

# **3D-IQA: Analysis of 3D views for no-reference and full-reference quality assessment**

Thesis submitted for the award of the Degree  
of

**Doctor of Philosophy**

in the Department of Computer Science and Engineering

by

**Sadbhawna**

(2018RCS0013)

Under the supervision of

**Dr. Vinit Jakhetiya**



विद्याधनं सर्वधनं प्रधानम्

भारतीय प्रौद्योगिकी  
संस्थान जम्मू

**INDIAN INSTITUTE OF  
TECHNOLOGY JAMMU**

**Indian Institute of Technology Jammu**

**Jammu 181221**

**September 2022**

# Declaration

I hereby declare that the matter embodied in this thesis entitled “**3D-IQA: Analysis of 3D views for no-reference and full-reference quality assessment**” is the result of investigations carried out by me in the Department of Computer Science, Indian Institute of Technology Jammu, India, under the supervision of **Dr. Vinit Jakhetiya** (IIT Jammu) and it has not been submitted elsewhere for the award of any degree or diploma, membership etc. In keeping with the general practice in reporting scientific observations, due acknowledgements have been made whenever the work described is based on the findings of other investigators. Any omission that might have occurred due to oversight or error in judgment is regretted. A complete bibliography of the books and journals referred in this thesis is given at the end of the thesis.

September 2022

Indian Institute of Technology Jammu

Sadbhawna

2018RCS0013

*To my late grand parents*

*Sh. Capt. Mohan Singh and Smt. Ikadshi Devi*

# Abstract

Image Quality Assessment (IQA) is the analysis of degradation in the images and the effect of these degradations on the overall perceptual quality. 3D views are a type of image which are gaining popularity these days and have applications in various domains such as Free-viewpoint Televisions (FTVs) and Virtual Reality (VR) for an immersive experience. Subsequently, their quality assessment is an important aspect of research in the computer vision domain. The 3D IQA methods can be divided into two categories: Full Reference (FR) and No Reference (NR) IQA metrics are based on the amount of information utilized from the reference image or the side images. Recent 3D synthesis algorithms produce distortions that are not pleasant to human visual systems (HVS), such as stretching artifacts, improper alignment, and various geometric/structural distortions. We analyzed these new types of distortions and how these distortions are different from distortions such as “black-holes,” which are obsolete. Building upon these observations, we have proposed three 3D-IQA algorithms in this thesis which are explained in detail below:

**1. No Reference 3D IQA (Stretching Artifacts Identification for No Reference IQA of 3D Synthesized Images):** Existing quality assessment algorithms consider identifying “black-holes” to assess the perceptual quality of 3D-synthesized views. However, advancements in rendering and inpainting techniques have made “black-holes” artifacts obsolete. Subsequently, 3D-synthesized views frequently suffer from stretching artifacts due to occlusion that, in turn, affects perceptual quality. Existing QA algorithms are found to be inefficient in identifying these artifacts. We found a relationship between the number of blocks with stretching artifacts in view and the overall perceptual quality. With this view, in our **first chapter**, we propose a Convolutional Neural Network (CNN) based algorithm that identifies the blocks with stretching artifacts and incorporates the number of blocks with the stretching artifacts to predict the quality of 3D-synthesized views. To address the challenge with the existing 3D-synthesized views dataset, which has few samples, we collect images from other related datasets to increase the sample size and generalization while training our proposed CNN-based algorithm. The proposed algorithm identifies blocks with stretching distortions and fuses them to predict perceptual quality without reference.

**2. Full Reference 3D IQA 1 (Perceptually Unimportant Information Re-**

**duction and Cosine Similarity based Full Reference IQA for 3D Images):** All IQA methods have their importance, whether reference-less or reference-based. The generation of 3D synthesized images produces a few pixel shifts between reference and 3D synthesized images; hence, they are not properly aligned. And as most full reference IQA methods start with taking the difference/residual of reference and 3D synthesized image, the different image contains much perceptually unimportant information due to this shifting. To address this, in the **second chapter** of the thesis, we propose to use the morphological operation (opening) in the residual image to reduce perceptually unimportant information between the reference and the distorted 3D synthesized image. The residual image suppresses the unimportant information and highlights the geometric distortions that significantly affect the overall quality of 3D synthesized images. We utilized the information in the residual image to quantify the perceptual quality measure and named this algorithm the Perceptually Unimportant Information Reduction (PU-IR) algorithm. At the same time, the residual image cannot capture minor structural and geometric distortions due to the usage of erosion operation. We extract the perceptually important deep features from the pre-trained VGG-16 architectures on the Laplacian pyramid to address this. The distortions in 3D synthesized images are present in patches, and the human visual system perceives even the small levels of these distortions. With this view, we proposed using cosine similarity to compare these deep features between reference and distorted images. We named this algorithm Deep Features extraction and comparison using Cosine Similarity (DF-CS) algorithm. Finally, the pooling is done to obtain the objective quality scores using simple multiplication to both PU-IR and DF-CS algorithms.

**3. Full Reference 3D IQA 2 (Context Region Identification based Quality Assessment of 3D Synthesized Views):** According to recently proposed 3D view synthesis algorithms, the choice of context region for the disocclusion plays a vital role in the perceptual quality of 3D synthesized views. The context region taken from the background of a view produces a perceptually better quality of 3D synthesized views than when the context region is taken from the foreground. With this view, **third chapter** of the thesis is the first effort toward identifying the context region and incorporating this information for the perceptual quality assessment of 3D synthesized views. We observed that the depth energy maps of the 3D synthesized views vary significantly with the change in the context region and subsequently can identify the context region. Hence, in

this work, we propose a new and efficient quality assessment algorithm based upon the variation in the depth of 3D synthesized and reference views, giving two-fold advantages:

1. It can predict the quality based on whether the context region is foreground or not.
2. It is also able to suggest the possible location of distortions. We have proposed two new algorithms for both situations when the context region is foreground or not. The overall predicted score is the direct multiplication of the quality score estimated when the context region is foreground or not.

## Acknowledgements

First of all, I wish to convey my deepest gratitude to my supervisor Dr. Vinit Jakhetiya for believing in me more than I believe in myself. I am very lucky and honored to be his first Ph.D. student. I never expected anyone could help me as much as he did. He worked hard to create platforms so that I could furnish. I will forever appreciate and recall all our discussions, academics and otherwise. I am grateful that he thought I was worth his valuable time. None of my paper submissions and revisions would have been possible if he did not consistently have my back, always helping and encouraging. I am thankful for his continued support and guidance as I embark on my academic journey. In him, I found a mentor for life.

My dissertation would be impossible without the unconditional support, love, inspiration, and encouragement of my parents, Sh. Jaipal Singh, Smt. Nisha Rani and my brother Sadbhav Singh. They provided me with an invaluable platform to grow from and a philosophy to live by and pursue my dreams.

Sh. Capt. Mohan Singh, Smt. Ikadasi Devi. I am always humbled by remembering the roots of their journey. I dedicate this thesis to them for their uncountable sacrifices and unconditional love for the whole family.

I spent a significant portion of six months at K|Lens GmbH with Dr. Sunil Jaiswal. I am deeply grateful for his kind support. Thank you for teaching me super-resolution, how to think concretely about research ideas, and trusting intuitions.

I would also like to thank Prof. Lalit Kumar Awasthi for his blessings and all the career advice since my undergraduate. My master's supervisor, Dr. Major Singh Gorayya, is also a source of inspiration for me throughout my life journey.

I will always appreciate the fruitful research discussions with Dr. Badri Narayan Subudhi and Dr. Sharath Chandra Guntuku. These discussions helped me a lot in my research work. I am also extremely grateful to have Dr. Ankit Dubey and Dr. Yamuna Prasad as my thesis committee members.

I am of full gratitude to my friends and family. I would not have been fortunate enough to be the first doctorate in our family tree had it not been for their support. Although the list is endless, I would like to thank my aunt Smt. Vinod Sheela and Smt. Ranjana Thakur for calling me regularly and reminding me to take care of my health and my studies.

I am extremely grateful to Divyansh for being on my side, encouraging me, keeping me sane at the same time, and being my core strength all these years. Spending time with my dear little Venkie has made me even stronger. As the first child I've seen so closely, his purest heart worked as a stress reliever for me in all the difficult times.

I also want to thank Deebha Mumtaz for being with me from the first day of my Ph.D. She has always been patient enough to listen to my preaches and all my probably unrealistic life goals. She taught me how to get people by understanding their perspectives. Another strength during my Ph.D. journey was Ms. Annu Maheshwari, who has supported me in all my ups and downs like an elder sister. Every Saturday evening we spent together, the delicious food she cooked felt like a second home to me.

I would also like to thank my labmates Aanchana Sharma, Akshita Sharma, Amreen Bashir and Sapna Thapar for their friendship, companionship, and kind support. I'll always enjoy the happiness and pain we went through together during my Ph.D. journey.

I am fortunate to have good friends like Ankita, Sheetal, Priya, Richa, Vibha, Priyanka, Kiran, Hemlata, Supriya, Surayya, Smriti, Archana, and Maheep, whom I met in my student life. Talking to them inspired me in my journey, and I wish all of them a successful and happy life ahead.

Last but not least, I would like to express my deepest regards to Indo-German Science and Technology Center (IGSTC) for awarding me with their fellowship to visit K|Lens, Germany, for six months between my Ph.D. Their untroubled administration made the visit smoother in getting this industrial exposure.



# Contents

<b>Contents</b>	<b>viii</b>
<b>List of Figures</b>	<b>x</b>
<b>List of Tables</b>	<b>xiv</b>
<b>List of Abbreviations</b>	<b>xvii</b>
<b>1 Introduction</b>	<b>2</b>
1.1 <b>Introduction to Image Quality Assessment of 3D Synthesized Views</b>	<b>2</b>
1.2 Depth-Image-Based-Rendering (DIBR)	3
1.3 3D IQA Datasets Survey	4
1.4 Evaluation Metrics	8
1.5 Literature Survey	9
<b>2 Chapter 3</b>	<b>16</b>
2.1 Motivation	16
2.2 Proposed Quality Assessment Metric	19
2.2.1 Stretching Identification using Deep Learning (SI-DL) Model	20
2.2.1.1 Data Collection	22
2.2.1.2 CNN Architecture	23
2.2.1.3 Stretching Artifact based Quality Score	26
2.2.2 Quality prediction using BIQI Metric	27
2.2.3 Scores Pooling	28
2.3 Experimental Results and Analysis	29
2.3.1 Evaluation Protocols	29

2.3.1.1	3D Synthesized Views Dataset . . . . .	29
2.3.1.2	Evaluation Criteria . . . . .	30
2.3.2	Parameters Sensitivity Analysis . . . . .	32
2.3.3	Performance Comparison and Analysis . . . . .	34
2.3.4	Statistical Significance Analysis . . . . .	37
2.3.5	Time Complexity Comparison . . . . .	38
2.4	Application in the enhancement of 3D views . . . . .	38
2.5	Conclusions and Future Work . . . . .	39
<b>3</b>	<b>Chapter 4</b>	<b>43</b>
3.1	Proposed Algorithm . . . . .	45
3.1.0.1	Perceptually Unimportant Information Reduction (PU-IR)	46
3.1.0.2	Deep Features extraction and comparison using Cosine Similarity (DF-CS) . . . . .	50
3.1.0.3	Scores Pooling . . . . .	54
3.2	Experimental Results & Analysis . . . . .	55
3.2.1	3D Synthesized images Dataset . . . . .	55
3.2.2	Performance Comparison and Analysis . . . . .	56
3.2.3	Statistical Significance . . . . .	59
3.2.4	Parameter Sensitivity Analysis . . . . .	60
3.3	Conclusions and Future work . . . . .	61
<b>4</b>	<b>Chapter 5</b>	<b>67</b>
4.1	Proposed Methodology . . . . .	69
4.1.1	Quality Score when context region is the foreground (Step 1) . . . .	70
4.1.2	Quality score when the context region is background (Step 2) . . .	72
4.1.3	Final Perceptual Quality Score Pooling . . . . .	74
4.2	Experimental Results . . . . .	75
4.2.1	Dataset and evaluation criteria for performance comparison . . . .	75
4.2.1.1	IETR dataset . . . . .	75
4.2.1.2	IVY dataset . . . . .	76
4.2.2	Performance Analysis . . . . .	77
4.2.3	Parameters Sensitivity Analysis . . . . .	78

4.2.4	Statistical Significance . . . . .	79
4.2.5	Existing 3D IQA algorithms as a plug-in for the proposed algorithm.	80
4.3	Conclusions and future work . . . . .	80
<b>5</b>	<b>Conclusions and Future Work</b>	<b>89</b>
5.1	Future Work . . . . .	89
5.2	Conclusions . . . . .	91
	<b>Bibliography</b>	<b>91</b>
	<b>List of Publications</b>	<b>106</b>

# List of Figures

1.1	This figure is taken from <a href="http://global.canon/en/news/2017/20170921.html">http://global.canon/en/news/2017/20170921.html</a>	2
1.2	Real-world example of 3D Views and its corresponding zoomed-in patches.	3
1.3	A general process of 3D synthesis using warping process and filling of black holes using inpainting algorithms. . . . .	4
1.4	Some example views from the three datasets with the primary distortions in them zoomed in. . . . .	6
1.5	Performance Evaluation Metrics . . . . .	7
2.1	Stretching artifact identification using [1]. The plot represents each column's average Horizontal Gradient (AHG) in the given view from IETR Dataset [2]. Purple and red colors indicate stretching artifacts and corresponding AHG in the middle and corner of the view, respectively. A red double arrow indicates Stretching Width (SW) as defined in [1] . . . . .	17
2.2	Coarse stretching region map and its column-wise mean values of the image in Figure 1, as proposed in [3]. . . . .	18
2.3	The workflow of the proposed method. . . . .	20
2.4	A view from IETR Dataset divided into blocks, the red window shows blocks with stretching artifacts. . . . .	20
2.5	Class-1: Examples of blocks (dimensions: $160 \times 160$ ) with Stretching Artifacts. . . . .	21
2.6	Class-2: Examples of blocks (dimensions: $160 \times 160$ ) without Stretching Artifacts. . . . .	22
2.7	An elaborated workflow of the proposed Stretching Identification using Deep Learning (SI-DL) Model. . . . .	24
2.8	Plot of variation in training and validation accuracy with every epoch. . .	26

2.9	Parameter Sensitivity. (a) and (b) shows the performance of the proposed algorithm with varied parameters $x$ , $y$ , and $u$ , $v$ for Natural Views and Synthetic Views, respectively. . . . .	33
2.10	Scatter Plot of DMOS values and objective scores of state-of-the-art IQAs. . . . .	35
2.11	Location identification of blocks with stretching artifacts using proposed SI-DL model. . . . .	38
3.1	Predicted distortion maps using various algorithms. (a) and (b) are a reference and its corresponding 3D synthesized image from IETR Dataset. (c)-(f) are the distortion maps of the image predicted using four existing algorithms in the literature. . . . .	44
3.2	Elaborated Perceptually Unimportant Information Reduction (PU-IR) method with highlighted (perceptually important and unimportant) distortions. . . . .	46
3.3	Perceptually Unimportant Information Reduction map ( $DO$ ) analysis using the histogram. . . . .	48
3.4	Laplacian Levels wise highlighted distortions in the reference and distorted images. RLL stands for Reference image Laplacian Level, and DLL stands for Distorted image Laplacian Level. Please note that RLL(2-4) and DLL(2-4) are of different resolutions, but these are shown in equal size for better visualization in this figure. . . . .	49
3.5	Effect of Laplacian on the distortions of an image from IETR Dataset. . . . .	49
3.6	An elaborated workflow of the proposed Deep Features fusion using Cosine Similarity (DF-CS). . . . .	51
3.7	Scatter Plot between DMOS Values and Objective Scores of the 3D synthesized IQAs for the IETR dataset. . . . .	57
3.8	Effect of variation of parameter $\gamma$ on the performance of the proposed PU-IR algorithm and the proposed overall algorithm (equation (13)). . . . .	58
3.9	Sensitivity analysis of parameters $\lambda_1$ and $\lambda_2$ on the performance of the proposed algorithm. . . . .	59
3.10	Performance variation of the proposed algorithm with change in Structuring Elements (SE). . . . .	66

4.1	Workflow diagram of the proposed algorithm when the context region is foreground. . . . .	68
4.2	Workflow Diagram of Step 2 of the proposed algorithm when context region is background. This diagram shows the images visually after applying operations described in Step 2, such as morphological dilation, masking out, and discrete cosine transform. . . . .	71
4.3	Masking out operation flow in detail. . . . .	73
4.4	Workflow Diagram of the proposed model. . . . .	75
4.5	Scatter Plot between DMOS Values and Objective Scores of different IQA methods. . . . .	76
4.6	Performance dependency of the proposed algorithm with variation in Structuring Elements. Here, ‘r’ is the radius, and ‘w’ is the width in terms of pixels. . . . .	79
5.1	(a). A synthesized view. (b). The failure (green arrows) of a random patch (red window) in a 3D synthesized view. Synthesized Using: [4] . . . . .	89
5.2	Step-wise flow of the proposed future work. . . . .	90

# List of Tables

1.1	Comparison of existing IQA datasets. . . . .	5
1.2	Summary of existing Full-Reference IQA's in the literature . . . . .	10
1.3	Summary of existing No-Reference IQA's in the literature . . . . .	13
2.1	Data collection description in detail . . . . .	23
2.2	Train-Validation splits. . . . .	23
2.3	The proposed VGG-16 fine-tuned architecture. FC stands for Fully Connected Layers. Conv stands for Convolution Kernel. . . . .	25
2.4	Performance metric values on validation data with varying loss functions .	26
2.5	Effect of varying block sizes on performance metrics using proposed pooling.	34
2.6	Performance comparison after pooling with the existing NR-IQA metrics. .	34
2.7	Performance comparison of various algorithms (in terms of PLCC, SROCC, KRCC, and RMSE). The table is arranged in descending order of PLCC. The symbol "-" indicates the unavailability of source code or reference resources, and "◇" indicates the results are taken directly from the original research papers. "Δ" indicates that official source codes were available at the time of experimentation, and "†" indicates the results are taken from experiments of research papers. . . . .	40
2.8	Performance metrics comparison individually for Natural and Synthetic View types for IETR Dataset. . . . .	41
2.9	Stage-wise performance evaluation of the proposed algorithm. . . . .	41
2.10	Performance comparison of various 3D IQA algorithms for IRCCyN Dataset and IVY Dataset. ("◇", "Δ", "†", "-" indicates same meaning as in Table 2.7.) . . . . .	41

2.11	Performance of the proposed algorithm on different DIBR synthesis algorithms used in IETR dataset. . . . .	42
2.12	Statistical Significance (SS) Table for comparison between the proposed algorithm and existing state-of-the-art IQAs. . . . .	42
2.13	Time taken (in seconds) by objective 3D IQA metrics. . . . .	42
3.1	Performance comparison of the objective IQA metrics on IETR Dataset sorted in descending order of PLCC values. - indicates that either the information is missing in the literature or the reference data is unavailable. * indicates the performance was evaluated on a subset of the IETR dataset in the source paper. . . . .	63
3.2	Step-wise performance analysis of the proposed algorithm. . . . .	64
3.3	Dependency of DF-CS algorithm on different Laplacian Levels. . . . .	64
3.4	Effect of various metrics on fusion of deep-features. PP stands for Performance Parameter . . . . .	64
3.5	Performance comparison of the proposed algorithm with existing algorithms on IRCCyN dataset, sorted in descending order of PLCC values. . .	65
3.6	Statistical Significance (SS) comparison of the proposed algorithm with existing state-of-the-art IQA algorithms for the IETR dataset. . . . .	65
4.1	The literature survey shows how depth information has been used earlier for the quality assessment of 3D synthesized images. . . . .	82
4.2	effect of context region as foreground and background on depth and energy map of the 3D-synthesized views. Here, Synthesized View-1 (SV-1) and Synthesized View-2 (SV-2) are the case of context region from foreground and background, respectively. CC stands for Correlation Coefficient . . . .	84
4.3	Analysis of the effect of context region as foreground (FG) and background (BG) on the depth maps. . . . .	85
4.4	comparison for the performance of the proposed algorithm with existing IQA algorithms on the IETR dataset. Unavailability of data is indicated using the '-' symbol. '↑' indicates a higher value is better while '↓' indicates a lower value is better. . . . .	86
4.5	Ablation study of the proposed algorithm. . . . .	87



4.6	Comparison of the proposed algorithm with existing algorithms on the IVY dataset for performance. Unavailability of data is indicated using the ‘-’ symbol. . . . .	87
4.7	Comparison of the proposed algorithm when different edge detection methods are used for edge detection. . . . .	87
4.8	The proposed algorithm as a plug-in to improve the performance of the existing algorithms on the IETR dataset. . . . .	88
4.9	Comparison of the proposed algorithm with different edge detection methods.	88
5.1	Performance of state-of-the-art IQA algorithms for the proposed test dataset. . . . .	90

# List of Abbreviations

Abbreviation	Description
AHG	Average Horizontal Gradient
AVG	Average Vertical Gradient
AR	Auto Regression
APT	Auto Regression Plus Thresholding
BIQI	Blind Image Quality Index
BRISQUE	Blind Referenceless Image Spatial Quality Evaluator
BG	Background
CS	Cosine Similarity
CNN	Convolutional Neural Network
CLGM	Combining Local and Global Measures
CODIF	COLOR-Depth Image Fusion
DIBR	Depth Image Based Rendering
DSCB	Distortion Specific Contrast Based
DCT	Discrete Cosine Transform
DF-CS	Deep Features Cosine Similarity
DO	Dilation Operation
DMOS	Differential Mean Opinion Score
DLL	Distorted Laplacian Level
EO	Erosion Operation
FVV	Free View-point Video
FG	Foreground
FR	Full Reference
GANs	Generative Adversarial Networks
HVS	Human Visual System
HHF	Hierarchical Hole Filling
HED	Holistically-Nested Edge Detection
IQA	Image Quality Assessment
IDEA	Instance DEgradation and global Appearance
IR	Image Restoration
KRR	Kernel Ridge Regression
KRCC	Kendall Rank Correlation Coefficient

Abbreviation	Description
LPIPS	Learned Perceptual Image Patch Similarity
LVGC	local variation and Global Change
LDI	Layered Depth Image
MSE	Mean Square Error
NR	No Reference
NIQE	No Reference Image Quality Evaluator
NVS	Novel View Synthesis
PLCC	Perason Linear Correlation Coefficient
PSNR	Peak Signal to Noise Ratio
PU-IR	Perceptually Unimportant Information Reduction
RLL	Reference Laplacian Level
SROCC	Spearman Rank Order Correlation Coefficient
SAMVIQ	Subjective Assessment Methodology for Video Quality
SSIM	Structural SIMilarity
SV	Synthesized View
SVR	Side View Reference
SIFT	Scale Invariant Feature Transform
SI-DL	Stretching Identification Deep Learning
SR	Super Resolution
SS	Statistical Significance
VR	Virtual Reality
VSRS	View Synthesis Reference Software

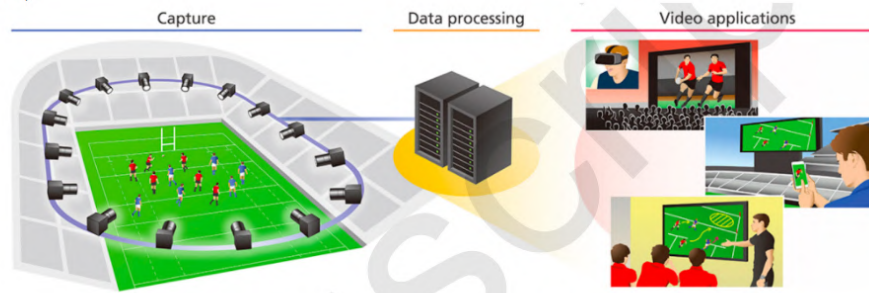
# Chapter 1

## Introduction

### 1.1 Introduction to Image Quality Assessment of 3D Synthesized Views

An appropriate 3D view can provide consumers with a more engaging and better immersive experience [5]. 3D-Television, Free-Viewpoint-Video (FVV), Virtual-Reality (VR), and 360° video are popular applications of 3D view synthesis widely used due to their realistic and interactive experience [6, 7]. Figure 1.1 is a visual description of an application of 3D view synthesis during sports events.

