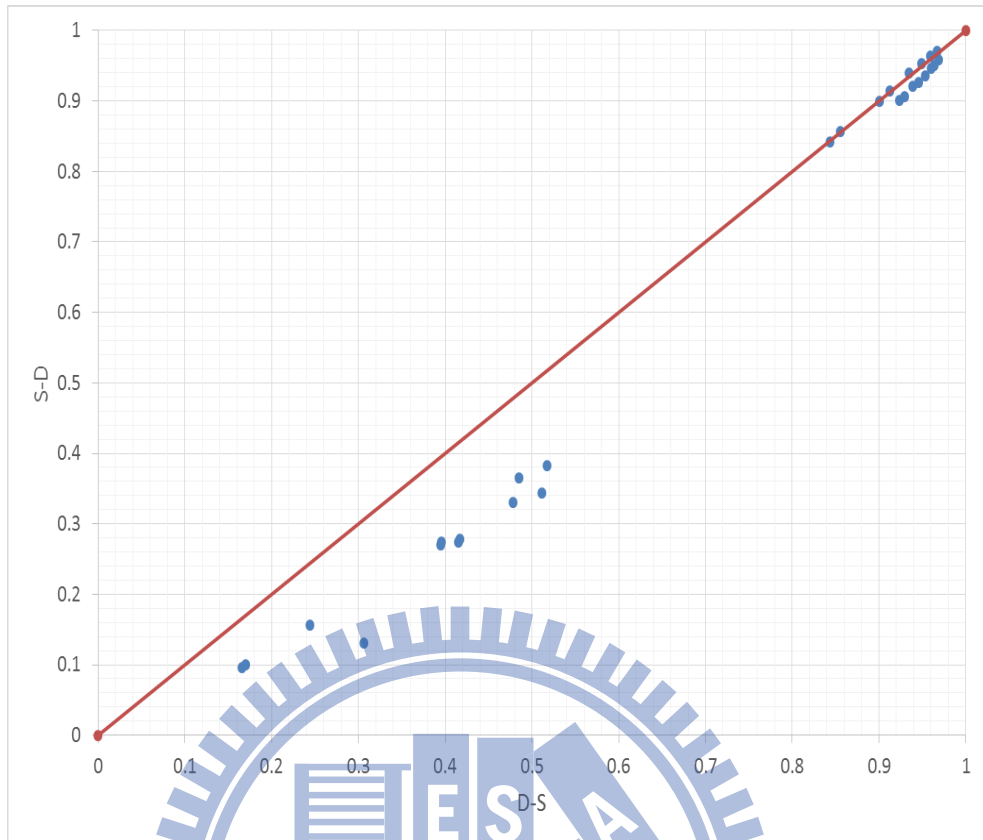


Figure 64. SSIM comparison between D-S and S-D images (3D scenario).