Free view-point videos (Eg. users can view sporting events from various different angles and viewpoints)



Example of Canon's Free Viewpoint Video System.

(a)

Figure 1.1: This figure is taken from <http://global.canon/en/news/2017/20170921.html>

The real-world 3D images look like Figure 1.2, which is generated using the Facebook 3D algorithm [8]. (a) and (c) are two viewpoints of the figure, whereas (b) and (d) are their zoomed-in views, respectively. As can also be observed from the figure, these zoomed-in

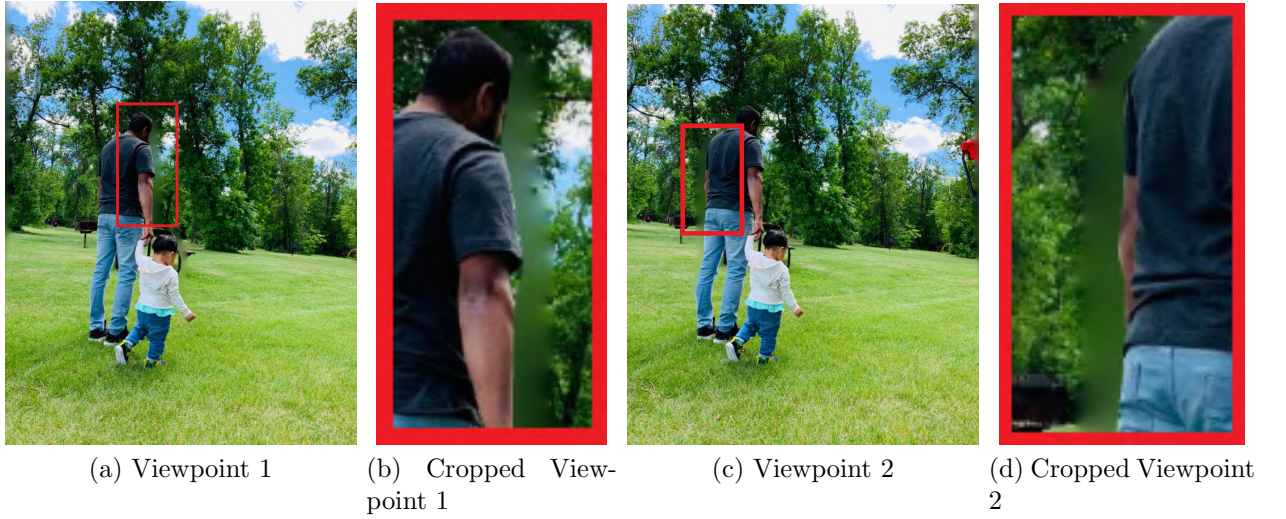


Figure 1.2: Real-world example of 3D Views and its corresponding zoomed-in patches.

figures show visually unpleasing distortions to the human visual system (HVS). Similarly, different rendering methods produce different unpleasing artifacts. Hence, there is a need for improvement in the area of Image quality assessment of 3D view synthesis.

## 1.2 Depth-Image-Based-Rendering (DIBR)

DIBR [9] is a powerful method used to represent and code new scenes in the process of 3D view synthesis. In the process of DIBR, a virtual/novel view of a scene is synthesized using still or moving images and their per-pixel depth information. Elaborating further, the DIBR is a process that consists of two following steps:

1. Reprojection of original image points into the 3D world. This can be done using depth information of the image.
2. Projection of 3D space points to virtual camera points at the desired location.

This process of projection (2D-to-3D) and (3D-to-2D) is termed 3D warping in Computer Graphics. 3D warping is followed by image inpainting techniques to fill the missing information in the warping process. This process of 3D synthesis is shown in Figure 1.3. This process of rendering new scenes introduces new types of artifacts such as cracks, ghosting, stretching, flickering and crumbling, etc. [5]. These artifacts differ from conventional ones, which occur in regular natural images. Also, with technological advancement

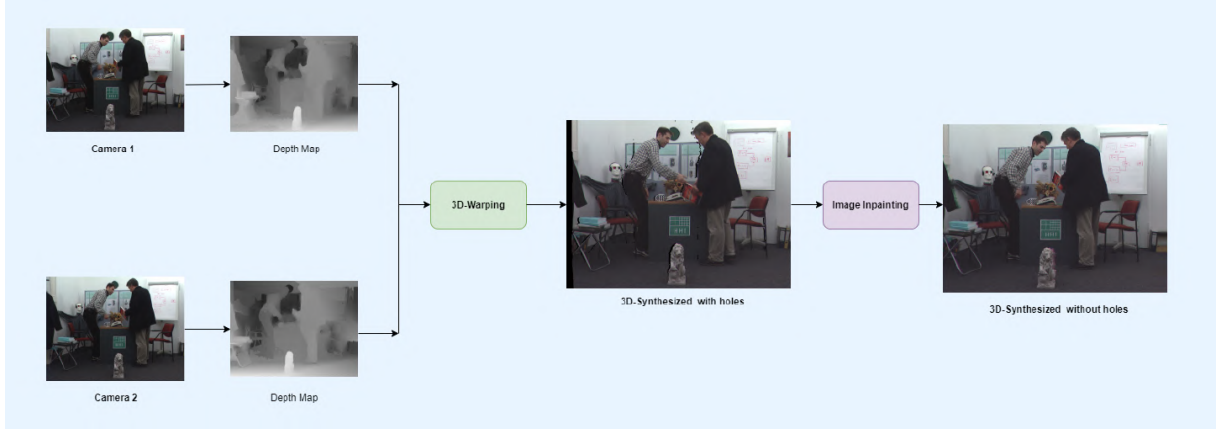


Figure 1.3: A general process of 3D synthesis using warping process and filling of black holes using inpainting algorithms.

and 3D-synthesis methods, some of these traditional 3D-synthesis artifacts, such as black holes [10], have become obsolete [2].

### 1.3 3D IQA Datasets Survey

**IETR DIBR Dataset:** This dataset comprises 140 synthesized views generated using ten reference views and corresponding subjective scores. Out of these ten reference views, 7 are Natural Views, and 3 are Synthetic Views. The views are rendered using 7 different DIBR methods M1: Criminisi’s [11], M2: LDI [12], M3: Ahn [13], M4: Luo’s [14], M5: HHF [15], M6: VSRS [16], M7: Zhu [17]. Among these 7 DIBR methods (M1 to M7), M7: Zhu is an inter-view 3D synthesis method, while M6: VSRS can be used as both inter-view and single-view 3D synthesis; all the other methods (M1 to M5) belong to the single-view 3D synthesis category. A brief overview of these methods is listed below:

i). Single-view 3D synthesis Methods:

- M1 (Criminisi’s [11]): This method is based on an exemplar-based texture synthesis technique. Patch priorities are computed using *confidence* parameter to improve the order of the pixel filling.
- M2 (LDI [12]): This algorithm uses an object-based Layered-Depth-Image (LDI) representation to obtain the synthesized view. Based on this representation’s foreground and background segmentation, the authors have proposed to render the 3D-synthesized view.

Dataset	IVC	IVY	IETR
Dataset proposed in year	2011	2016	2019
Number of synthesized Views	84	84	140
3D synthesis algorithms used	4	7	7
Year of most recent 3D algorithms	2010	2014	2016
Obsolete distortions?	Yes	Yes	Yes
Type of obsolete distortion	Black Holes	Ghosting	Stretching

Table 1.1: Comparison of existing IQA datasets.

- M3 (Ahn [13]): A depth-based 3D synthesis method was proposed by Ahn *et al.* using patch-based texture synthesis.
- M4 (Luo’s [14]): This method is proposed based on background reconstruction. A random walker segmentation technique was employed using a detected initial seed.
- M5 (HHF [15]): Sohl *et al.* proposed two approaches, Hierarchical Hole-Filling (HHF) and Depth Adaptive Hierarchical-Hole-Filling for filling dis-occluded regions.

ii). Inter-view 3D synthesis Methods:

- M6 (VSRS [16]): The MPEG 3D video Group has adopted View-Synthesis-Reference-Software (VSRS) as a standard. A post filter is applied on depths to solve depth discontinuities. Then the holes are filled using inpainting.
- M7 (Zhu [17]): In this algorithm, Zhu *et al.* proposed to identify the background pixels and unoccluded background around the holes. Finally, these holes are filled using depth-enhanced Criminisi’s method and simple block-average filling method.

**IRCCyN/IVC Dataset:** [18] This dataset comprises of 84 synthesized views. These views are generated from 3 different references. All three reference views are Natural Views. A total of 7 rendering methods are used in this dataset. The point to be noted in this dataset was that one rendering method was to fill the holes with the black pixels simply. Hence, “black-holes” are a dominant distortion of this dataset.

**IVY Dataset:** The IVY dataset [19] is also a test dataset of the stereo images. The authors used publicly available datasets to select the reference images, such as the Middlebury dataset (Aloe, Dolls, Reindeer, and Laundry), video-plus-depth sequences provided by MPEG 3DV ad hoc group ( Lovebird1, Newspaper, and Bookarrival). Further, the stereo images were generated using four view synthesis algorithms:



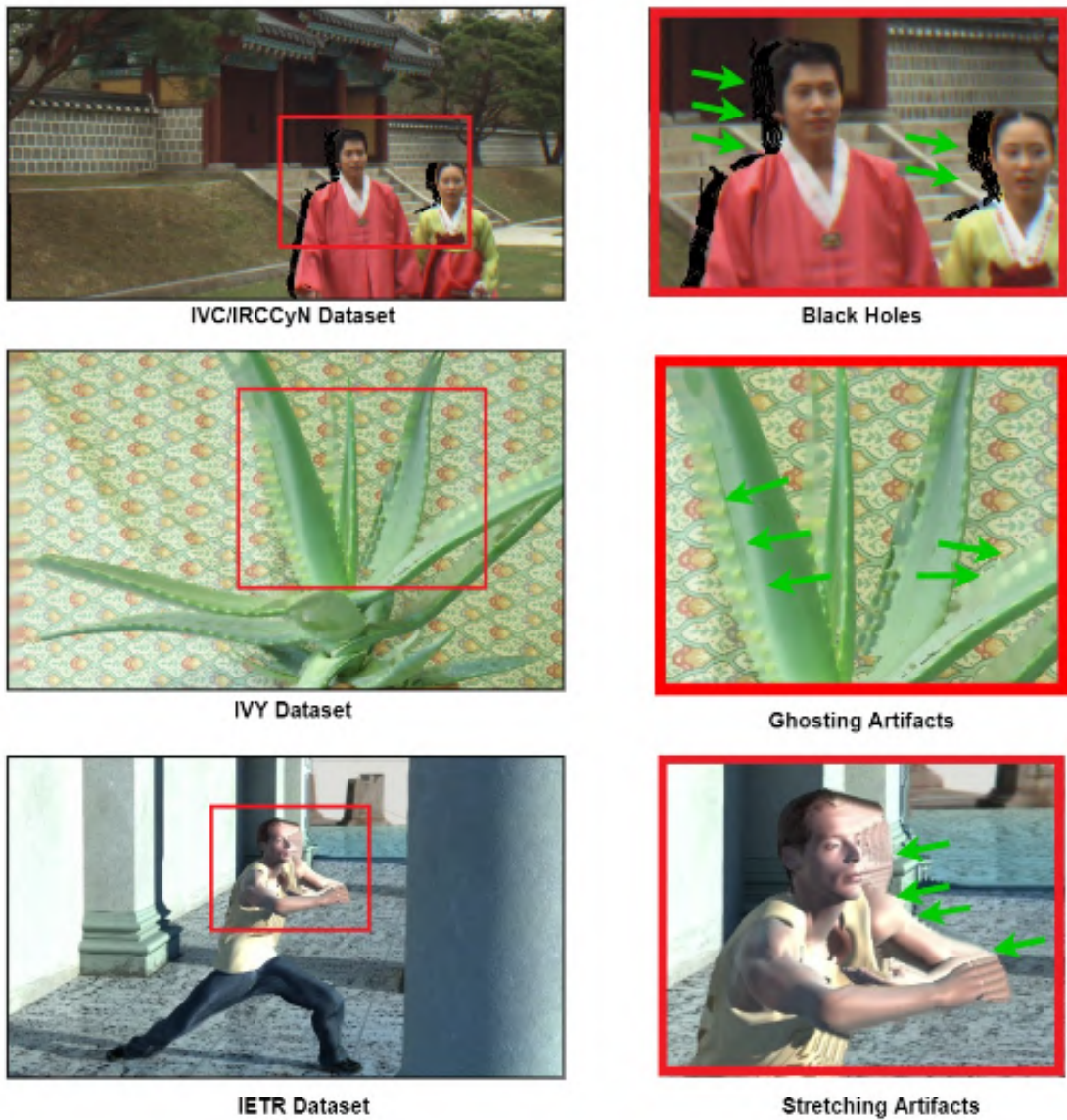


Figure 1.4: Some example views from the three datasets with the primary distortions in them zoomed in.

1. MPEG VSRS version 3.5
2. Criminisi's method [11]
3. Ahn's method [13]
4. inter-view consistent inpainting method

**MCL-3D Dataset:** In MCL-3D Dataset [20], nine image and depth views are selected. Then different distortions are applied to the images or the depths before rendering the stereoscopic images. These distortions include down-sampling blur, additive white



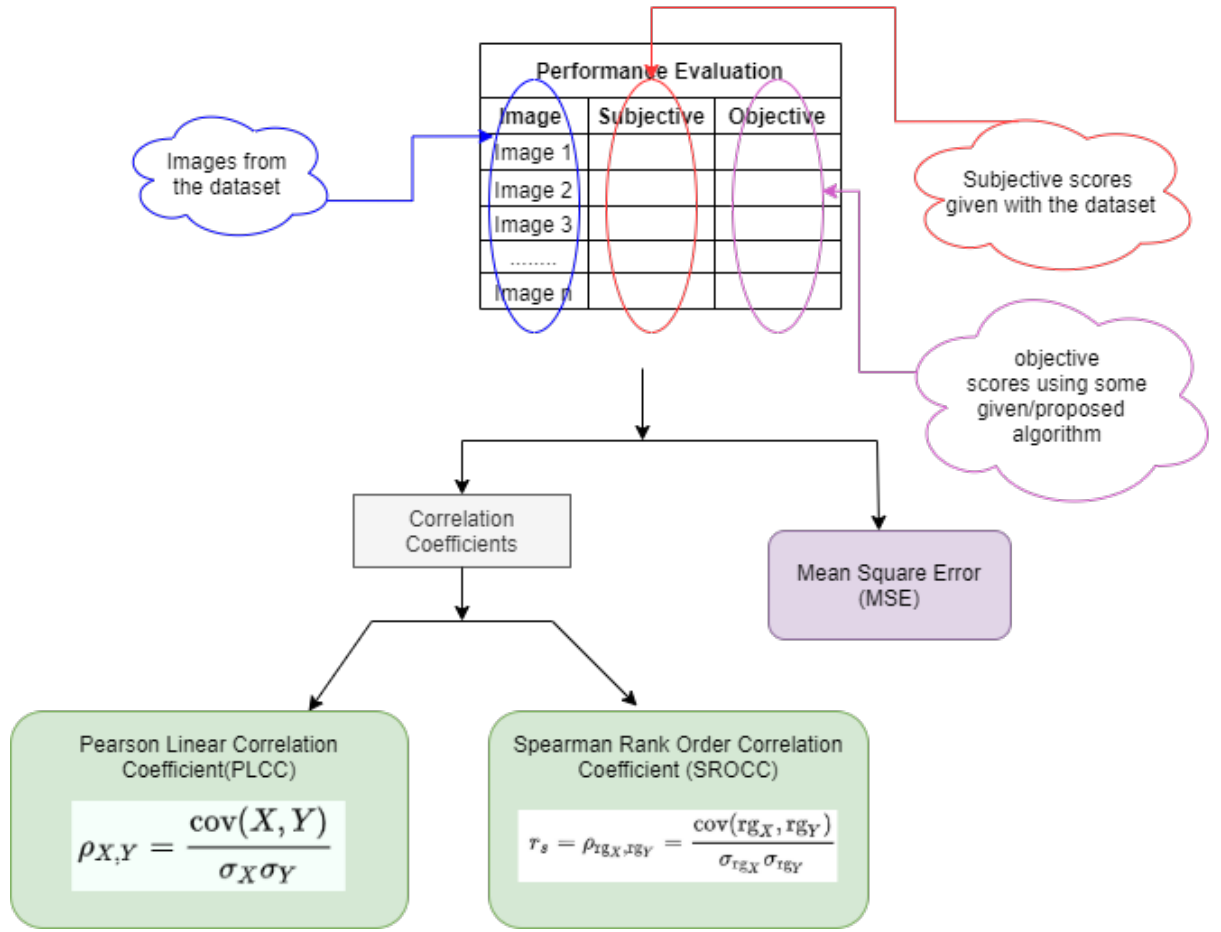


Figure 1.5: Performance Evaluation Metrics

noise, JPEG and JPEG-2000 (JP2K) compression, Gaussian blur, and transmission error. Hence, the dataset contains 693 image pairs of resolutions ranging from 1024x728 to 1920x1080. Mean Opinion The score (MOS) was computed using subjective testing.

**3D IQA Datasets Discussion:** Amongst all the datasets discussed above, the IETR dataset is the state-of-the-art 3D IQA dataset. In IRCCyN and MCL-3D black holes are the main distortions that are obsolete these days [2]. Moreover, in synthesizing 3D views of the IVY dataset, only four types of old algorithms are used. Hence this thesis mainly focuses on distortions present in the IETR dataset, which are stretching, blockiness, blurriness, etc. Some example views from three datasets, i.e., IRCCyN/IVC, IVY, and IETR datasets in Figure 1.4.

## 1.4 Evaluation Metrics

The performance of any image quality assessment metric can be explained in Figure 1.5. For this process, mainly three methods are used. In this section, we will describe them one by one.

1. **Pearson-Linear-Correlation-Coefficient (PLCC):** It is also named Pearson R statistical test. PLCC can measure the strength of two variables. The formulation of PLCC is as follows:

$$PLCC = \frac{\sum(a_i - \bar{a})(b_i - \bar{b})}{\sum(a_i - \bar{a})^2(b_i - \bar{b})^2} \quad (1.1)$$

where  $a_i$  are the values of the x-variable of the first sample.

$\bar{a}$  is mean of the values  $a_i$ .

$b_i$  are the values of the y-variable of the second sample.

$\bar{b}$  is the mean of the values  $b_i$ .

2. **Spearman-Rank-Correlation-Coefficient (SRCC):** SRCC uses a monotonic function to describe the relationship between two variables. This is done by measuring the non-parametric rank correlation.

$$SRCC = 1 - \frac{6 \sum(a_i^2)}{k(k^2 - 1)} \quad (1.2)$$

Here,  $a_i$  is the difference between the ranks of two variables.

k is the total number of observations in the variables.

3. **Kendall-rank-correlation-coefficient(KRCC):** KRCC is commonly referred to as  $\tau$  coefficient. It is used to measure the ordinal association between two variables statistically.

$$\tau = \frac{(A - B)}{(A + B)} \quad (1.3)$$

Here, A is the number of concordant pairs

B is the number of discordant pairs.

4. **Root-Mean-Square-Error(RMSE):** RMSE uses the euclidean distance to calculate the distance of true values from the predicted values. RMSE value can be calculated using the following formula:

$$RMSE = \sqrt{\left(\frac{\sum_{i=1}^K (a_i - \hat{a})^2}{K}\right)} \quad (1.4)$$

Here,  $i$  is variable.  $K$  is the number of non-missing data points.  $a_i$  are true values.  $\hat{a}_i$  are predicted values.

## 1.5 Literature Survey

The 3D IQA methods can be divided into two types based on the amount of information of reference views available when evaluating quality metrics.

### 1. Full-Reference IQAs:

- **SSPD [21]:** Mahmoudpour *et al.* [35] proposed to quantify the local differences using feature matching technique. Further, the gradient difference in image superpixels is measured to quantify the global loss.
- **IDEA [22]:** Li *et al.* [22] proposed Instance-DEgradation-and-global Appearance (IDEA) with the purpose that local distortions generally occur around instance contours and hence dominate the occurrence of distortions. These local distortions are then measured using discrete orthogonal moments, and the global distortions are measured using super-pixel representations.
- **LPIPS [23]:** LPIPS: Learned-Perceptual-Image-Patch-Similarity (LPIPS) [23] metric is based on deep features trained on the ImageNet dataset. We used their trained model on VGG features for comparison.
- **LOGS [36]:** LOGS stands for LOfcal-Geometric-distortions-in-disoccluded-regions-and-global-Sharpness [36]. SIFT flow-based warping is first used to detect the disoccluded regions. Then the global sharpness is quantified using a re-blurring-based strategy.
- **Tian's [25]:** Tian *et al.* [37] observed the statistical features of the reference image and its corresponding 3D-synthesized image using wavelet sub-bands to

Table 1.2: Summary of existing Full-Reference IQA's in the literature

<b>IQA</b>	<b>For</b>	<b>Publisher</b>	<b>Year</b>	<b>Remarks</b>	<b>Drawback</b>
SSPD [21]	3D Images	IEEE SPL	2021	SURF feature mapping	Time Consuming, doesn't calculate shift efficiently.
IDEA [22]	3D Images	IEEE TMM	2021	Instance Degradation and global appearance	performs poorly on IETR
LPIPS [23]	Natural Images	IEEE CVPR	2017	Pre-trained Deep Features	block-wise not helping in 3D
LOGS [24]	3D Images	IEEE TIP	2017	Local Geometric and Global Sharpness	performs poorly on IETR, parameters
Tian's [25]	3D Images	IEEE TIP	2017	Wavelet Sub-bands	hand crafted features, not generalizable
MP-PSNR [26]	3D Images	IEEE QoMEX	2012	Morphological Pyramids	poor performance
MW-PSNR [27]	3D Images	IEEE QoMEX	2012	Morphological Wavelets	poor performance
MS-SSIM [28]	Natural Images	IEEE TIP	2008	Multi-Scale Structural Similarity	Focuses global structural distortions only.
SSIM [29]	Natural Images	IEEE TIP	2004	Structural Similarity	Focuses global structural distortions only.
SC-IQA [30]	3D Images	IEEE VCIP	2018	Shift Compensation	cannot detect shift completely
LMS [31]	3D Images	Elsevier	2019	Structural representation	hand crafted features, not generalizable
PRSI [32]	3D Images	IEEE ICIP	2019	Perceptual Representations of Structural Information	Black-holes oriented
SR-VQA [33]	3D Videos	IEEE TIP	2019	Sparse Representation	hand crafted features, not generalizable
EM-IQA [34]	3D Images	IEEE MM	2017	Elastic Metric	Black-holes oriented

detect the artifacts in the views.

- **MP-PSNR [26]:** Sandić-Stanković *et al.* [38] proposed to decompose the views into multi-scale pyramids using morphological pyramids for quality prediction.
- **MW-PSNR [27]:** Sandić-Stanković *et al.* [27] used morphological filters to maintain low-level features such as edges over multiple levels. These levels are obtained using wavelet decomposition.
- **SC-IQA [30]:** In this work, the authors analyzed that there is object shifting and geometric distortions in 3D synthesized views. They proposed a shift-compensation-based-image-quality-assessment-metric(SC-IQA) using a global geometric shift calculation. This shift is calculated using SURF and RANSAC homography approaches. A visual saliency map is also used as a weighting function to calculate the overall perceptual quality.
- **LMS [31]:** Low-level-and-Mid-level-Structural-representation (LMS) constructs a scale space to mimic the hierarchical property of the human visual system (HVS). Then the statistics of gradient orientation are integrated with the statistics of gradient intensity for the low-level structural representation. Finally, a distance between the low-level and mid-level features is calculated and used as a quality score.
- **PRSI [32]:** Perceptual-Representations-of-Structural-Information(PRSI) is based on hierarchical representation within HVS. This metric is based on low-level contour descriptors, mid-level category descriptors, and task-oriented non-natural structure descriptors.
- **SR-VQA [33]:** Sparse-Representation-Based-Video-Quality-Assessment-for-Synthesized-3D-Videos(SR-VQA). This method treats the video as a 3D volume data as spatial and temporal domains. Then gradient and strong edges of the depth map are key features. These features can detect the location of flicker distortions. Further, a rank pooling method is used to pool all the temporal layer scores. This score is the flicker distortion. The final quality score is based on this flicker distortion measurement.
- **EM-IQA [34]:** It is an elastic-metric-based-image-quality-assessment(EM-

IQA). In this work, a local distortion region is first selected, and then deformations of curves are quantified. This metric can measure the difference in stretching or bending between two curves.

## 2. No-Reference IQAs:

- **CODIF [39]:** Li *et al.* [39], proposed Color-Depth-Image-Fusion-(CODIF) which is a NR-3D IQA algorithm. An image fusion base on wavelet is proposed in CODIF to emulate the relationship between color and depth images. Then their statistical features were used to learn the proposed quality prediction model.
- **GANs-NRM [40]:** In [52], Suiyi *et al.* proposed a Generative-Adversarial-Networks(GAN) based NR-3D-IQA metric called, GANs-NRM that used a Bag-of-Distortion-Word (BDW) feature extraction method on synthetic data rendered using a GANs-based context renderer.
- **Wang’s [53]:** In [54], Wang *et al.* proposed NR 3D quality assessment for videos which could also be translated for the image domain. The authors measured the motion difference between consecutive frames in the optical flow method to quantify the temporal inconsistency. Then the pixel differences of optical flow fields are weighted using structural similarity values.
- **Yan’s [42]:** To measure quality-aware features considering ”local-variation-and-global-change (LVGC)”, Yan *et al.* [42] developed a 3D-synthesized views quality assessment algorithm. In the LVGC metric, structure and chromatic luminance features are extracted to measure local and global variations.
- **Yue’s [3]:** Yue *et al.*, [3] proposed NR quality metric which focuses on two types of distortions, i.e., sharpness and geometric distortions. Further, distorted regions and stretching are detected by calculating the local similarity. The overall sharpness is measured using its downsampled image. Finally, linear pooling is done to merge them.
- **NIQSV+ [43]:** In [1], Tian *et al.* proposed a No-reference-Image-Quality-assessment-method-for-3D-Synthesized-Views (NIQSV+). The quality of synthesized views is measured by quantifying specific distortions such as black

Table 1.3: Summary of existing No-Reference IQA's in the literature

<b>IQA</b>	<b>For</b>	<b>Publisher</b>	<b>Year</b>	<b>Remarks</b>	<b>Drawback</b>
CODIF [39]	3D Images	IEEE TMM	2021	Color-Depth Image Fusion	Pre-DIBR IQA Algorithm
GANs-NRM [40]	3D Images	IEEE TIP	2020	Generative	Trained on black-hole artifacts
Wang's [41]	3D Images	IEEE TIP	2020	Wavelet Transform	IRCCyN Dataset oriented, parameters
Yan's [42]	3D Images	IEEE TIP	2020	Local Structure and Global naturalness	IRCCyN Dataset oriented, parameters
Yue's [3]	3D Images	IEEE TIP	2019	Local and Global Measures	Designed for predicting black holes.
NIQSV+ [43]	3D Images	IEEE TIP	2019	Stretching and Black hole identification	Designed for predicting black holes.
Jakhetiya's [44]	3D Images	IEEE TIP	2018	Auto Regression + Thresholding (APT)	Designed for predicting black holes.
BRISQUE [45]	Natural Images	IEEE TIP	2013	Spatial Domain-based NSS	Focuses global structural distortions only.
BIQI [46]	Natural Images	IEEE SPL	2012	Wavelet Transform based NSS	Focuses global structural distortions only.
NIQSV [1]	3D Images	IEEE ICASSP	2017	Edge Image using morphology	simple but Poor Performance
MNSS [47]	3D Images	IEEE TB	2020	new NSS features oriented for 3D views	Hand-crafted features are not reliable
$NR_MWT$ [48]	3D Videos	IEEE TIP	2019	Wavelet sub-bands with threshold	Hand-crafted features
GDIC [49]	3D Images	IEEE ICASSP	2018	Wavelet sub-bands	Hand-crafted features
CSC-NRM [50]	3D Images	IEEE ICIP	2018	Convolutional Sparse Coding (CSC)	simple but Poor Performance
SIQA-CFP [51]	3D Images	IEEE ICIP	2019	Contextual Multi-Level Feature Pooling	deep features but poor performance

holes, blurry regions, stretching artifacts (among-out-of-field-areas (OOFAs) [8, 55, 56]), and finally merged using pooling.

- **APT [57]:** Gu *et al.* [57], suggested an auto-regression-plus-thresholding(APT) based algorithm for NR IQA of 3D-synthesized views. In this method, the local image description is predicted using auto-regression over small windows and then thresholding.
- **Jakhetiya's [44]:** [44] used Kernel-Ridge-Regression(KRR) as a global predictor in this method. This predictor estimates the geometric distortions for the quality assessment of 3D-synthesized views.
- **BIQI [46]:** The BIQI metric works in two steps and is based on Natural-Scene-Statistics(NSS). It computes the quality scores and the probabilities of the occurrence of five types of distortions in an image, namely, JPEG, JP2K, Fast fading (FF), Gaussian Blur (Blur), and white noise (WN).
- **NIQSV [1]:** NIQSV (No-reference-Image-Quality-assessment-of-Synthesized-Views). In the NIQSV metric, authors predict the quality based on the edges of the objects. This metric is based only on morphological operations such as opening and closing. Then this edge detected image is used as a quality score.
- **MNSS [47]:** multiscale-natural-scene-statistical-analysis(MNSS) is a combination of two new natural-scene-statistics(NSS) models oriented for 3D views. First is the variations in the degree of self-similarity from natural images at different scales. Second is the statistical regularity-based features. These features decide the final quality score.
- **NR-MWT [48]:** No-Reference-Morphological-Wavelet-with-Threshold (NR-MWT) is based on the fact that there is an increase in high-frequency content in 3D views. The selected areas with high wavelet sub-band are quantified for this purpose, followed by a threshold.
- **GDIC [49]:** Geometric-Distortions-and-Image-Complexity(GDIC) decomposes the views into into wavelet sub-bands using discrete wavelet transform. Then edge similarity is computed between the low-frequency sub-band and high-frequency sub-bands. Then, an auto-regressive filter is combined with a bi-lateral filter to compute the image complexity. Lastly, the quality score is



computed by normalizing geometric distortions.

- **CSC-NRM [50]:** Convolutional-Sparse-Coding(CSC) is based on computing a sparse representation with the sum of a set of convolutions for learning. These dictionary filters are used to create a codebook for learning.
- **SIQA-CFP [51]:** Synthesized-Image-Quality-Assessment-with-Contextual-Multi-Level-Feature-Pooling(SIQA-CFP). It is based upon a contextual multilevel feature pooling module. This module can encode the low- and high-level features. A deep pre-trained ResNet extracts these features.

We concluded the following points from the extensive literature survey conducted in this chapter:

1. Even contemporary 3D view synthesis algorithms produce distortion in the vicinity of object boundaries.
2. Existing No Reference (NR)/Full Reference(FR) IQAs focus on detecting black-holes and perform satisfactorily on the IRCCyN dataset while failing to detect the new types of distortions present in the recently proposed IETR dataset.
3. Existing algorithms are not incorporating the fundamental properties of the 3D synthesized views, such as shifting and context information.

Considering all these points, we proposed three different 3D IQA in this thesis, which are the next three chapters.

- Chapter 2: No Reference 3D IQA (Stretching Artifacts Identification for No Reference IQA of 3D Synthesized Images)
- Chapter 3: Full Reference 3D IQA 1 (Perceptually Unimportant Information Reduction and Cosine Similarity based Full Reference IQA for 3D Images)
- Chapter 4: Full Reference 3D IQA 2 (Context Region Identification based Quality Assessment of 3D Synthesized Views)

# Chapter 2

## No Reference 3D IQA

### Stretching Artifacts Identification for No Reference IQA of 3D Synthesized Images

From the extensive literature survey conducted above, we observed that very few of the existing 3D-synthesized IQA algorithms consider stretching artifacts as noticeable distortions for 3D-synthesized views. The evaluation dataset used in those works was IRCCyN/IVC Dataset [18], where the dominant distortion is black-holes. However, black holes are always filled in 3D-synthesized views, so they are no longer ‘holes.’ Consequently, evaluating how these areas are filled is crucial. black-holes are considered obsolete in the newly proposed IETR dataset [2]. Moreover, stretching artifacts mainly occur in the newly proposed 3D synthesis algorithms. Hence, there is a drastic difference in the performance of these algorithms on these two datasets.

#### 2.1 Motivation

In the literature, two existing methods i.e., NISQV+ [1] and Yue [3] try to identify the stretching artifacts, and their performance is limited due to the violation of some assumptions (described below).

In the NR IQA method for 3D-synthesized Views (NIQSV+) [1], Tian *et al.* detected stretching using Average Horizontal and Vertical Gradients (AHG/AVG) of each column/row of the view. Authors have proposed the use of the Sobel operator for detecting the AHG/AVG of a synthesized view as:

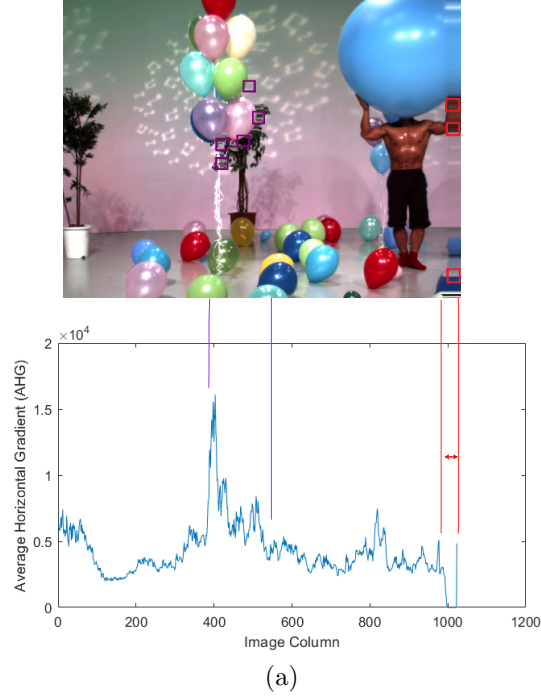


Figure 2.1: Stretching artifact identification using [1]. The plot represents each column’s average Horizontal Gradient (AHG) in the given view from IETR Dataset [2]. Purple and red colors indicate stretching artifacts and corresponding AHG in the middle and corner of the view, respectively. A red double arrow indicates Stretching Width (SW) as defined in [1]

$$AHG = \frac{\vec{\nabla}_h}{H} = \frac{I * G_h}{H}, AVG = \frac{\vec{\nabla}_v}{H} = \frac{I * G_v}{H} \quad (2.1)$$

where,  $I$  is the  $Y$  component (from the  $YC_bC_r$  color space) of the synthesized view,  $\vec{\nabla}_v$ ,  $\vec{\nabla}_h$ , are the vertical and horizontal gradients obtained using  $G_v$ ,  $G_h$ , Sobel vertical and horizontal gradient operators.  $H$  is the height of the synthesized view. NIQSV+ assumes that stretching occurs only on the right or left part of the view or in a completely horizontal or vertical direction. Therefore, they calculated the gradient of each row or column, and if for a particular row or column gradient magnitude is low, it suggests that this column or row has the stretching distortions. Finally, they proposed calculating the width of the stretching artifacts to estimate the quality score. Unfortunately, the assumption made in NIQSV+ is not true in the IETR dataset, as stretching artifacts generally occur near the occluding objects with no guarantee that stretching will happen in a complete row or column. To justify these arguments, Figure 2.1 shows the gradient magnitude in the horizontal direction for a view (from the IETR dataset) with stretching artifacts<sup>1</sup>. From this figure, it is clear that the NIQSV+ algorithm cannot detect stretching occurring near

<sup>1</sup>This plot is generated using the official code provided by the authors of [1].

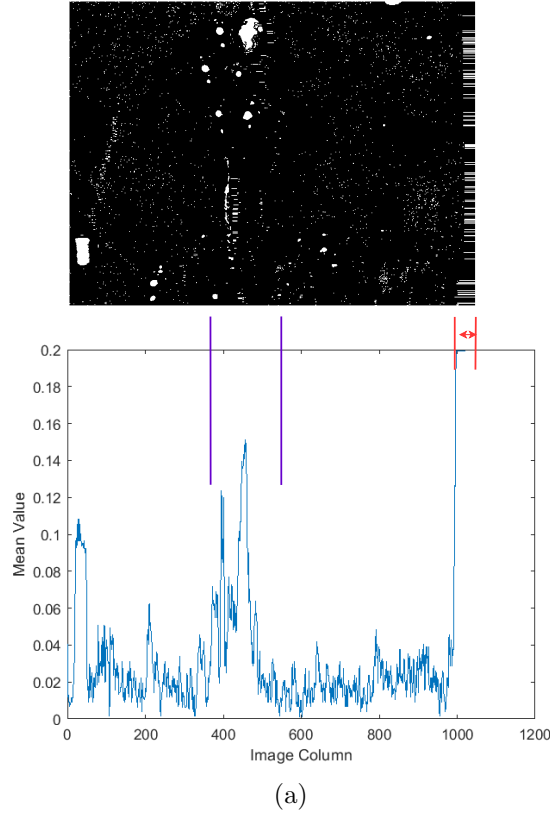


Figure 2.2: Coarse stretching region map and its column-wise mean values of the image in Figure 1, as proposed in [3].

objects which are not in the entire row and column. When the performance of NIQSV+ is analyzed on the IETR dataset, it performed poorly [2].

In [3], Yue *et al.* proposed to evaluate the stretching strength in two steps. First, the authors identify the regions with stretching artifacts by drawing a coarse stretching region map of a 3D-synthesized view. For this map, authors have used a uniform rotation invariant Local Binary Pattern (LBP),  $LBP_K^{ri}$ , which can be expressed as,

$$LBP_K^{ri} = \begin{cases} \sum_{i=0}^{K-1} s(I(n_i), I(n_c)), & \text{if } \Delta_K \leq 2 \\ K + 1, & \text{otherwise} \end{cases} \quad (2.2)$$

where  $I(n_i)$ ,  $i = 0, 1, 2, \dots, K - 1$  denotes the values of surrounding  $K$  symmetric pixels,  $I(n_c)$  is the value of center pixel. The relationship between two pixels  $a, b$  is calculated using  $s(a, b)$  as,

$$s(a, b) = \begin{cases} 1, & \text{if } a \geq b \\ 0, & \text{if } a < b \end{cases} \quad (2.3)$$

Also,  $\Delta_K$  is the number of bit-wise transitions defined as,

$$\Delta_K = \|s(I(n_0), I(n_c)) - s(I(n_{K-1}), I(n_c))\| + \sum_{i=1}^{K-1} \|s(I(n_i), I(n_c)) - s(I(n_{i-1}), I(n_c))\| \quad (2.4)$$

An example of such a coarse stretching map can be seen in Figure 2.2<sup>2</sup>. The white region in the coarse map primarily indicates the stretching region. Further, the stretching region is detected by calculating the average values and the map columns, followed by thresholding. The authors have empirically proposed to set the threshold value  $T$  to be 0.2. However, let's carefully analyze the mean values of each column in Figure 2. The stretching regions in the middle (approximately column 400 to 600) of the 3D view are not identified through mean values. Moreover, this algorithm calculates the stretching strength for only the right or left side of the view.

To address this gap, we propose a lightweight patch-based CNN model to detect the blocks with stretching artifacts in a 3D-synthesized view. The proposed model detects the stretching artifacts around the occluded region even if they do not occur in the entire row or column.

The contributions of the proposed work are listed below:

1. Demonstrate the effect of the stretching artifacts on the perceived quality and propose an efficient method of detecting the blocks with stretching artifacts.
2. The proposed model can efficiently predict the quality of 3D-synthesized views, and the performance of the proposed algorithm is better than the existing NR IQA algorithms. In addition, the proposed algorithm can predict the stretching artifact's location, which may further be useful to enhance the perceptual quality of 3D-synthesized views.

## 2.2 Proposed Quality Assessment Metric

This section introduces a new lightweight CNN model to identify the blocks in the view with stretching artifacts and further use the count of identified blocks for predicting the

---

<sup>2</sup>The map and the plot is inspired from [3] whose codes are not open-sourced. Hence we have implemented this idea of identifying stretching ourselves.

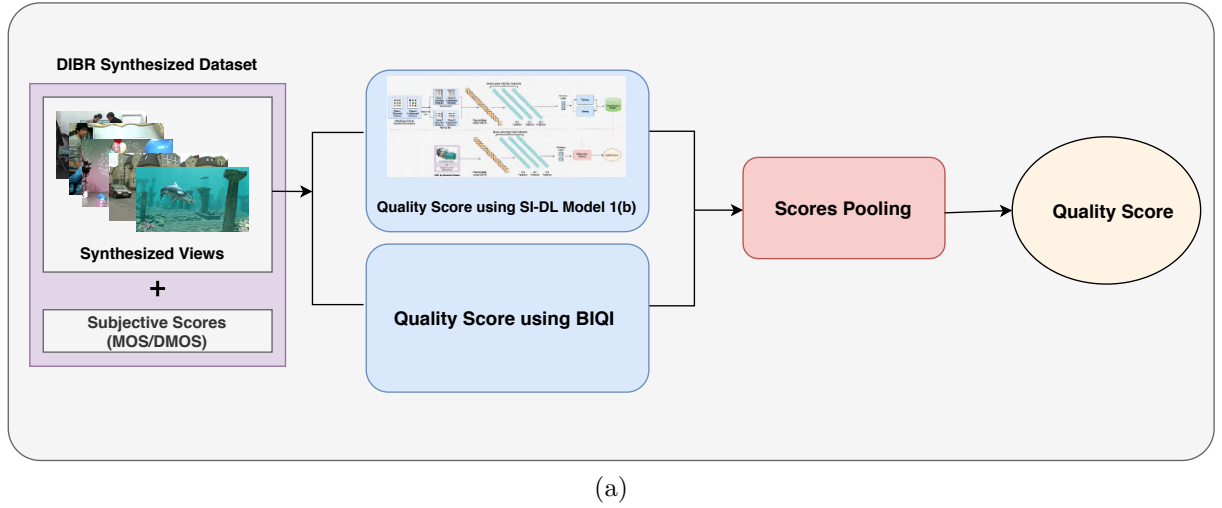


Figure 2.3: The workflow of the proposed method.