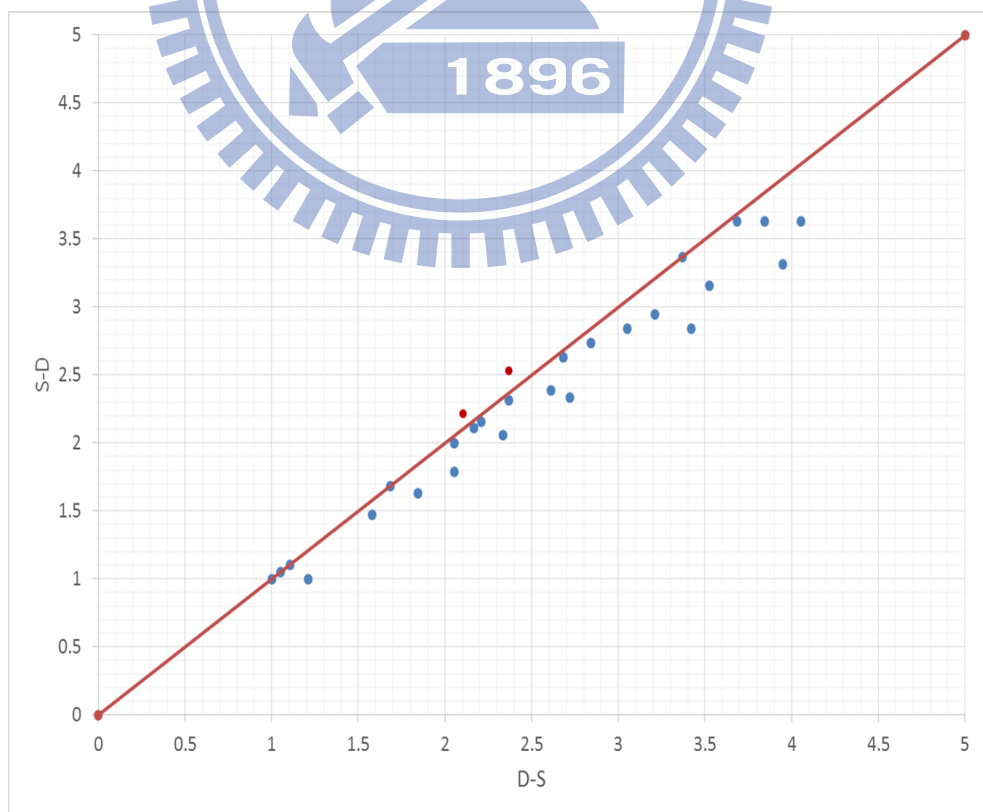


Figure 65. Subjective comparison between D-S and S-D images (3D scenario).

# Chapter 7 Conclusions and Future Works

## 7.1. Conclusions

Objective quality assessment (QA) model is a desirable tool in multimedia applications. It reduces the complexities in measuring the subjective image quality. As 3D contents and applications are widely used, developing objective QA models to assess 3D image becomes an important issue. The ISO/IEC Moving Picture Expert Group (MPEG) is specifying 3D video coding (3DVC) based on the multiple-view plus depth (MVD) format. Each view contains a color image and its corresponding depth map. At the receiver, the virtual view images are generated by a view synthesis algorithm. The virtual view synthesizing process introduces new artifacts, such as shift, ghost, and sticker. These artifacts are different from those of the D distortion. Thus, the conventional pixel-by-pixel 2D objective QA model could not predict the quality of synthesized images precisely. In this thesis, we proposed an objective image QA model for depth map error induced distortions. The model evaluates two scores, the image quality score (IQS) and the edge structural distortion (ESD). IQS computes 2D color quality of the synthesized image and considers object shift. ESD estimates the degree of structural artifact by employing the Hausdorff distance. The final score of the proposed model is obtained by combining IQS and ESD together in the pooling stage. The experimental results show that proposed model has higher PLCC and SROCC (both increase about 20%) and less RMSE and outlier ratio than the 2D models.

For color distortion, we designed a database with two distortion systems based on the

MVD format. They are the distortion-synthesis (D-S) and synthesis-distortion (S-D). From the experimental results, we found out the D-S system has higher SSIM quality score than the S-D system for the Gaussian noise distortion. They show similar subjective quality for the JPEG coding and Gaussian blur. We also found out if we blur the left and right color images before the rendering process, a so-called sticker artifact would appear around the object boundaries.

## 7.2. Future Work

This thesis proposed an objective QA image model to assess the synthesized virtual images with depth map errors. Also, we did some researches on synthesized image quality evaluation due to (texture) color distortion. Developing a QA model to assess color distortion is the next essential work. To generalize the 3D image QA model, a model to evaluate artifacts due to depth and color distortion has to be constructed to replace the 2D QA model. Furthermore, the QA model for 3D video is also an interesting topic to study. In this case, in addition to the depth information, the temporal information needs to be considered.



# References

- [1] K. Muller and P. Merkle, "Challenges in 3D video standardization," *Visual Communications and Image Processing (VCIP)*, pp.1-4, 6-9 Nov. 2011.
- [2] P. Hanhart, F. D. Simone, M. Rerabek, and T. Ebrahimi, "3DV: Objective quality measurement for the 2-view case scenario," Doc. M23908, ISO/IEC JTC1/SC29/WG11, San Jose, USA, February 2012.
- [3] P. Hanhart, F. D. Simone, M. Rerabek, and T. Ebrahimi, "3DV: Alternative metrics to PSNR," Doc. M24807, ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland , May 2012.
- [4] P. Hanhart and T. Ebrahimi, "3DV: Quality assessment of stereo pairs formed from two synthesized views," Doc. M25818, ISO/IEC JTC1/SC29/WG11, Stockholm, SE, 16–20 July 2012.
- [5] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," in *Proceedings of SPIE Stereoscopic Displays and Virtual Reality Systems XI*, 2004, vol. 5291, p. 93104.
- [6] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. BT.500-11, ITU-R, 2002.
- [7] Z. Wang and A. C. Bovik, "Modern Image Quality Assessment. Synthesis Lectures on Image, Video & Multimedia Processing," Morgan & Claypool Publishers, 2006.
- [8] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr. 2004.
- [9] Z. Wang, E.P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conf. Signals, Systems, and Computers*, pp. 1398-1402, 2003.
- [10] Z. Wang, and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing*

*Letters*, vol. 9, no. 3, pp. 81-84, March 2002.