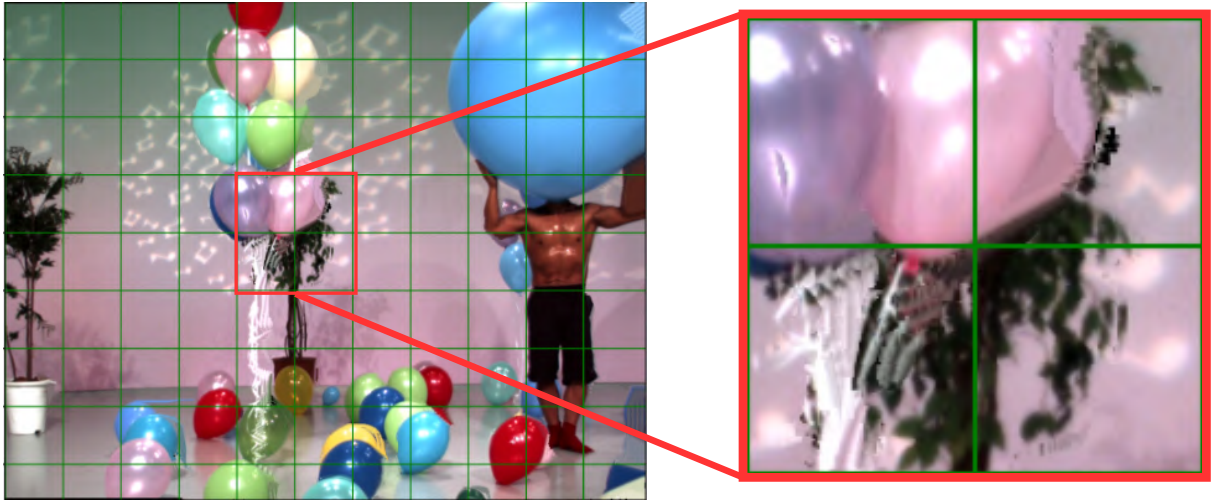


Figure 2.4: A view from IETR Dataset divided into blocks, the red window shows blocks with stretching artifacts.

perceptual quality score. Further, to seize other distortions such as blurring and blocking, this score is fused with the quality score obtained using the Blind Image Quality Index (BIQI) [46], an existing NR IQA metric. We propose to integrate these scores using the technique of pooling. The main workflow of the proposed method is outlined in Figure 2.3.

### 2.2.1 Stretching Identification using Deep Learning (SI-DL) Model

A stretching artifact is an annoying distortion due to improper inpainting while rendering 3D synthesized views. 3D synthesized views are generally contaminated with stretching, flickering, ghosting, and crumbling artifacts [1], and researchers have proposed different mechanisms to independently identify these artifacts. In this work, cohesively, we call

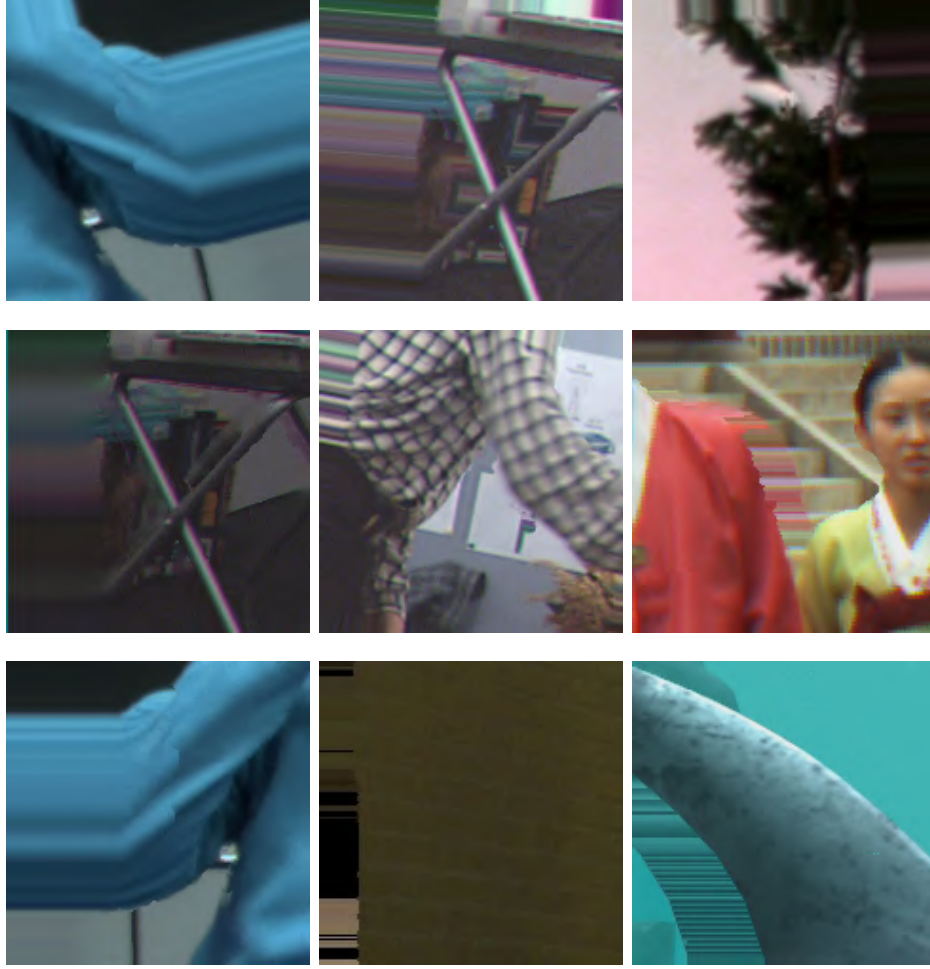


Figure 2.5: Class-1: Examples of blocks (dimensions:  $160 \times 160$ ) with Stretching Artifacts.

them stretching artifacts and propose to have a single model to identify these.

Initially, we observed in the IETR dataset [2] that there is a direct relationship between the stretching artifacts and overall subjective scores. However, the available data in IETR dataset is not enough to identify the exact magnitude of stretching artifacts. We analyzed and divided the images into blocks (of size  $160 \times 160$ ). Visually, we classified them into two categories, blocks with stretching artifacts and without stretching artifacts. We asked five expert subjects to participate in this experiment for these visual classifications. Next, we calculated the correlation coefficient between the number of blocks with stretching artifacts and their corresponding subjective scores. The calculated correlation coefficient was 0.55. This empirical study motivated us to objectively identify stretching using an efficient CNN model.

The steps involved in the proposed Stretching Identification using Deep Learning (SI-DL) Model are as follows:



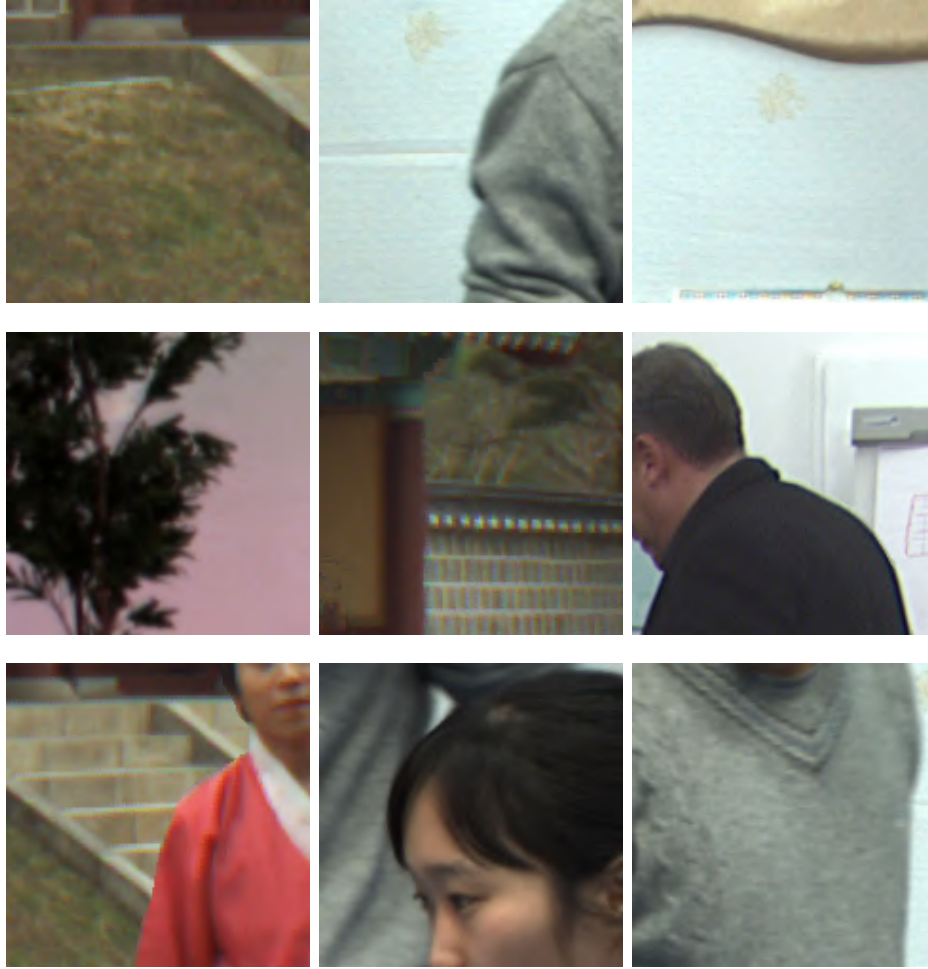


Figure 2.6: Class-2: Examples of blocks (dimensions:  $160 \times 160$ ) without Stretching Artifacts.

### 2.2.1.1 Data Collection

The IETR dataset has 140 views which are insufficient to train a CNN model that can generalize well. To overcome this issue, we propose to collect the views with stretching artifacts from the publicly available 3D datasets, namely: IRCCyN/IVC Dataset [?], MCL-3D Dataset [58], VRTS Dataset [59]. To avoid overfitting our CNN Model, we did not include blocks from the IETR dataset. We collected 214 related views from these datasets and further divided them into blocks of size  $160 \times 160$ . The effect of changing the size of the blocks on the overall performance is shown in Table 2.5. These blocks were then categorized into two classes, Class 1: Blocks with stretching artifacts, and Class 2: Blocks without stretching artifacts. Five subjects with expertise in the visual perception domain classified the blocks into two classes using majority voting. An example of highlighted blocks with stretching artifacts is shown in Figure 2.4. The detail of data collection is shown in Table 2.1. Further, Figures 2.5 and 2.6 show examples of the blocks in Classes



Table 2.1: Data collection description in detail

Dataset	Total Views	Selected Views	Number of blocks
IRCCyN/IVC [18]	96	26	532
MCL-3D [58]	693	154	3152
VRTS [59]	100	34	695

Table 2.2: Train-Validation splits.

Class	Class-1: Blocks with SA	Class-2: Blocks without SA
Total Blocks	2157	2216
Training Blocks (80 %)	1725	1773
Validation Blocks (20 %)	432	443

1 and 2, respectively.

### 2.2.1.2 CNN Architecture

The prepared dataset contains two categories: blocks with stretching artifacts and remaining blocks without stretching artifacts. To further increase the size of the dataset, we used the popular techniques of data augmentation, and transfer learning [60] in the proposed model. Random rotations, random zoom, width shifts, height shifts, and random horizontal flips are used for data augmentation. The created dataset had class imbalance, as in 3D synthesized views, the blocks with distortions are less frequent than those without distortions. Hence, the data augmentation techniques are only applied to the class with distorted blocks to prevent the CNN model’s class imbalance while creating the dataset. In total, 4379 blocks were included in the dataset. The distribution of these blocks into two classes and further into the training and validation set is shown through the following Table 2.2:

(i). Transfer Learning:

Weights of Deep convolutional networks trained on high-level image classification tasks work surprisingly well for numerous other tasks such as image super-resolution [61] and image synthesis [62]. In [23], Zhang *et al.* analyzed these deep features and concluded that these features effectively correspond to human perceptual judgments, as well. Similarly, instead of training the CNN model from scratch for our classification task, we utilized feature maps using the VGG-16 [63] deep learning model trained on Imagenet dataset [64].

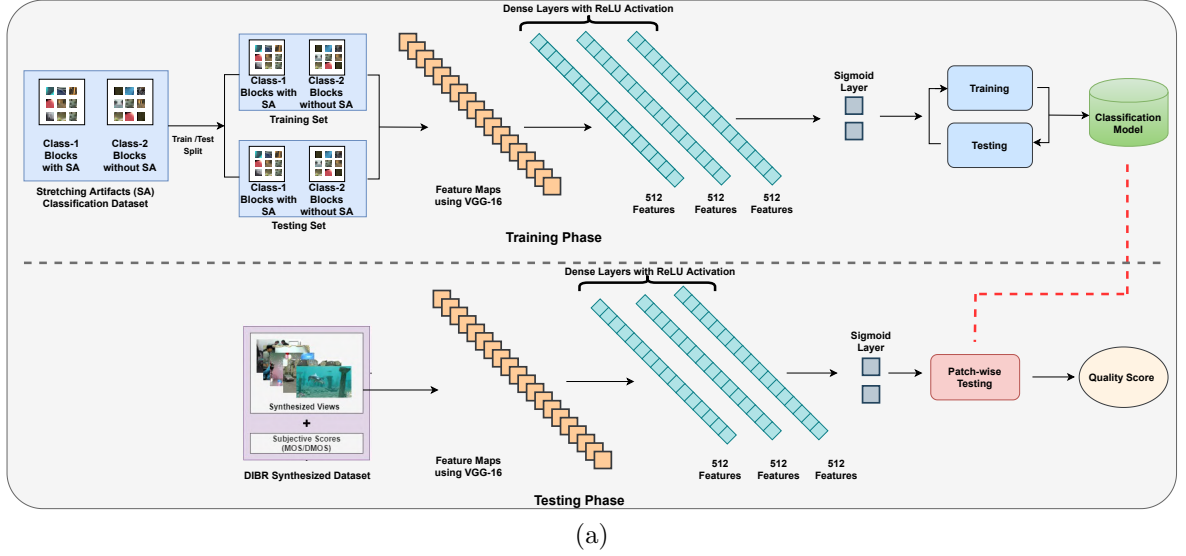


Figure 2.7: An elaborated workflow of the proposed Stretching Identification using Deep Learning (SI-DL) Model.

We also experimented with other DL architectures, such as Resnet [65], Inception V3 [66], and concluded that similar to the LPIPS metric [23], VGG-16 features work best for our classification task. The weights of the layers of VGG-16 are then fine-tuned for the proposed classification dataset. From our analysis, we have observed that freezing the middle two blocks of VGG-16 for feature extraction gives the most relevant features for classification. The extracted features are subjected to three Dense layers with 512 features each. Also, every layer is followed by the Rectified Linear Units (ReLU) activation function to add non-linearity and dropout of 0.3 factor to avoid over-fitting. For binary classification purposes, sigmoid activation is adopted in the last layer. The complete description of the proposed CNN architecture is tabulated in Table 2.3.

(ii). Training Process:

During the training of the proposed model, Binary Cross-Entropy Loss and Stochastic Gradient Descent (SGD) are used as a loss function and optimization algorithm, respectively. All the loss functions work satisfactorily for this task. Further, we analyzed the performance of four loss functions, Binary Cross-Entropy Loss, Poisson Loss, Squared Hinge Loss, and Hinge Loss, for the proposed CNN architecture. Table 2.4 shows the model performance on the validation data using these loss functions. The Binary Cross-Entropy Loss gives the best performance, so we used the model with this loss function as the final prediction model. Let  $c_1$  and  $c_2$  be the two classes, i.e., blocks with stretching artifacts and blocks without stretching artifacts, respectively. Let  $g_1$  and  $g_2$  are the

Table 2.3: The proposed VGG-16 fine-tuned architecture. FC stands for Fully Connected Layers. Conv stands for Convolution Kernel.

	Layer	Description	Output Shape	Trainable?	Activation
VGG-16 Layers	Input	Rescale Image	$160 \times 160 \times 3$	True	ReLU
	Block 1	$2 \times \text{Conv} - 64$	$112 \times 112 \times 64$	True	ReLU
	Block 2	$2 \times \text{Conv} - 128$	$56 \times 56 \times 128$	False	ReLU
	Block 3	$2 \times \text{Conv} - 256$	$28 \times 28 \times 256$	False	ReLU
	Block 4	$2 \times \text{Conv} - 512$	$14 \times 14 \times 512$	True	ReLU
	Block 5	$2 \times \text{Conv} - 512$	$7 \times 7 \times 512$	True	ReLU
Additional Layers	FC-1	dense-512	512	True	ReLU
	Dropout-1	dropout-0.3	-	-	-
	FC-2	dense-512	512	True	ReLU
	Dropout-2	dropout-0.3	-	-	-
	FC-3	dense-512	512	True	ReLU
	Dropout-3	dropout-0.3	-	-	-
	FC-4	dense-2	2	True	Sigmoid

ground-truth encoded value of each class  $c_1$  and  $c_2$ , respectively. Let  $s_1$  and  $s_2$  be the scores predicted from the proposed model. The cross-entropy loss function  $\mathcal{L}_{CE}$  can be finally formulated as,

$$\mathcal{L}_{CE} = -g_1 \log(s(s_1)) - (1 - g_2) \log(1 - s(s_2)) \quad (2.5)$$

where  $s(x)$  is the sigmoid activation function calculated as  $s(x) = 1/(1 + e^{-x})$ . Further, adopted learning-rate value for SGD optimization is  $10^{-5}$ . Let  $P_{class}(i)$  be the predicted class of the  $i^{th}$  patch of the view; it is given by,

$$P_{class}(i) = \begin{cases} 0, & \text{if class } \epsilon c_1 \\ 1, & \text{if class } \epsilon c_2 \end{cases} \quad (2.6)$$

The complete architecture of the proposed SI-DL model with training process visualization has been shown in Figure 2.7. The training was done on NVIDIA Tesla K80 in mini-batches of size 30 for 50 epochs. Figure 2.8 shows the training and validation accuracy at each epoch. In addition, we use four metrics, accuracy, precision, recall, and F1 score, to show the effectiveness of the proposed trained model, tabulated in Table 2.4. All the values are above 94% using binary cross-entropy loss, which further validates the

Table 2.4: Performance metric values on validation data with varying loss functions

Loss Function	Accuracy	Precision	Recall	F1-Score
Binary Cross-Entropy Loss	0.9474	0.9494	0.9474	0.9473
Poisson Loss	0.9399	0.9400	0.9399	0.9399
Squared Hinge Loss	0.9294	0.9363	0.9294	0.9289
Hinge Loss	0.9159	0.9160	0.9159	0.9159

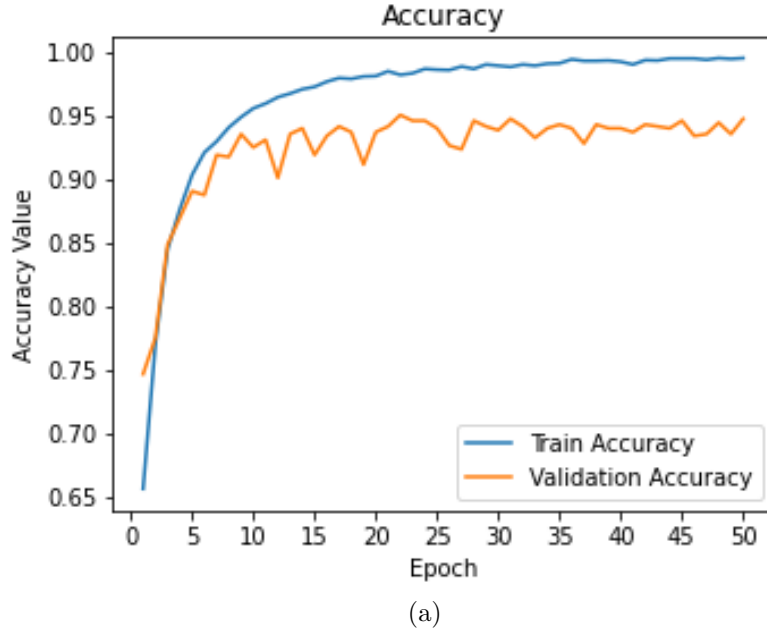


Figure 2.8: Plot of variation in training and validation accuracy with every epoch.

performance of the proposed model.

### 2.2.1.3 Stretching Artifact based Quality Score

A similar CNN architecture as SI-DL is used for testing purposes, where testing is done on each block of the 3D synthesized view from the dataset being tested. Each block is then tested for class type using the proposed SI-DL CNN Model. The final quality score  $Q_{SA}$  of the 3D-view is proposed by calculating the average number of blocks with stretching artifacts and is given by,

$$Q_{SA} = \frac{\sum_{i=1}^m [P_{class}(i) = 0]}{m} \times 100 \quad (2.7)$$

where  $m$  is the total number of blocks in the 3D-synthesized view. It is worth mentioning that for the training part, none of the images from the IETR dataset were used. The classification accuracy for the IETR dataset is 82% (testing accuracy), and the correlation

coefficient between the number of blocks with stretching artifacts and the subjective score of the IETR dataset is 0.4327, which is much better than most of the algorithms. These empirical results suggest that the proposed block classification model for stretching identification is fairly accurate.

### 2.2.2 Quality prediction using BIQI Metric

The proposed CNN-based quality prediction can only handle geometric distortions (such as stretching artifacts, flickering, and crumbling). It can not handle structural distortions (such as blurring and blocking). These structural distortions occur because of the improper rendering of the 3D-synthesized views. In the literature, there are many NR IQA metrics proposed for the identification of such artifacts in natural images such as BIQI [46], NIQE [67], and BRISQUE [45]. In [68], it was analyzed that the BIQI metric [46] can efficiently identify several kinds of distortions in 3D synthesized views. Based on this observation, we have employed the BIQI metric to identify structural distortions in the proposed algorithm. We also analyzed the effect of using NIQE and BRISQUE on the performance of the proposed algorithm in Table 2.4, which concludes the superiority of BIQI for this task. The BIQI metric works in two steps and is based on Natural Scene Statistics (NSS). It computes the quality scores  $q_j$  and the probabilities  $p_j$  of the occurrence of five types of distortions in an image, namely, JPEG, JP2K, Fast fading (FF), Gaussian Blur (Blur), and white noise (WN). The final quality score  $Q_{BIQI}$  is calculated as,

$$Q_{BIQI} = \sum_{j=1}^n p_j q_j \quad (2.8)$$

where,  $j = 1 - 5$  are the five types of distortions.

First, the image is decomposed using the Daubechies 9/7 wavelet transform over three scales and orientations. These decompositions are then subjected to a Generalized Gaussian Distribution (GGD). The vector obtained from mean, standard deviation, and shape parameters over all scales and orientations acts as the feature vector for the images. Finally, Support Vector Regression (SVR) based testing is employed for the given images on a pre-trained model provided with BIQI software release to get the final quality score,  $Q_{BIQI}$ . A detailed methodology used in the BIQI metric can be found in [46].

### 2.2.3 Scores Pooling

As discussed in the previous subsection, the quality of each 3D synthesized view is predicted using the proposed SI-DL and the BIQI algorithm. We aim to optimally pool the capabilities of these two algorithms to obtain the final score. As the number of blocks with stretching artifacts increases in the SI-DL algorithm, the perceptual quality degrades. Hence, there is an inversely proportional relationship between the subjective and predicted scores by (7). From (8), it is evident that the relationship between the scores given by the BIQI algorithm and the subjective score is directly proportional. Thus, it is imperative to systematically merge these scores since they are on different scales and have diverse relationships.

As suggested in the DSCB algorithm [68], the perceptual characteristics of natural views and synthetic views<sup>3</sup> are slightly different [69], it is beneficial to identify whether the particular view belongs to a natural or a synthetic view type. Subsequently, to effectively merge these scores, we introduce a pooling solution. The final quality score,  $Q_f$  of the proposed no-reference 3D-IQA is obtained as follows:

$$Q_f = \begin{cases} \frac{Q_{BIQI}^x + \epsilon}{Q_{SA}^y + \epsilon}, & \text{if } I_z \in I_N \\ \frac{Q_{BIQI}^u + \epsilon}{Q_{SA}^v + \epsilon}, & \text{if } I_z \in I_S \end{cases} \quad (2.9)$$

where  $I_z$  is the  $z^{th}$  view of the dataset.  $I_N$  and  $I_S$  are the set of Natural and Synthetic Views, respectively. These sets are obtained using the framework proposed in the recently proposed DSCB [68] algorithm. There is an inverse proportion of subjective scores with stretching artifacts. Hence the  $Q_{SA}$  is in the denominator of this equation.  $x, y, u, v$  are the positive non-zero constants that balance the variations in scales and diversity of the different obtained scores.  $\epsilon$  is a constant with small values used to avoid dividing by zero conditions.

---

<sup>3</sup>Note that here “natural view” is a 3D image synthesized from a reference image, which is a natural image, whereas “synthetic view” is 3D image synthesized from a reference image which is a synthetic image (a computer-generated image).

## 2.3 Experimental Results and Analysis

### 2.3.1 Evaluation Protocols

#### 2.3.1.1 3D Synthesized Views Dataset

The performance of the proposed metrics is evaluated by employing them on the publicly available dataset IETR DIBR [2].

**IETR DIBR Dataset:** This dataset comprises 140 synthesized views generated using ten reference views and corresponding subjective scores. Out of these ten reference views, 7 are Natural Views, and 3 are Synthetic Views. The views are rendered using 7 different DIBR methods M1: Criminisi’s [11], M2: LDI [12], M3: Ahn [13], M4: Luo’s [14], M5: HHF [15], M6: VSRS [16], M7: Zhu [17]. Among these 7 DIBR methods (M1 to M7), M7: Zhu is an inter-view 3D synthesis method, while M6: VSRS can be used as both inter-view and single-view 3D synthesis; all the other methods (M1 to M5) belong to the single-view 3D synthesis category. A brief overview of these methods is listed below:

i). Single-view 3D synthesis Methods:

- M1 (Criminisi’s [11]): This method is based on an exemplar-based texture synthesis technique. Patch priorities are computed using *confidence* parameter to improve the order of the pixel filling.
- M2 (LDI [12]): This algorithm uses an object-based Layered Depth Image (LDI) representation to obtain the synthesized view. Based on this representation’s foreground and background segmentation, the authors have proposed to render the 3D-synthesized view.
- M3 (Ahn [13]): A depth-based 3D synthesis method was proposed by Ahn *et al.* using patch-based texture synthesis.
- M4 (Luo’s [14]): This method is proposed based on background reconstruction. A random walker segmentation technique was employed using a detected initial seed.
- M5 (HHF [15]): Sohl *et al.* proposed two approaches, Hierarchical Hole-Filling (HHF) and Depth Adaptive Hierarchical Hole-Filling for filling dis-occluded regions.

ii). Inter-view 3D synthesis Methods:

- M6 (VSRS [16]): The MPEG 3D video Group has adopted View Synthesis Reference Software (VSRS) as a standard. A post filter is applied on depths to solve depth discontinuities. Then the holes are filled using inpainting.
- M7 (Zhu [17]): In this algorithm, Zhu *et al.* proposed to identify the background pixels and unoccluded background around the holes. Finally, these holes are filled using depth-enhanced Criminisi's method and simple block-average filling method.

It may also be noted that warping without further rendering is not used in this dataset, considering this method obsolete. Hence, there are no views with black-holes artifacts in this dataset, which makes it different from the existing DIBR datasets (IRCCyN/IVC Dataset [18], MCL-3D Dataset [58], VRTS Dataset [59]).

### 2.3.1.2 Evaluation Criteria

We followed the standard criteria for evaluating the correlation between the objective scores obtained using different metrics and the given subjective scores in the IETR dataset. The four criteria used are SROCC, PLCC, RMSE, and KRCC. A better IQA metric attains larger values of PLCC, SROCC, and KRCC and a lower value of RMSE. The scores given by different algorithms are mapped to subjective scores via the following non-linear equation as follows:

$$g(X) = x_1 \left( \frac{1}{2} - \frac{1}{1 + e^{x_2(X-x_3)}} \right) + x_4 X + x_5 \quad (2.10)$$

where  $X$  and  $g(X)$  are the objective and corresponding subjective scores, respectively.  $x_1, x_2, x_3, x_4, x_5$  are the five parameters to be fitted non-linearly.

To comprehensively judge the performance of the proposed algorithm, we compare the proposed algorithm with a total of 16 state-of-the-art Image Quality Assessment (IQA) metrics, including 10 No-Reference (NR) and 6 Full-Reference (FR) metrics. These algorithms are designed for two types of images, i.e., DIBR Synthesized Views, and Natural Images. A comprehensive description of these algorithms is given below:

i). No-Reference (NR) Metrics:

- GANs-NRM: GANs-NRM [52] was proposed by Suiyi *et al.* in which bag of words features was extracted for quality evaluation from a Generated Adversarial Network (GAN) rendered synthetic dataset.



- DSCB: In this method, Sadbhawna *et al.* [68], have assessed the quality of 3D-synthesized views. The distortions are incorporated with the properties of the human visual system (HVS) to generate the final quality score. Authors have also proposed to evaluate natural and synthetic views separately since their properties differ.
- BIQI: BIQI is proposed by Moorthy *et al.* [46] for quality prediction of natural images. This method is based on natural scene statistics (NSS). It is a two-step framework. Firstly the image is subjected to a Daubechies 9/7 wavelet transform and is parametrized using Gaussian distribution. Second, support vector regression (SVR) predicts the scores.
- Wang's: Wang *et al.* [70] measured geometric distortion and global sharpness by decomposing the image into wavelet subbands and further comparing the high-frequencies and low-frequencies. Image complexities are also computed in this quality assessment method of 3D-synthesized views.
- APT: Gu *et al.* proposes this method [57] for 3D-synthesized images. It is based on getting the local image description using autoregression (AR) and following this prediction with thresholding.
- Jakhetiya's: Jakhetiya *et al.* [44] used Kernel Ridge Regression (KRR) as a global predictor in this method. This predictor estimates the geometric distortions for the quality assessment of 3D-synthesized views.
- OMIQA: Jakhetiya *et al.* [71] proposed OMIQA for 3D-synthesized views using non-linear median filtering for predicting outliers to detect the geometric and structural distortions.
- NIQSV+: Tian *et al.* [1] proposed a NR IQA metric for 3D-synthesized views to identify distortions such as blurry regions, stretching, and black-holes, typically related to 3D views.
- Yue's: Combining Local and Global Measures (CLGM) [3] accounts for two types of distortions, i.e., sharpness and geometric. To detect stretching (a typical geometric distortion), the authors proposed to estimate the similarity between a region and

its equal-size adjacent region. The disoccluded regions are predicted by using local similarity. Finally, these scores are pooled together linearly.

ii). Full-Reference (FR) Metrics:

- SSPD: Mahmoudpour *et al.* [35] proposed to quantify the local differences using the feature matching technique. further, the gradient difference in image superpixels is measured to quantify the global loss.
- LOGS: It stands for LOfcal-Geometric-distortions-in-disoccluded-regions-and-global-Sharpness [36]. SIFT flow-based warping is first used to detect the disoccluded regions. Then the global sharpness is quantified using a re-blurring-based strategy.
- Tian's: Tian *et al.* [37] observed the statistical features of the reference image and its corresponding 3D-synthesized image using wavelet sub-bands to detect the artifacts in the views.
- LPIPS: Learned-Perceptual-Image-Patch-Similarity(LPIPS) [23] metric is based on deep features trained on the ImageNet dataset. We used their trained model on VGG features for comparison.
- MP-PSNR: Sandi'c-Stankovi'c *et al.* [38] proposed to decompose the views into multi-scale pyramids using morphological pyramids for quality prediction.
- MW-PSNR: Sandi'c-Stankovi'c *et al.* [27] used morphological filters to maintain low-level features such as edges over multiple levels. These levels are obtained using wavelet decomposition.

### 2.3.2 Parameters Sensitivity Analysis

As shown in equation (9), the proposed algorithm uses predominantly four parameters ( $x$ ,  $y$ ,  $u$ , and  $v$ ), and these parameters control the contribution of the proposed SI-DL algorithm and the BIQI algorithm. The other parameter  $\epsilon$  is a small non-zero positive value used to avoid division by zero. We chose the value of  $\epsilon = 1$  for our experiment. The following observations are made to choose the value of these four parameters.

1. From Table 2.8, it is visible that the proposed algorithm (SI-DL) does not perform as well as the BIQI algorithm for the natural images; consequently,  $x > y$ .

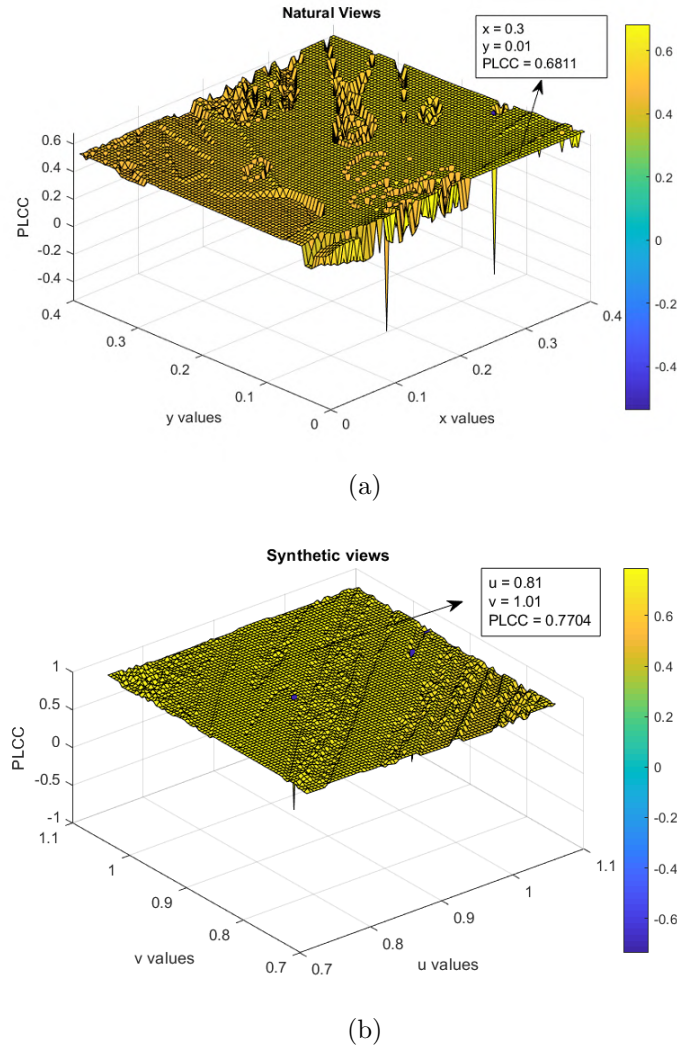


Figure 2.9: Parameter Sensitivity. (a) and (b) shows the performance of the proposed algorithm with varied parameters  $x$ ,  $y$ , and  $u$ ,  $v$  for Natural Views and Synthetic Views, respectively.

2. The BIQI algorithm is designed for the quality prediction of natural images, and it is expected that it cannot perform well for synthetic images. While the proposed SI-DL algorithm is performing much better than the BIQI algorithm and overall pooling, the contribution of the SI-DL algorithm should be higher than the BIQI algorithm. Subsequently, parameter  $v$  should have a higher value than the  $u$ .

Based upon the above arguments and extensive empirical analysis, we have chosen the value of these parameters ( $x$ ,  $y$ ,  $u$ , and  $v$ ) to be 0.31, 0.01, 0.81, and 1.01, respectively. In Figure 2.9, the dependency of the proposed algorithm on these parameters has been shown. From these figures, it is visible that slightly varying these parameters does not significantly affect the proposed method's performance.

In the proposed SI-DL algorithm, the blocks are identified with stretching artifacts

Table 2.5: Effect of varying block sizes on performance metrics using proposed pooling.

Block-sizes	IETR Dataset			
	PLCC	SROCC	KRCC	RMSE
$128 \times 128$	0.6457	0.5598	0.3727	0.1532
$160 \times 160$	0.7087	0.6672	0.4726	0.1749
$192 \times 192$	0.5936	0.5414	0.3731	0.1429

Table 2.6: Performance comparison after pooling with the existing NR-IQA metrics.

Metric	IETR Dataset		
	PLCC	SROCC	RMSE
Proposed( $Q_{SA}$ ) with BIQI [46]	0.7087	0.6672	0.1749
Proposed( $Q_{SA}$ ) with BRISQUE [45]	0.5331	0.4153	0.2098
Proposed( $Q_{SA}$ ) with NIQE [67]	0.5038	0.3792	0.2142

to assess the perceptual quality of 3D images. The blocks cannot be too big, as bigger blocks can have more than one object, and only some parts can have stretching artifacts. On the other side, smaller blocks do not represent the image properties. With this view, we have chosen  $160 \times 160$  blocks to train the deep-learning model. The performance of the proposed algorithm by varying the block size is given in Table 2.5, and it is visible that the proposed algorithm performs best when the block size is  $160 \times 160$ .

The proposed algorithm uses BIQI [46] algorithm for the quality prediction of structural distortions. To further validate the superiority of BIQI over other popular NR IQA metrics such as BRISQUE [45] and NIQE [67], we analyzed the performance of pooling these metrics in the proposed algorithm. Table 2.4 contains this detailed comparison, which also validates the superiority of the BIQI metric in terms of performance parameters.

### 2.3.3 Performance Comparison and Analysis

A detailed comparison of the proposed metric with the existing NR and FR IQA metrics is shown in Table 2.7. The proposed method is significantly superior to all the compared existing methods. The proposed metric obtains the PLCC, SROCC, KRCC, RMSE values of 0.7087, 0.6672, 0.4726 and 0.1749 respectively. The proposed method performs almost equal to SSPD metric [35] but can be considered superior to the SSPD metric because

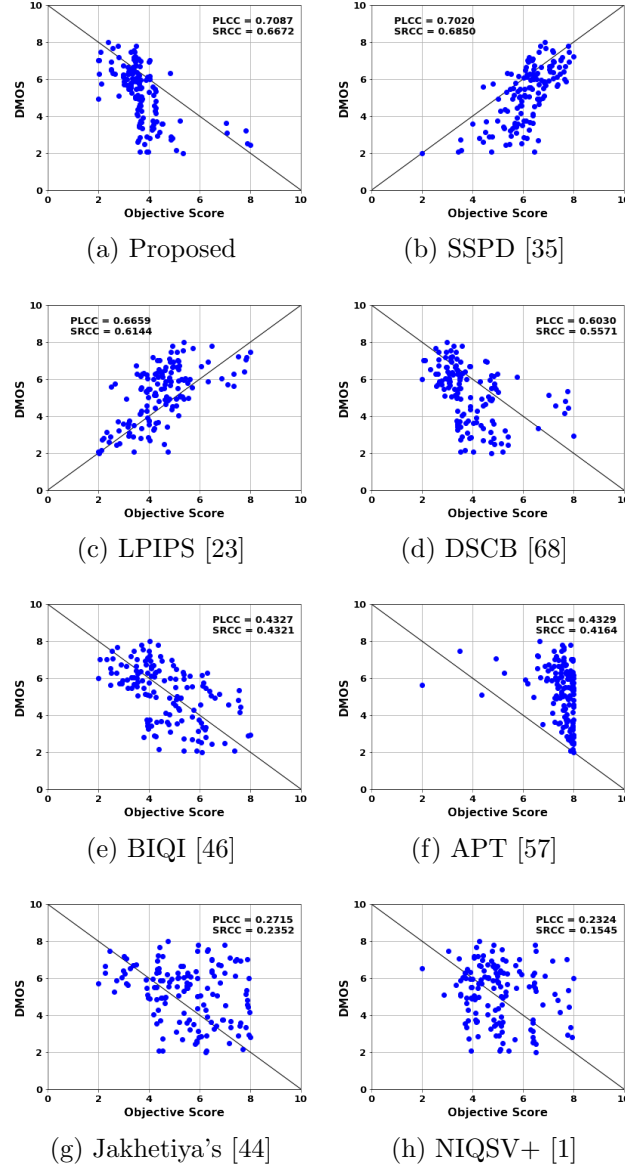


Figure 2.10: Scatter Plot of DMOS values and objective scores of state-of-the-art IQAs.

SSPD is a FR 3D IQA metric while the proposed is an NR 3D IQA metric. Additionally, the SSPD metric takes approximately 28 seconds to predict the quality score of one view, which is very large compared to the proposed metric (see Table 2.13). Further, the proposed method achieves 9.7%, and 16.84% gain in PLCC and SROCC from GANs-NRM [52] metric, which is an NR 3D IQA metric. The proposed method achieves a 2.99% and 6.56% increase in PLCC and SROCC compared to the recently proposed Yan's metric [42]. Yan's metric is a training-based algorithm, trained using random forest regression on the IETR dataset. So, similar to Yan's metric, we also applied random forest regression for mapping scores (scores using the SI-DL model and BIQI) to subjective scores, and the proposed method achieves above 10% better performance (in terms of

PLCC and SROCC) than Yan’s metric. It is worth noting that the proposed method with and without training on the IETR dataset performs significantly better than Yan’s. The proposed method outperforms all the other NR IQA methods, whether designed for DIBR Views or Natural Images.

As discussed in [68], the human visual system is more sensitive to natural than synthetic images. To show the effectiveness of the proposed algorithm for both the natural and synthetic images, separate results for both types of images are shown in Table 2.8. As the table shows, the proposed method obtained 0.6811, 0.6045, 0.4197, and 0.1709 values of PLCC, SROCC, KRCC, and RMSE, respectively, for the 98 natural views IETR datasets. The proposed method also achieved 0.7704, 0.7273, 0.5470, and 0.1766 values of PLCC, SROCC, KRCC, and RMSE, respectively, for 42 synthetic views from the IETR dataset. Table 2.8 shows that the proposed algorithm performs well for both types of images.

The proposed algorithm has several stages: stretching artifact identification, BIQI algorithm, classification, and pooling. The stage-wise performance is shown in Table 2.9. The inclusion of each stage significantly enhances the performance of the proposed algorithm.

The proposed quality metric is designed to identify stretching artifacts and is not focused on identifying black holes. Nevertheless, we tested our method on the IRCCyN dataset (whose most dominant distortion is black holes) [18] and IVY dataset [19]. The detailed analysis can be seen in Table 2.10. It may also be noted that while performing this experiment, we have taken care of generality by not including the patches from the dataset which is being tested. Moreover, the parameters settings and pooling method are different for different datasets. The performance of the proposed algorithm is slightly poorer than the QA algorithms specifically designed for the 3D synthesized views with black holes (IRCCyN dataset). We have also compared the proposed algorithm with the existing algorithms on the IVY dataset [19], and the proposed algorithm is performing better than the existing algorithms except IDEA [22]. At the same time, the performance of the proposed algorithm is much better than the IDEA on the IETR and IRCCyN datasets.

Further, we have evaluated the individual performance of the proposed algorithm for the 7 DIBR synthesis algorithms (M1-M7) used in the IETR dataset, and the results

are shown in Table 2.11. It can be inferred from this table that apart from the whole IETR dataset, the proposed SI-DL and overall algorithm works satisfactorily for M1-M7 individually also. Although, the performance of SI-DL is limited for some algorithms such as M1, M2, and M5. One possible reason is that SI-DL is specially designed for stretching artifact identification, and these algorithms do not significantly produce such distortions. Nevertheless, the proposed overall pooling method overcomes this limit of SI-DL, as can also be seen in Table 2.11.

We have also adopted scatter plots for intuitively comparing quality assessment methods. The scatter plot between the predicted objective score and the subjective scores given in the IETR dataset is shown in Figure 10. We compared the plots of seven relevant and recently developed different methods such as SSPD [35], LPIPS [23], DSCB [68], BIQI [46], APT [57], Jakhetiya's [44], NIQSV+ [1]. It can be depicted from Figure 2.9 that the predicted scores from the proposed method converge better as compared to other quality metrics.