- [11] H.R. Sheikh and A.C. Bovik, "Image information and visual quality," *IEEE Transactions on Image Processing*, vol.15, no.2, pp. 430- 444, Feb. 2006.
- [12] D.M. Chandler and S.S. Hemami, "VSNR: A wavelet-based visual signal-to-noise-ratio for natural images," *IEEE Transactions on Image Processing*, vol. 16, pp.2284-2298, 2007.
- [13] Video Quality Experts Group, "Final Report from the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment," *VQEG*, August 2003.
- [14] G. Lavoue, E. Drelie Gelasca, F. Dupont, A. Baskurt, and T. Ebrahimi, "Perceptually driven 3D distance metrics with application to watermarking," *In Proceedings of SPIE*, vol. 6312, 2006.
- [15] L. Goldmann, F. De Simone, and T. Ebrahimi, "Impact of acquisition distortions on the quality of stereoscopic images," *5th International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM)*, Scottsdale, USA, 2010.
- [16] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, "Quality assessment of stereoscopic images," *EURASIP Journal on Image and Video Processing*, 2008.
- [17] M. Urvoy, J. Gutiérrez, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricordel, P. Le Callet, and N. García, "NAMA3DS1-COSPAD1 : Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," *Fourth International Workshop on Quality of Multimedia Experience*, Yarra Valley, July 2012.
- [18] A. K. Moorthy, C.-C. Su, A. Mittal, and A. C. Bovik, "Subjective evaluation of stereoscopic image quality," *Signal Processing: Image Communication*, 2012.
- [19] E. Bosc, R. Pépion, P. Le Callet, M. Köppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, "Towards a new quality metric for 3-D synthesized view assessment," *IEEE Journal on Selected Topics in Signal Processing*, 2011.
- [20] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the

- Hausdorff distance,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.15, no.9, pp.850,863, Sep. 1993.
- [21] M. P. Dubuisson and A. K. Jain, “A modified Hausdorff distance for object matching,” *International Conference on Pattern Recognition (ICPR)*, pages A:566–568, Jerusalem, Israel, 1994.
- [22] A. K. Moorthy and A. C. Bovik “Visual importance pooling for image quality assessment,” *IEEE J. Special Topics in Signal*, vol.3, no.2, pp.193-201, Apr. 2009.
- [23] J. Park, K. Seshadrinathan, S. Lee, and A. C. Bovik, “Spatio-temporal quality pooling accounting for transient severe impairments and egomotion,” *IEEE International Conference on Image Processing (ICIP)*, 11-14 Sept. 2011.
- [24] ISO/IEC JTC1/SC29/WG11 “Description of core experiments in HEVC based 3D video coding,” Doc. N12354, Geneva, Switzerland , December 2011.
- [25] M. Domanski, T. Grajek, K. Klimaszewski, M. Kurc, O. Stankiewicz, J. Stankowski, and K. Wegner, “Poznan multiview video test sequences and camera parameters,” ISO/IEC JTC1/SC29/WG11 m17050, Xian, China, October 2009.
- [26] “Nagoya university ftv test sequences,” <http://www.tanimoto.nuee.nagoya-u.ac.jp/>
- [27] G. M. Um, G. Bang, N. Hur, J. Kim, and Y. S. Ho, “3D video test material of outdoor scene,” ISO/IEC JTC1/SC29/WG11, Archamps, France, Tech. Rep. M15371, Apr. 2008.
- [28] Y. S. Ho, E. K. Lee, and C. Lee, “Multiview video test sequence and camera parameters,” ISO/IEC JTC1/SC29/WG11, Archamps, France, Tech. Rep. M15419, Apr. 2008.
- [29] [https://hevc.hhi.fraunhofer.de/svn/svn\\_3DVCSsoftware/](https://hevc.hhi.fraunhofer.de/svn/svn_3DVCSsoftware/)
- [30] MeTriX MuX Visual Quality Assessment Package, [http://foulard.ece.cornell.edu/gaubatz/metrix\\_mux/](http://foulard.ece.cornell.edu/gaubatz/metrix_mux/)
- [31] F. Lu, H. Wang, X. Ji, and G. Er, “Quality assessment of 3D asymmetric view coding using spatial frequency dominance model,” *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 4-6 May 2009.

- [32] H.R. Sheikh, Z.Wang, L. Cormack, and A.C. Bovik, "LIVE image quality assessment database release 2," <http://live.ece.utexas.edu/research/quality>.
- [33] H.R. Sheikh, M.F. Sabir, and A.C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440-3451, Nov. 2006.
- [34] Z. Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol.13, no.4, pp. 600- 612, April 2004.