### 2.3.4 Statistical Significance Analysis

Besides numerical comparisons using PLCC, SROCC, KRCC, and RMSE parameters, statistical significance analysis is another widely adopted comparison method. For this purpose, F-Test is done between the objective scores obtained using the proposed method and the scores obtained using the state-of-the-art quality assessment methods. The F-Test is based upon the variance-based hypothesis [72], where the score of the F-test,  $F_{score}$  is given by,

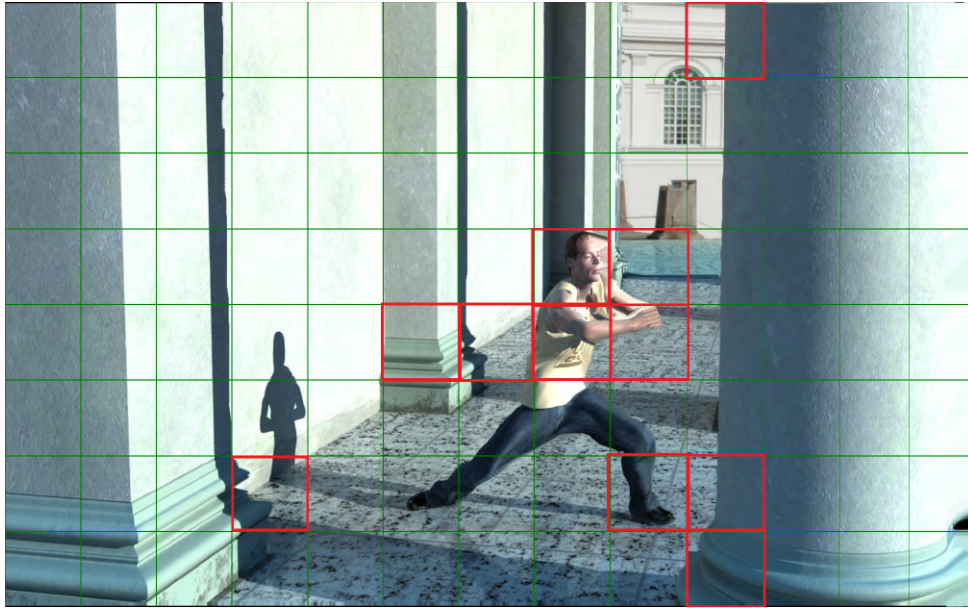
$$F_{score} = \frac{\sigma_{m1}^2}{\sigma_{m2}^2} \quad (2.11)$$

here,  $\sigma_{m1}, \sigma_{m2}$  are the RMSE values of the two metrics ( $m1, m2$ ) being tested. All the tested IQA metrics with their F-test values are listed in Table 2.12. '+1' indicates that m1 is statistically superior to m2, '0' indicates that the two are equally competitive, whereas '-1' indicates the statistical inferiority of m1 to m2. Table 2.12 shows that the proposed method is statistically superior to all the compared no-reference IQAs.

### 2.3.5 Time Complexity Comparison

The computational complexity of the proposed algorithm is in line with the requirements needed to address real-time applications. To empirically validate this, we calculated the time for predicting the perceptual quality of a 3D-synthesized image of resolution  $1024 \times 768$  from the IETR dataset, and the results are reported in Table 2.13. The proposed algorithm takes approximately 0.646 seconds to predict the perceptual quality of the 3D synthesized image on a system with configuration, Intel(R) Core(TM) i7-8700 CPU, 16 GB RAM, and NVIDIA GeForce GT730 Graphics. Further, we can interpret from the table that although the time taken by the proposed algorithm is higher than NIQSV+ [1], OMIQA [71], DSCB [68] algorithms. At the same time, all these algorithms perform poorly in the experimentation compared to the proposed algorithm.

## 2.4 Application in the enhancement of 3D views



(a)

Figure 2.11: Location identification of blocks with stretching artifacts using proposed SI-DL model.

The proposed algorithm efficiently predicts the quality of 3D-synthesized views. Along with predicting the quality, our SI-DL algorithm can accurately predict the location of blocks with stretching artifacts. Figure 2.11 illustrates the performance of the proposed algorithm on a view from the IETR dataset (for detecting the blocks with stretching artifacts). In this figure, blocks with red boundary show that it has stretching artifacts.



After detecting these blocks, any of the enhancement algorithms (such as [44], [56], [55]) can be applied to remove stretching artifacts. Henceforth, the proposed algorithm can assess the perceptual quality of 3D synthesized views and help enhance their perceptual quality.

## 2.5 Conclusions and Future Work

In this paper, we proposed a novel algorithm for no-reference quality assessment of 3D synthesized views, which is a challenging problem due to the presence of cohesive stretching artifacts. Several algorithms were proposed to assess the quality of 3D-synthesized views in the literature. Still, they perform poorly when employed for distortions in more recent datasets due to their inability to identify the stretching artifacts efficiently. We observe a relationship between the perceptual quality of 3D synthesized images and the number of blocks with stretching artifacts and propose to estimate these blocks via a CNN-based architecture. In contrast, none of the images from the IETR dataset are used while training. Our approach outperforms the existing methods in terms of correlation between the subjective and objective scores. One of the limitations of our approach is its inability to detect the level of stretching distortions in the identified block, which we plan to address in our future work.

Table 2.7: Performance comparison of various algorithms (in terms of PLCC, SROCC, KRCC, and RMSE). The table is arranged in descending order of PLCC. The symbol "Δ" indicates the unavailability of source code or reference resources, and "◇" indicates the results are taken directly from the original research papers. "Δ" indicates that official source codes were available at the time of experimentation, and "†" indicates the results are taken from experiments of research papers.

	IQA	Designed for	IETR Dataset			
			PLCC	SROCC	KRCC	RMSE
NR IQAs	Proposed	DIBR-Synthesized Views	0.7087	0.6672	0.4726	0.1749
	Yan's◇ [42]	DIBR-Synthesized Views	0.6881	0.6261	0.4660	0.1750
	GANs-NRM◇ [52]	DIBR-Synthesized Views	0.6460	0.5710	-	0.1980
	DSCBΔ [68]	DIBR-Synthesized Views	0.6030	0.5571	0.3677	0.1978
	BIQIΔ [46]	Natural Images	0.4327	0.4321	0.2898	0.2223
	Wang's◇ [70]	DIBR-Synthesized Views	0.4338	0.4254	-	0.2244
	APTΔ [57]	DIBR-Synthesized Views	0.4329	0.4164	0.2830	0.2235
	JakhetiyaΔ [44]	DIBR-Synthesized Views	0.2715	0.2352	0.1607	0.2386
	OMIQAΔ [71]	DIBR-Synthesized Views	0.2705	0.2331	0.1593	0.2387
	NIQSV+Δ [1]	DIBR-Synthesized Views	0.2324	0.1545	0.1083	0.2411
	Yue's† [3]	DIBR-Synthesized Views	0.1146	0.08605	-	0.2463
	SSPDΔ [35]	DIBR-Synthesized Views	0.7020	0.6850	0.4952	0.1790
	LOGS◇ [36]	DIBR-Synthesized Views	0.6687	0.6683	-	0.1845
	Tian's◇ [37]	DIBR-Synthesized Views	0.6685	0.5903	-	0.1844
FR IQAs	LPIPSΔ [23]	Natural Images	0.6659	0.6144	0.4386	0.1850
	MP-PSNR† [38]	DIBR-Synthesized Views	0.6190	0.5809	0.3802	0.1947
	MW-PSNR† [27]	DIBR-Synthesized Views	0.5389	0.4875	0.3364	0.2088

Table 2.8: Performance metrics comparison individually for Natural and Synthetic View types for IETR Dataset.

Step	Natural Views				Synthetic Views			
	PLCC	SROCC	KRCC	RMSE	PLCC	SROCC	KRCC	RMSE
BIQI	0.5700	0.5347	0.3667	0.1908	0.2623	0.2894	0.1870	0.2672
SI-DL (Proposed)	0.4543	0.4099	0.2962	0.2079	0.5217	0.5019	0.3609	0.2363
Pooling (proposed)	0.6811	0.6045	0.4197	0.1709	0.7704	0.7273	0.5470	0.1766

Table 2.9: Stage-wise performance evaluation of the proposed algorithm.

Stage	IETR Dataset			
	PLCC	SROCC	KRCC	RMSE
BIQI	0.4327	0.4321	0.2898	0.2223
Stretching Identification (Proposed)	0.4307	0.3610	0.2530	0.2237
BIQI + Stretching Identification (Proposed)	0.6109	0.5652	0.3897	0.1963
BIQI + Stretching Identification + Classification (Proposed)	0.7087	0.6672	0.4726	0.1749

Table 2.10: Performance comparison of various 3D IQA algorithms for IRCCyN Dataset and IVY Dataset. ("◇", "△", "†", "- " indicates same meaning as in Table 2.7.)

	Metric	NR/FR	IRCCyN Dataset			IVY Dataset		
			PLCC	SROCC	RMSE	PLCC	SROCC	RMSE
	Proposed	NR	<b>0.7710</b>	0.6896	0.4240	0.5459	0.5396	11.9349
	NIQSV+ <sup>△</sup> [1]	NR	0.7114	0.6668	0.4679	0.2191	0.2990	24.0530
	APT <sup>△</sup> [57]	NR	0.7307	0.7157	0.4546	0.5240	0.4748	20.9961
	OMIQA <sup>△</sup> [71]	NR	0.7678	0.7036	0.4266	0.2637	0.1131	13.7566
	IDEA <sup>◇</sup> [22]	FR	0.6652	0.4986	0.3533	0.6311	0.6132	19.0379
	GANs-NRM <sup>◇</sup> [52]	NR	0.7940	0.7720	0.4100	-	-	-
	Wang's <sup>◇</sup> [70]	NR	0.7995	0.7869	0.4000	-	-	-
	Jakhetiya's <sup>△</sup> [44]	NR	0.8054	0.7598	0.3946	0.5211	0.2288	12.1467

Table 2.11: Performance of the proposed algorithm on different DIBR synthesis algorithms used in IETR dataset.

DIBR Algorithm	Proposed SI-DL				Overall			
	PLCC	SROCC	KRCC	RMSE	PLCC	SROCC	KRCC	RMSE
M1 (Criminisi's) [11]	0.1529	0.0577	0.0161	0.0906	0.5520	0.4451	0.2842	0.0765
M2 (LDI) [12]	0.1234	0.1605	0.1013	0.1059	0.4195	0.3857	0.2630	0.0854
M3 (Ahn) [13]	0.5080	0.3837	0.2995	0.1498	0.8003	0.7865	0.5895	0.1043
M4 (Luo's) [14]	0.3386	0.0535	0.0106	0.1209	0.3950	0.3759	0.2526	0.1181
M5 (HHF) [15]	0.0922	0.0552	0.0162	0.1020	0.6719	0.5083	0.3684	0.0759
M6 (VSRS) [16]	0.4325	0.2569	0.1816	0.1559	0.4742	0.4652	0.3149	0.1522
M7 (Zhu) [17]	0.6336	0.1581	0.0899	0.0109	0.5785	0.5033	0.3333	0.0528

Table 2.12: Statistical Significance (SS) Table for comparison between the proposed algorithm and existing state-of-the-art IQAs.

Metrics	SSPD [35]	LPIPS [23]	DSCB [68]	BIQI [46]	APT [57]	Jakhetiya [44]	NIQSV+ [1]
Proposed	0	0	+1	+1	+1	+1	+1

Table 2.13: Time taken (in seconds) by objective 3D IQA metrics.

Metric	Proposed	SSPD [35]	NIQSV+ [1]	DSCB [68]	APT [57]	Jakhetiya [44]	OMIQA [71]
Time (in seconds)	0.6457	28.868	0.26	0.575	101.66	4.285	0.041

# Chapter 3

## Full Reference 3D IQA 1

### Perceptually Unimportant Information Reduction and Cosine Similarity based Full Reference IQA for 3D Images

We also observed from the literature that there are very few optimal Full-Reference 3D IQAs in the literature. And as both NR and FR 3D IQA have their own advantages in multiple applications, we proposed a new full reference 3D IQA in this and the next chapter. One of the primary reasons behind these algorithms' sub-optimal performance is their inability to highlight the distortions accurately. To overcome this limitation, we propose a novel FR IQA algorithm to efficiently identify the distortions present in 3D-synthesized images and eventually predict their perceptual quality. The proposed algorithm mainly works in two parts. We anticipated the geometric distortions using the reference and synthesized image in the first part of our approach. Rendering 3D synthesized image shifts pixels in the image. Due to these shifts and geometric distortions, the different image contains minuscule information perceptually unimportant for the 3D synthesized image's quality assessment. We propose using the morphological operation (opening) for unimportant information reduction to calculate the quality score. In the second part of our approach, cosine similarity between pre-trained VGG-16 [63] features on the Laplacian pyramid of the original and synthesized image is utilized to predict the quality score. Then these scores are efficiently pooled to get the final quality score. The main contributions of the proposed algorithm are enumerated below:

1. We use a simple morphological operation (opening) for unimportant information reduction between original and distorted 3D synthesized images. The residual image

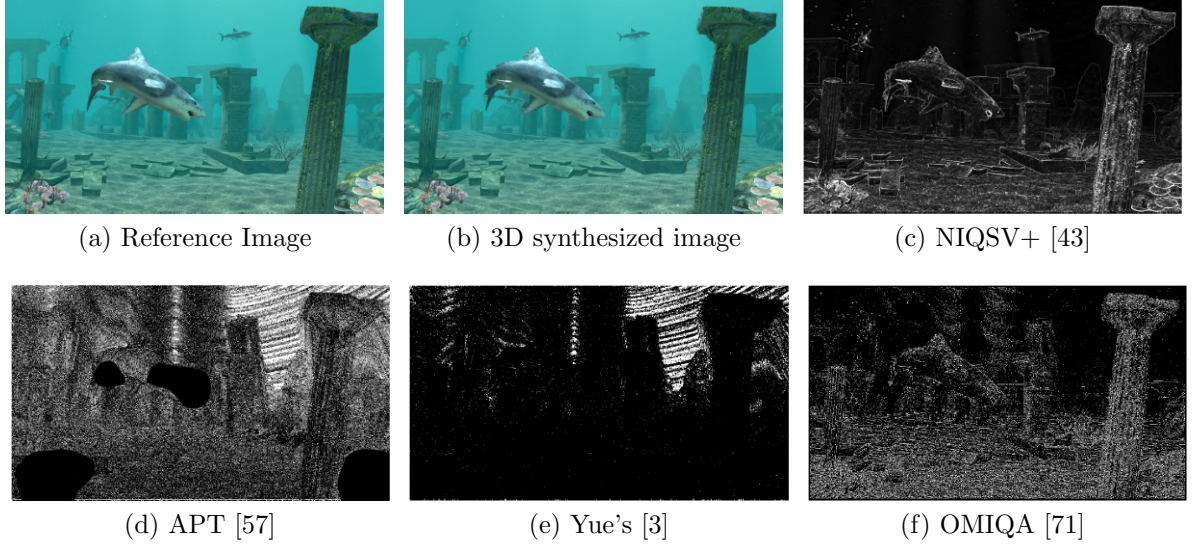


Figure 3.1: Predicted distortion maps using various algorithms. (a) and (b) are a reference and its corresponding 3D synthesized image from IETR Dataset. (c)-(f) are the distortion maps of the image predicted using four existing algorithms in the literature.

after a morphological operation can highlight the geometric distortions well. We propose to use this image to quantify the quality score of 3D synthesized images.

2. Geometric and structural distortions are quantified using the multi-level Laplacian pyramid-based deep features.
3. Finally, we compare these deep features of the reference and the distorted image, which are perceptually important via cosine similarity. Cosine similarity identifies the similarity of deep features within each other rather than giving importance to the magnitude of difference across them.

Algorithms proposed in prior works identify geometrical distortions, such as stretching, black-holes, blurring, flickering and crumbling, using different techniques such as super-pixel gradients [21], shift-compensation [73], Kernel Ridge Regression [74], Autoregressive Modelling [57], Median filtering [71], Horizontal/Vertical average gradient [43], Local Binary Pattern [3] etc. Most of these algorithms perform well on the IRCCyN dataset [18]. The algorithms work on two assumptions: a) black-hole artifacts exist in the 3D synthesized images; b) correlation between neighboring pixels in the geometrically distorted region is pretty low, and these distortions can be highlighted using any prediction algorithm. However, with evolution in 3D rendering, black holes have become obsolete while stretching artifacts are predominantly present in the rendered images, as observed in the

IETR dataset [2]. As existing algorithms cannot highlight stretching artifacts accurately, their performance is suboptimal on the IETR dataset, also shown in this survey by Tian *et al.* [75].

To illustrate the above arguments, Fig. 3.1(a) and 3.1(b) show a reference and a 3D-synthesized image, respectively, from the IETR dataset (in which the prominent distortion is around the boundary of the shark). Correspondingly, Figs. 3.1(c) – (f) show the identified artifacts using several existing algorithms (such as NIQSV+ [43], APT [57], Yue [3], and OMIQA [71])<sup>1</sup>. In this figure, a higher gray level suggests the location of geometric distortions, and a dark level suggests that geometric distortions are not present in this region. These distortion maps are pooled to estimate the perceptual quality of 3D synthesized images. It can be noticed from the figure that the existing algorithms are unable to identify geometric distortions accurately. There are two primary reasons for their sub-optimal performance.

1. Existing algorithms assume that distortions are abrupt and the correlation between neighboring pixels is low. On the contrary, with the growth of better inpainting algorithms, distortions in 3D synthesized images are smooth (as shown in Fig. 3.1(b)), and correlation among the neighboring pixels is high even in the distortion region.
2. NIQSV+ [43] and Yue’s [3] algorithms assume that stretching artifacts mainly arise at the left and right margin of the image and that they occur in the entire row or column. This assumption is invalid since the stretching artifacts can also arise near the objects due to the occlusion.

## 3.1 Proposed Algorithm

We propose an algorithm that can accurately identify the artifacts in 3D synthesized images and eventually judge their quality. First, we use a morphological operation (opening) in the residual image to efficiently identify the distortion in a 3D synthesized image. Then, to capture the other vital features affecting the quality of a 3D image, we use a pre-trained Convolutional Neural Network (CNN), i.e., VGG-16 [63] on the Laplacian pyramid. Using the cosine similarity, we calculate the quality score based on the similarity between the

---

<sup>1</sup>Note that all the maps are generated using the official codes released by the authors except for Yue’s which we implemented ourselves.

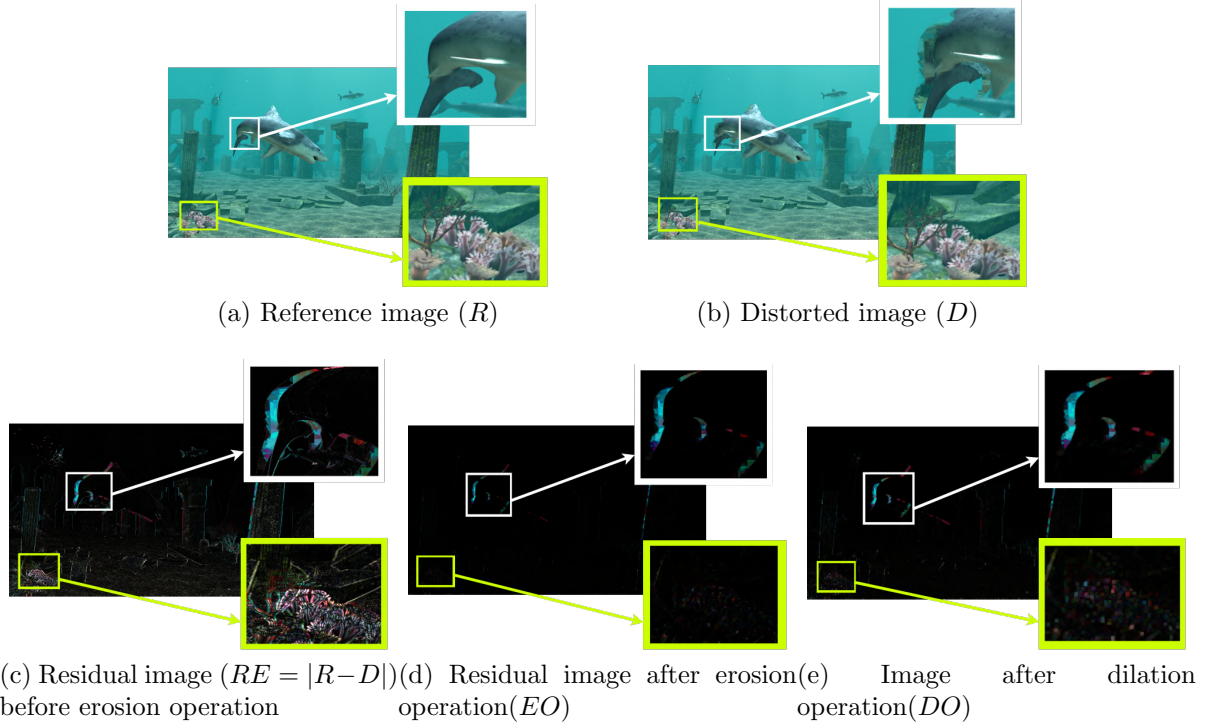


Figure 3.2: Elaborated Perceptually Unimportant Information Reduction (PU-IR) method with highlighted (perceptually important and unimportant) distortions.

feature maps of reference and the distorted image. Lastly, we pooled the scores calculated using both algorithms to get the final quality scores. In the following sub-sections, the detailed methodology is explained:

#### 3.1.0.1 Perceptually Unimportant Information Reduction (PU-IR)

3D synthesized images are generally contaminated with geometric distortions, while structural distortions are predominantly not present in these images [57]. Consequently, the difference between the reference and distorted image can provide substantial useful information related to perceptual quality. This argument can be validated by comparing the performance of PSNR on the IETR dataset [2] and the LIVE dataset [76]. With this view, we calculated the residual image ( $RE$ ) by taking the respective difference of Y, Cb, and Cr components of reference and distorted images as:

$$RE = |R - D| \quad (3.1)$$

Here,  $R$  and  $D$  are the references and the corresponding distorted image. Example reference and distorted image from the IETR dataset are shown in Fig. 3.2(a) and 2(b),



respectively. In Fig. 3.2(c), the residual image is shown, and it can be observed that the residual image (highlighted using a white window) can efficiently highlight the important geometric distortions. Simultaneously, the reference and distorted images are slightly shifted due to improper rendering. This shifted information is visible in the residual image (highlighted using a fluorescent yellow window). This information from the shift cannot be considered distortion and does not contribute to the overall perceptual quality. This is why the performance of PSNR is promising but not optimal for the quality assessment of 3D synthesized images. Removing the perceptually unimportant information from the shift between the reference and the distorted images is required. In the literature, this shift compensation has been done using the SURF key-point detection [21, 73], sparse representation [77], SIFT-flow based warping [24], etc. We propose a high-speed and efficient algorithm that removes unimportant information using classical image processing techniques in this work. Due to the slight shift in the reference and distorted image, only thin object boundaries are visible in the residual image (as shown in Fig. 3.2(c)). This perceptually unimportant information can be further suppressed using the morphological erosion operation ( $\ominus$ ) [74]. The erosion ( $\ominus$ ) of RE by S replaces the value of RE at a pixel (x, y) by the minima of the values of RE over a structuring element S. So, eroded image (EO) is obtained via:

$$EO = RE \ominus S = \min_{(i,j) \in S} \{RE(x+i, y+j)\} \quad (3.2)$$

We have used  $5 \times 5$  square window as the structuring element in this work. Fig. 3.2(d) shows that a simple erosion operation can remove the unimportant perceptual information arising from the shift between the reference and distorted image. With this view, erosion operation can effectively reduce unimportant information arising due to the shift. However, erosion operation shrinks the geometric artifacts slightly due to the  $5 \times 5$  structuring element. To identify the geometric distortion, we propose applying the dilation operation ( $\oplus$ ) further after the erosion operation (2). The dilation ( $\oplus$ ) of EO by S replaces the value of EO at a pixel (x, y) by the maxima of the values of EO over a structuring element S.

$$DO = EO \oplus S = \max_{(i,j) \in S} \{EO(x-i, y-j)\} \quad (3.3)$$

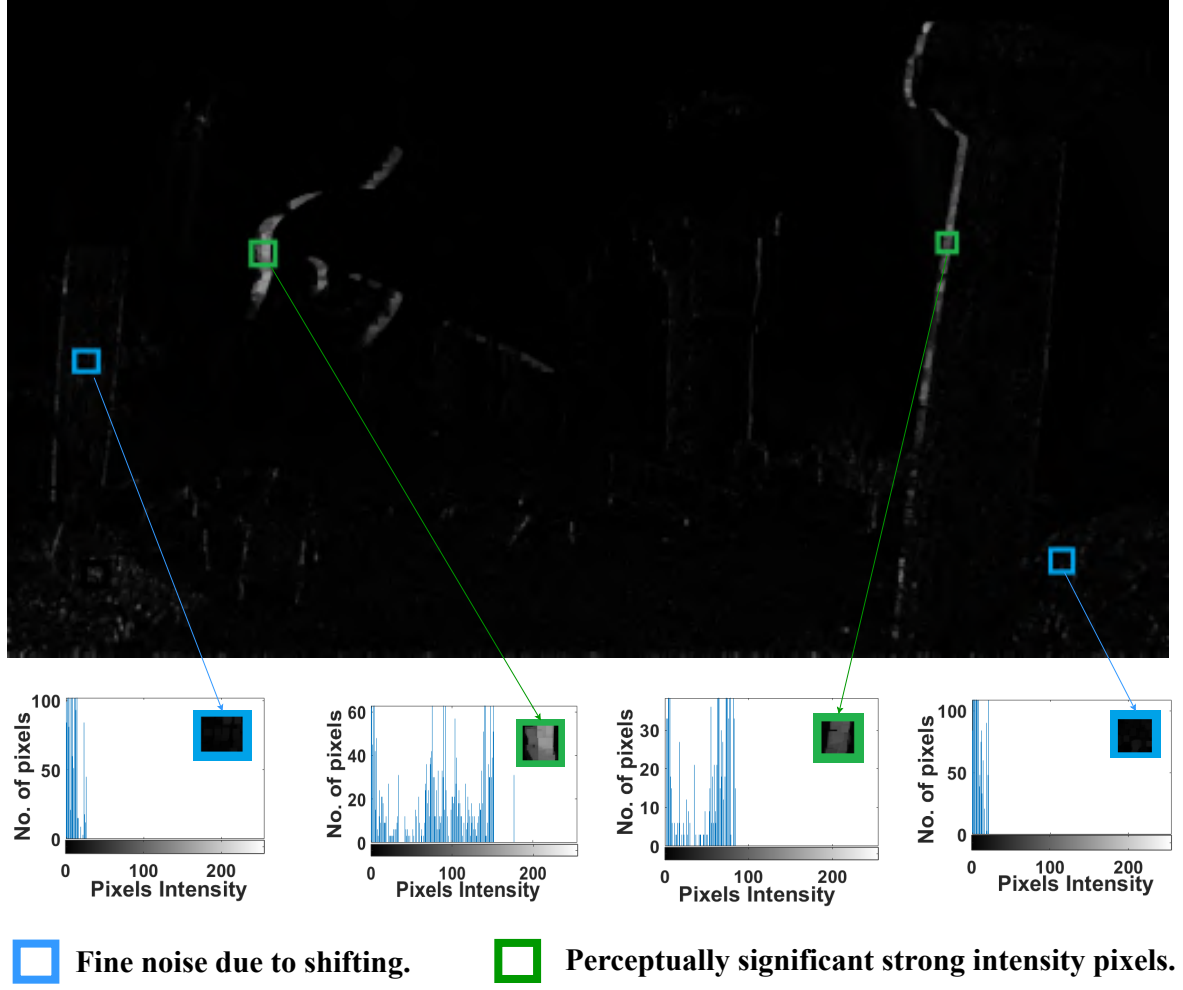


Figure 3.3: Perceptually Unimportant Information Reduction map ( $DO$ ) analysis using the histogram.

In Fig. 3.2(e), we have shown the residual image after the dilation operation. This figure shows that the proposed algorithm can highlight geometric distortions, and this information can be infused to predict perceptual quality. The final quality score is calculated using the mean of the reference and the residual image after the morphological operation ( $DO$ ). The quality score using the proposed PU-IR algorithm is calculated as:

$$Q_{PU-IR} = \log \frac{\text{mean}(DO)}{\text{mean}(R)} \quad (3.4)$$

Here  $\text{mean}$  represents the average value of the corresponding image. This quality score is scaled to avoid negative values of scores.

The residual image after morphological operation still contains small perceptually unimportant information (shown in Fig 3.2(e) highlighted using a fluorescent yellow window). This information should not be considered to estimate the objective quality of 3D-synthesized images. Hence, if the number of pixels in a residual image is more than

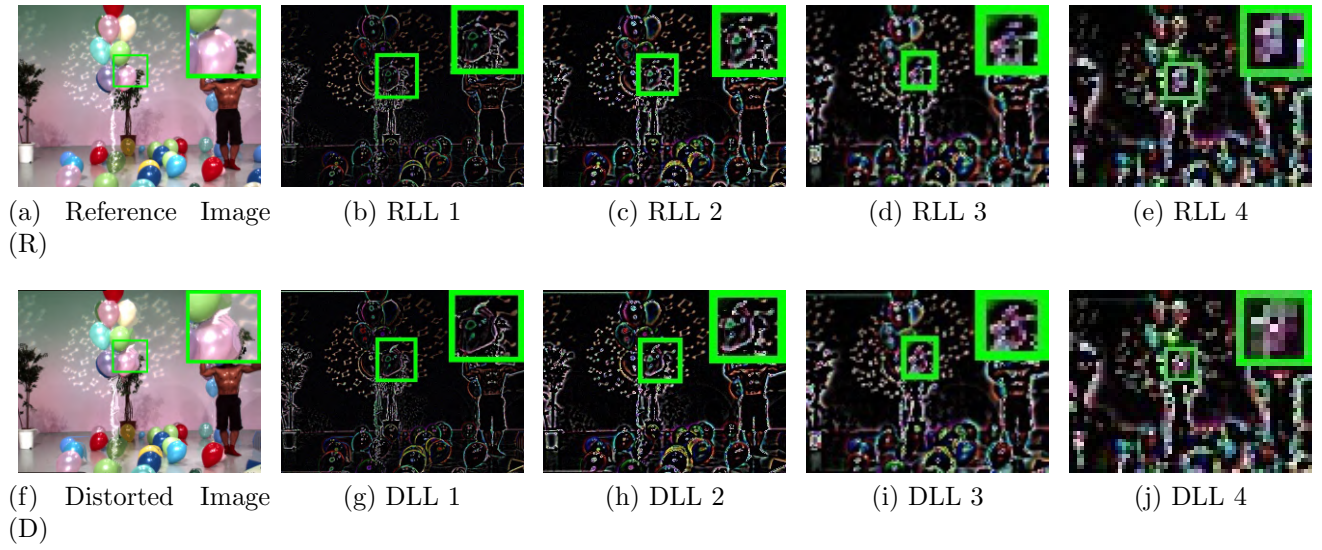


Figure 3.4: Laplacian Levels wise highlighted distortions in the reference and distorted images. RLL stands for Reference image Laplacian Level, and DLL stands for Distorted image Laplacian Level. Please note that RLL(2-4) and DLL(2-4) are of different resolutions, but these are shown in equal size for better visualization in this figure.

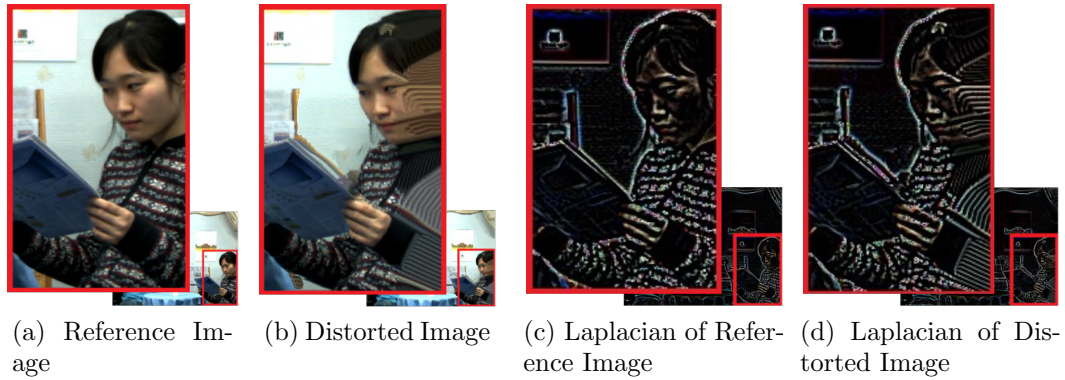


Figure 3.5: Effect of Laplacian on the distortions of an image from IETR Dataset.

a threshold, significant geometric distortions are present, and perceptual quality should be poor and vice-versa. With this view, the residual image after dilation operation (DO) can be further used to enhance the proposed PU-IR algorithm. From Fig. 3.3, it can be seen that the intensity range of pixels in the visible distortions (green window) is much greater than the range in fine noise (blue window), which is not perceptually significant. With this view, we propose to enhance the proposed PU-IR algorithm further using the number of pixels with strong geometric strengths as:

$$Q_{PU-IR} = \begin{cases} Q_{PU-IR'} \times \lambda_1, & \text{if } count > \frac{(m \times n)}{\gamma} \\ Q_{PU-IR'} \times \lambda_2, & \text{otherwise} \end{cases} \quad (3.5)$$

Here,  $\lambda_1$  and  $\lambda_2$  are positive non-zero constant parameters used to enhance the proposed PU-IR algorithm and estimate the perceptual quality of 3D synthesized images. The parameter ‘count’ is the number of pixels with strong intensity in the residual image after dilation operation (DO).  $m$  and  $n$  are the dimensions of the image in  $x$  and  $y$ -direction, respectively. The  $\gamma$  is a positive non-zero constant. The motivation behind (5) is to amplify the impact of information in the residual image if strong geometric distortions are present in the 3D synthesized views and vice-versa. With this view, the value of  $\lambda_1$  should be higher than  $\lambda_2$ , as strong geometric artifacts affect the perceptual quality more than the weaker ones. The sensitivity analysis of  $\lambda_1$ ,  $\lambda_2$ , and  $\gamma$  parameters is shown in the next section.

Although the proposed PU-RI algorithm is quite simple, it can remove unimportant information from the shift between the reference and distorted image. Even though the proposed algorithm can mainly incorporate geometric distortions, it cannot effectively identify structural artifacts such as blurring and noise contamination.

### 3.1.0.2 Deep Features extraction and comparison using Cosine Similarity (DF-CS)

In the proposed PU-IR algorithm, the quality of 3D synthesized images is calculated using the ratio of the mean of the residual image after dilation operation (DO) to the reference image. However, two significant drawbacks are associated with this algorithm, which need to be rectified.

1. In the PU-IR algorithm, erosion operation is used to remove the perceptually unimportant information in its process. Simultaneously, PU-IR removes small structural distortions such as noise and other artifacts.
2. Also, the PU-IR algorithm identifies the overall absolute geometric distortions but can not guarantee that these geometric distortions affect the perceptual quality.

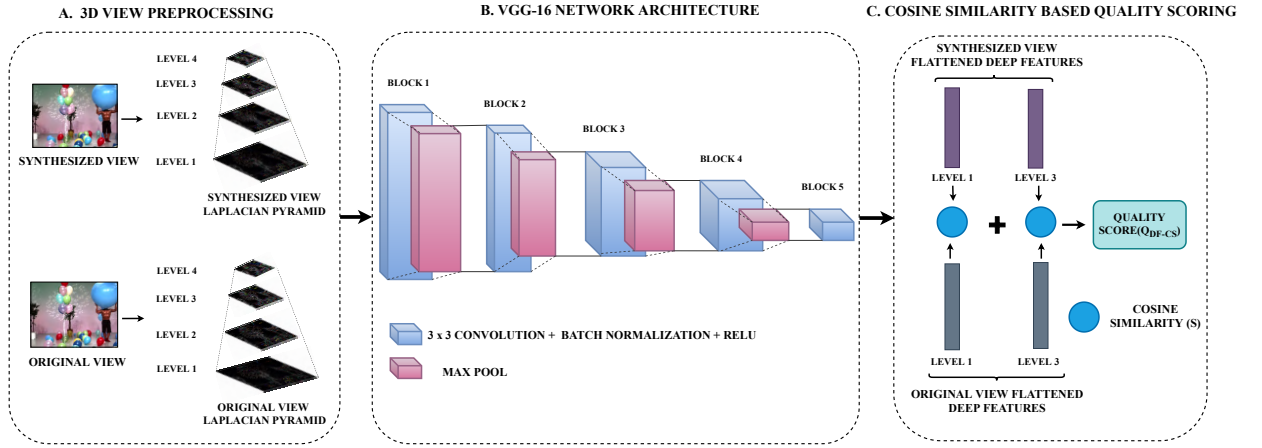


Figure 3.6: An elaborated workflow of the proposed Deep Features fusion using Cosine Similarity (DF-CS).

We propose a new algorithm based on the Laplacian pyramid and cosine similarity to overcome these issues. The Laplacian pyramid [78] is created using the Gaussian pyramid. In the Gaussian pyramid higher-level image (low resolution) is created from a lower-level image (high resolution) by blurring and down-sampling. To be more precise, the image at the next level of the Gaussian pyramid is generated by blurring using the Gaussian kernel and then downsampling the image from the current level by removing every even-numbered row and column. We first build a Gaussian pyramid of image  $I$  with  $L$  levels  $\{G_l\}_{l=1}^L$  ( $G_1 = I$ ), where  $G_l(x, y)$  is obtained as,

$$G_l(x, y) = (G_{l-1}(x, y) * f) \downarrow 2 \quad (3.6)$$

Here, “\*” is the convolution operation,  $\downarrow$  is the downsampling operation which is done by removing the alternative rows and columns, and  $f$  is the 2-D Gaussian kernel as:

$$f = \frac{1}{16} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix} \quad (3.7)$$

We follow this pattern as we go up the pyramid (i.e., resolution decreases). A level in the Laplacian pyramid is formed by the difference between the same level of the Gaussian pyramid and the interpolated version of its upper level in the Gaussian pyramid. Let  $L_l$  be the  $l_{th}$  level of the Laplacian pyramid formed using Gaussian levels  $(G_l)$  and  $(G_{l+1})$ ,

formulated as,

$$L_l = G_l - (G_{l+1} \uparrow 2) \quad (3.8)$$

Here  $\uparrow$  represents the interpolation operation, and interpolation is done using the bilinear interpolation algorithm. The first level of the Laplacian pyramid is the simple difference between the original image and the interpolated image from the above level in the Gaussian pyramid. The lower levels (such as levels 1 and 2 (See Fig. 3.4)) of the Laplacian pyramid are the difference between the color image and the moderately blurred versions of the color image. The synthesized view's level 1 and 2 contains structural and geometric distortions. Using these levels of the Laplacian pyramid can resolve the first issue associated with the proposed PU-IR algorithm. Simultaneously, levels 3 and 4 (See Fig. 3.4) of the Laplacian pyramid are images consisting of band-pass images and some low-frequency residuals. Any information available in these levels of synthesized view will be the substantial geometric distortions and overcome the second issue associated with the proposed PU-IR algorithm. Examples of cropped references and distorted images are shown in Fig. 3.5(a) and 3.5(b), respectively. Further, their corresponding Laplacian features at Level 1 are highlighted in Fig. 3.5(c) and 3.5(d), respectively, and from this figure, it can be seen that Laplacian can highlight the distortions.

Once the Laplacian pyramid is obtained, we move to the next feature extraction step. It is well known that VGG-16/19 architecture can extract the perceptually important features [23]. Further, its pre-trained architecture can be directly used to estimate the perceptual quality of natural images. Similarly, we also propose to extract features from the VGG-16 architecture. Notably, the input to the VGG-16 architecture is the specific levels of the Laplacian pyramid rather than color images. The weights of the pre-trained VGG network are fixed; it cannot detect the perceptually relevant features related to the 3D synthesized views (it can be seen via observing the performance of the LPIPS algorithm [23] on the IETR dataset). With this view, we propose to use the Laplacian pyramids of distorted and synthesized images to capture both structural and geometrical distortions and feed them into the network. The Laplacian pyramid works progressively at different scales towards the quality assessment of 3D synthesized images. The complete framework of the proposed DF-CS algorithm is shown in Fig 3.6.

CNN network trained on Imagenet dataset [64] for image classification surprisingly performs well for image representation [79,80]. This is the reason why feature extracted

from pre-trained CNN architectures have been used for style transfer [81], image super-resolution [82], image reconstruction [83]. Similarly, for efficient feature extraction, we have also used VGG-16 architecture [63]. Among different deep architectures in the literature, the deep features obtained from VGG-16 architecture are more relevant to the perceptual quality assessment of images. This argument is also justified by Zhang *et al.* in their detailed study of the power of the deep features as a quality metric in [23].

We propose to extract features from the fourth convolution layer of block five of VGG-16 for the distorted image and the original image for their different Laplacian levels. The Laplacian images are directly fed into the VGG with their original sizes without any initial pre-processing. Let  $F$  be the set of feature vectors (f) obtained from the VGG-16 architecture represented as

$$F = \{f_c^{or}, f_c^{sy}, f_{L_1}^{or}, f_{L_1}^{sy}, \dots, f_{L_n}^{or}, f_{L_n}^{sy}\} \quad (3.9)$$

Here *or*, *sy* stand for original and synthesized images, respectively. While, *c* stands for the RGB color components of the images and  $L_1, \dots, L_n$  are the  $n$  Laplacian levels.

Then these features are converted into a one-dimensional vector, i.e., the flattened vector. After obtaining the feature vector for the distorted and reference image, the next step is to combine these features to determine the final quality. Generally, the geometric distortions arise close to the high-frequency regions, and these distortions also occur in patches rather than evenly distributed in the whole image. With this view, the human visual system perceives a low level of geometrical distortions. So, it can be argued that most regions with geometric distortion have the same effect on the overall perceptual quality [57]. Therefore, we propose identifying the number of features that are similar to each other instead of calculating the magnitude of the difference between these features. On the contrary, the frequently used mean-square error (MSE) gives weightage to the more deviating features and ignores the less deviating features. For this, the cosine similarity has been recently used in several computer vision problems such as object tracking [84], image classification [85], palm-print recognition [86], binary search [87]. In these problems, the need is to identify the number of deep features that are similar. We have also used cosine similarity to fuse the deep features estimated on the Laplacian pyramid with this view. Henceforth, we propose to find the similarity between the deep features of different Laplacian pyramid levels of the synthesized image  $f^{sy}$  and the original image  $f^{or}$  as given

below:

$$\mathcal{S} = 1 - \cos \theta_{f^{or} f^{sy}} = 1 - \frac{f^{or} \cdot f^{sy}}{\|f^{or}\| \|f^{sy}\|} \quad (3.10)$$

where “ $\|\cdot\|$ ” and “ $\cdot$ ” represents the vector’s euclidean norm and the dot product between two vectors. “*Cosine Similarity*” is the term used for the cosine of the angle between the vectors  $f^{or}$  and  $f^{sy}$  and it is indicated by  $\cos \theta_{f^{or} f^{sy}}$ . The higher cosine similarity value indicates the closeness of vectors to each other, which further indicates better perceptual quality. Simultaneously, a lower cosine similarity value indicates that feature vectors differ and subsequently poorer the perceptual quality. Hence, cosine similarity is directly proportional to the quality score. To inverse this relationship, we subtracted the  $\cos \theta_{f^{or} f^{sy}}$  from 1.

Hence, we obtain the final set of Cosine Similarities ( $CS$ ) of different levels of the Laplacian pyramid as,

$$CS = \{\mathcal{S}_{L_1}, \dots, \mathcal{S}_{L_n}\} \quad (3.11)$$

After an exhaustive study, we propose to use only two levels of the Laplacian pyramid (first and third level) to evaluate the quality score of 3D-synthesized images. Including more levels does not significantly affect the proposed algorithm’s overall performance (as shown in Table 3.3). The quality score ( $Q_{DF-CS}$ ) using the proposed DF-CS algorithm is obtained as,

$$Q_{DF-CS} = \mathcal{S}_{L_1} + \mathcal{S}_{L_3} \quad (3.12)$$

### 3.1.0.3 Scores Pooling

The proposed PU-IR can identify the geometric distortions and quantify the perceptual quality based on these distortions. Further, the proposed DF-CS is based upon quantifying scores based on structural and geometric distortions. To merge the benefits of the proposed PU-IR and DF-CS, the final perceptual quality score is estimated by multiplying the quality scores obtained by the proposed DF-CS and PU-IR algorithm as:

$$Q = Q_{DF-CS} \times Q_{PU-IR} \quad (3.13)$$



Here,  $Q_{DF-CS}$  and  $Q_{PU-IR}$  are the perceptual quality scores estimated using the proposed DF-CS and PU-IR algorithms, respectively.

## 3.2 Experimental Results & Analysis

### 3.2.1 3D Synthesized images Dataset

We evaluated our model on the publicly available the IETR 3D Dataset [2] and IRCCyN Dataset [18]. The IETR dataset consists of a total of 140 synthesized images, along with ten reference images. These images are rendered using seven rendering methods, i.e. A1-A7 [88–94]. The subjective scores of each image are given in the form of Differential Mean Opinion Score (DMOS) in the dataset according to the Subjective Assessment Methodology for Video Quality (SAMVIQ) convention [95]. The rendering methods A1-A7 are explained in brief as:

- A1 [88]: It is an exemplar-based image inpainting method proposed by Criminisi *et al.*. The patch priorities are computed using a *confidence* to optimize the filling order further.
- A2 [89] is an object-based Layered Depth Image (LDI) representation. The authors have utilized the foreground-background segmentation approach to improve this method's inpainting.
- A3 [90]: In this method, texture synthesis is done using patches based upon the disocclusion filling method utilizing the depth map information.
- A4 [91]: Authors use a background reconstruction-based approach in this method. Morphological operations followed by canny edge detection are utilized for background reconstruction. Further, hole filling is done in this pre-processed image.
- A5 [92]: In this method, disocclusion of holes is done using depth adaptive Hierarchical Hole Filing (HHF).
- A6 [93]: The MPEG 3D video Group has adopted this method of Depth Image-Based Rendering (DIBR) by Tanimoto *et al.*, this software is also known as View Synthesis Reference Software (VSRS). A post-filter approach is used on the depth map to render the new scenes.

- A7 [94]: Zhu *et al.* identified the relevant pixels in the background around the holes to fill the occluded regions.

The four evaluation criteria used for performance comparison of the proposed algorithm with the existing IQA algorithms are SRCC, PLCC, KRCC, and RMSE. Higher PLCC, SRCC, KRCC, and lower RMSE values indicate a better IQA metric. The calculated objective scores are mapped to subjective scores using a five-parameter non-linear mapping as,

$$f(X) = \alpha_1 \left( \frac{1}{2} - \frac{1}{1 + e^{\alpha_2(X - \alpha_3)}} \right) + \alpha_4 X + \alpha_5 \quad (3.14)$$

where  $\alpha_i, i = 1, 2, 3, 4, 5$  are the five parameters to be fitted.  $X$  and  $f(X)$  is the objective score, and its corresponding mapped subjective score, respectively.

### 3.2.2 Performance Comparison and Analysis

Table 3.1 shows that the proposed algorithm's performance is compared to 24 existing IQA algorithms in the literature, including 8 FR IQAs, 13 NR IQA, and 3 SVR IQA. As can also be observed from the table, the proposed algorithm obtains the best overall performance. It is superior to all other competing IQA algorithms, whether oriented for 3D images or natural images. The proposed metric obtains 0.7965, 0.7909, 0.5992, and 0.1499 of PLCC, SRCC, KRCC, and RMSE, respectively. It exceeds by 9.7 % and 14.5 % in terms of PLCC and SRCC from state-of-the-art CODIF Metric [39], which is a no-reference 3D-IQA metric. Also, the proposed quality evaluation exceeds 13.4 % and 15.4 % in terms of PLCC and SRCC from SSPD Metric [21], which is a full-reference 3D-IQA metric.

The proposed algorithm combines several modules, i.e., PU-IR, DF-CS, and merging using the proposed pooling mechanism. To validate the inclusion of these modules individually and progressively, we have shown the step-wise performance of these modules in Table 3.2. It can be seen from the table that Proposed PU-IR and Proposed DF-CS individually also perform better than all the existing algorithms. The Proposed PU-IR metric obtains 0.7295, 0.7302, 0.5332, and 0.1696 of PLCC, SRCC, KRCC, and RMSE, respectively and exceeds by 0.4 % and 5.76 % in terms of PLCC and SRCC from state-of-the-art CODIF Metric [39]. The proposed DF-CS algorithm achieves 0.7848, 0.7676,

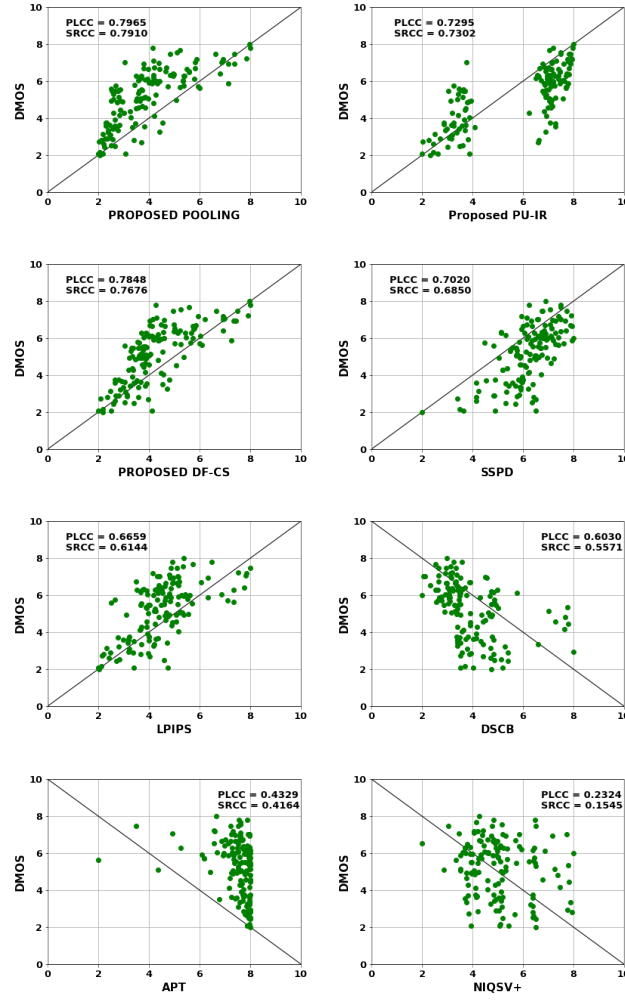
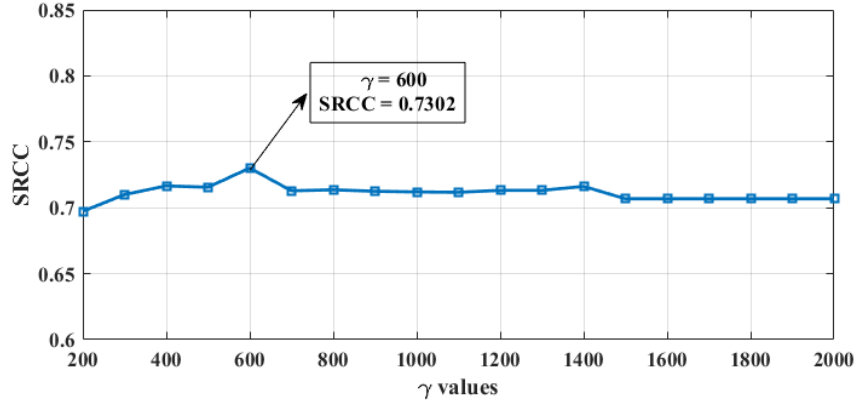


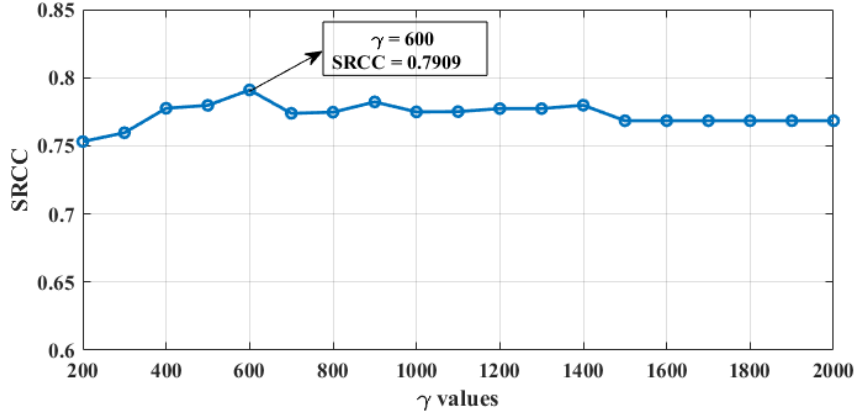
Figure 3.7: Scatter Plot between DMOS Values and Objective Scores of the 3D synthesized IQAs for the IETR dataset.

0.5753, and 0.1537 values of PLCC, SRCC, KRCC, and RMSE, respectively, 8.09 % and 11.18 % better in terms of PLCC and SRCC than the existing CODIF algorithm. Further, the final proposed pooling performs even better than the proposed PU-IR', PU-IR, and DF-CS, with at least 3.03 % gain in SRCC. These results also validate the proposed fusion method given in (13).

In Table 3.3, the proposed DF-CS algorithm's performance is given when VGG-16 based deep-features are extracted from different levels of Laplacian pyramids. From this table, it can be observed that all the levels of the Laplacian pyramid individually as well as in combination give satisfactory results, and the linear combination of Level-1 ( $\mathcal{S}_{L_1}$ ) and Level-3 ( $\mathcal{S}_{L_3}$ ) gives the best results in comparison to individual levels. It is also concluded through the extensive study by analyzing all the possible combinations of these levels. With this view, in the proposed DF-CS algorithm, Level-1 ( $\mathcal{S}_{L_1}$ ) and Level-3 ( $\mathcal{S}_{L_3}$ ) were



(a) Proposed PU-IR



(b) Proposed Final

Figure 3.8: Effect of variation of parameter  $\gamma$  on the performance of the proposed PU-IR algorithm and the proposed overall algorithm (equation (13)).

used to predict the perceptual quality of 3D synthesized images.

In the proposed DF-CS algorithm, we asserted that cosine similarity between deep features works better than extensively used other measures such as Structural SIMilarity (SSIM) [29] and Peak Signal to Noise Ratio (PSNR). To manifest the same, in Table 3.4, we have shown the performance of all these measures for the IETR Dataset. The table suggests that cosine similarity performs better for the quality assessment of 3D synthesized images than all the other measures. These results also show the importance of cosine similarity in the proposed algorithm.

The scatter plots of objective versus subjective scores for different 3D IQA metrics are compared in Fig. 3.7. The scatter plots are used to compare IQAs intuitively. We compared both proposed methods to five existing 3D IQA metrics: SSPD [21], DSCB [68], APT [57], NIQSV+ [43], LPIPS [23]. It can be depicted from the figure that scores using

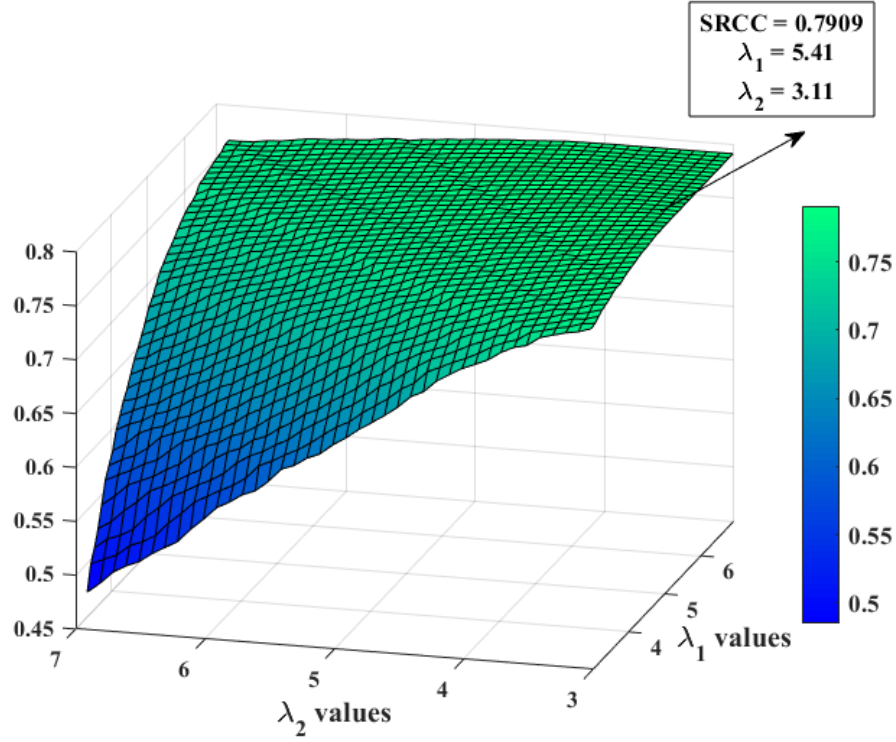


Figure 3.9: Sensitivity analysis of parameters  $\lambda_1$  and  $\lambda_2$  on the performance of the proposed algorithm.

proposed methods converge better than other metrics. Hence, all the predicted objective scores maintain more consistency with subjective ratings than state-of-the-art 3D IQA metrics.

The proposed algorithm is designed to identify all distortions, excluding black holes, as they are obsolete. Despite that, we tested the proposed algorithm for the performance on the IRCCyN dataset [18], in which the most dominant distortion is black holes. The proposed model is also found to be performing satisfactorily for this dataset. The results are shown in Table 3.5. The values of parameters  $\lambda_1$ ,  $\lambda_2$ ,  $\gamma$ , and laplacian levels are optimized for the IRCCyN dataset. As depicted in the table, the proposed algorithm's performance is comparable to the 3D IQA algorithms specifically designed for views with black holes. Also, the performance of these compared algorithms is inferior to the IETR dataset (Table 3.1). These results show that the proposed algorithm is generalized to datasets other than the IETR dataset.

### 3.2.3 Statistical Significance

Another widely adopted comparison method is statistical significance analysis to compare various quality assessment algorithms, and for this purpose, F-Test is commonly used. It

is based upon the variance hypothesis [101] between scores obtained using the proposed algorithm and the existing algorithms. The score of the F-test,  $F_s$ , is given by,

$$F_s = \frac{\sigma_{m_1}^2}{\sigma_{m_2}^2} \quad (3.15)$$

here,  $\sigma_{m_1}, \sigma_{m_2}$  indicates the RMSE values of the two metrics ( $m_1, m_2$ ) being tested. Then, a threshold  $F_c$  is calculated based on the number of images in each database with a confidence level of 90 %. If  $F_s > F_c$ , metric  $m_2$  performs statistically better than metric  $m_1$  (indicated by ‘+1’ in the Table 3.6). If  $F_s > 1/F_c$ , metric  $m_2$  performs statistically inferior to metric  $m_1$  (indicated by ‘-1’ in the Table 3.6). Otherwise, the two metrics are statistically competitive (indicated by ‘0’ in Table 3.6). Table 3.6 lists the tested IQA metrics with their F-Test values. It can be depicted from the table that the proposed algorithm is statistically better than all the compared 3D IQAs.

### 3.2.4 Parameter Sensitivity Analysis

The proposed PU-IR algorithm (5) used three parameters  $\lambda_1$ ,  $\lambda_2$ , and  $\gamma$ . The parameter  $\lambda_1$  and  $\lambda_2$  control the contribution of the proposed PU-IR algorithm in the situation of whether perceptually significant geometric distortions are available or not. To show the dependency of the proposed algorithm on these parameters, we have done the following two empirical studies:

1. The effect of variation of values of  $\lambda_1$  and  $\lambda_2$  on SRCC is shown through a 3D mesh graph (Fig. 3.9). From this figure, it can be analyzed that any value of  $\lambda_1 > \lambda_2$  gives very similar results. This empirical study validates our arguments for equation (5) that it is required to amplify the impact of geometric distortions and reduce perceptually unimportant information available in the residual image (DO).
2. Further, another parameter  $\gamma$  is used as a threshold for deciding between the two cases, as discussed in (5). To analyze the sensitivity of proposed pooling on  $\gamma$  values, we varied them from 200 to 2000. We showed the proposed algorithm’s corresponding performance in Fig. 3.8 for proposed PU-IR and for proposed final pooling. Moreover, for any positive value of  $\gamma$ , the performance of the proposed pooling is more significant than 0.7500 SRCC, which is still better than the state-of-the-art algorithms.

3. Fig. 3.10 shows the variation in performance by changing the structuring elements to ‘disk’ and ‘square’ of different ‘radius’ and ‘width,’ respectively. These plots also depict that the performance is not much changed with these parameters also.

The computational power for predicting a 3D image of  $1024 \times 768$  resolution from the IETR dataset is calculated individually for both the proposed PU-IR and DF-CS algorithms. The proposed PU-IR algorithm takes approximately 0.2 seconds to predict the perceptual quality of the 3D synthesized image. Further, the proposed DF-CS takes 0.3 seconds on a GPU (Nvidia Quadro P2000 with 16GB of RAM). Overall the proposed algorithm takes less than 1 second to process the 3D synthesized images, which is well suited for real-time applications.

### 3.3 Conclusions and Future work

This work proposed a fast and efficient algorithm for identifying geometrical and structural distortions and quality assessment of 3D synthesized images. The proposed algorithm is based on the idea that even a simple difference image can give perceptually important information and shift information. With this view, our proposed method removes the perceptually unimportant information via the morphological operation (opening). We compared the deep features extracted from the last layer of the pre-trained VGG-16 architecture to further refine the distortion. It is interesting to note that the deep features are extracted on the Laplacian pyramid of the reference and distorted 3D synthesized images. As the literature suggests, cosine similarity performs better than the mean square error if the application observes the similarity between two vectors. With this view, we also compared deep features using the cosine similarity. Finally, both algorithms were fused to estimate the perceptual quality of 3D synthesized images. The time taken by the proposed algorithm is approximately 1 second compared to 25 seconds taken by the existing algorithm SSPD.

In the proposed algorithm, the features of VGG-16 are used to estimate the perceptually relevant deep features and, subsequently, the perceptual quality of 3D synthesized images. All the features are not perceptually essential and should not be used for quality estimation purposes. For this purpose, future work should first identify perceptually important deep features and use them for quality estimation. One possible strategy could be

to calculate the eigenvalues of these deep features and use those above a certain threshold.



Table 3.1: Performance comparison of the objective IQA metrics on IETR Dataset sorted in descending order of PLCC values. - indicates that either the information is missing in the literature or the reference data is unavailable. \* indicates the performance was evaluated on a subset of the IETR dataset in the source paper.

S. No.	IQA Metric	Oriented For	No-Reference(NR)/ Full-Reference(FR)/ Side-View Reference(SVR)	PLCC	SRCC	KRCC	RMSE
1.	Proposed	3D-synthesized	FR	<b>0.7965</b>	<b>0.7909</b>	<b>0.5992</b>	<b>0.1499</b>
2.	CODIF* [39]	3D-Synthesized	NR	0.7260	0.6904	0.5033	0.1063
3.	Sadbhawna's [96]	3D-synthesized	NR	0.7087	0.6672	0.4726	0.1729
4.	SSPD [21]	3D-synthesized	FR	0.7020	0.6850	-	0.1790
5.	Yan's [97]	3D-synthesized	NR	0.6881	0.6261	0.4660	0.1750
6.	Tian's [25]	3D-synthesized	NR	0.6685	0.5903	-	0.1844
7.	LPIPS [23]	Natural Images	NR	0.6659	0.6144	0.4386	0.1850
8.	SC-IQA [73]	3D-synthesized	FR	0.6620	0.5960	-	0.1850
9.	GANs-NRM [52]	3D-synthesized	NR	0.6460	0.5710	-	0.1980
10.	LOGS [24]	3D-synthesized	SVR	0.6280	0.6160	-	0.1930
11.	MP-PSNR [26]	3D-synthesized	FR	0.6190	0.5809	0.3802	0.1947
12.	FSIM [98]	Natural Images	FR	0.6052	0.4755	0.3243	0.1973
13.	PSNR	Natural Images	FR	0.6012	0.5809	0.4024	0.1985
14.	DSCB [68]	3D-synthesized	NR	0.6030	0.5571	0.3677	0.1978
15.	GMSD [28]	Natural Images	FR	0.5560	0.4787	0.3257	0.2070
16.	MW-PSNR [27]	3D-synthesized	FR	0.5389	0.4875	0.3364	0.2088
17.	Wang's [41]	3D-synthesized	NR	0.4338	0.4254	-	0.2244
18.	APT [57]	3D-synthesized	NR	0.4329	0.4164	0.2830	0.2235
19.	BIQI [46]	Natural Images	NR	0.4327	0.4321	0.2898	0.2223
20.	SSIM [29]	Natural Images	FR	0.4016	0.2395	0.2647	0.2275
21.	SIQE [99]	3D-synthesized	SVR	0.3144	0.3418	-	0.2353
22.	DSQM [100]	3D-synthesized	SVR	0.2977	0.2369	-	0.2367
23.	OMIQA [71]	3D-synthesized	NR	0.2705	0.2331	0.1593	0.2387
24.	NIQSV + [43]	3D-synthesized	NR	0.2324	0.1545	0.1083	0.2411
25.	Yue's [3]	3D-synthesized	NR	0.1146	0.0860	-	0.2463

Table 3.2: Step-wise performance analysis of the proposed algorithm.

Metric	PLCC	SRCC	KRCC	RMSE
<b>Proposed (Pooling)</b>	<b>0.7965</b>	<b>0.7909</b>	<b>0.5992</b>	<b>0.1499</b>
Proposed DF-CS	0.7848	0.7676	0.5753	0.1537
Proposed PU-IR	0.7295	0.7302	0.5332	0.1696
Proposed PU-IR'	0.7199	0.7000	0.5096	0.1703

Table 3.3: Dependency of DF-CS algorithm on different Laplacian Levels.

Metric	PLCC	SRCC	KRCC	RMSE
Without Laplacian	0.7098	0.6998	0.5174	0.1750
Laplacian Level 1 ( $\mathcal{S}_{L_4}$ )	0.6538	0.6012	0.4317	0.1876
Laplacian Level 2 ( $\mathcal{S}_{L_3}$ )	0.7415	0.7061	0.5180	0.1663
Laplacian Level 3 ( $\mathcal{S}_{L_2}$ )	0.7665	0.7414	0.5529	0.1592
Laplacian Level 4 ( $\mathcal{S}_{L_1}$ )	0.7634	0.7491	0.5624	0.1601
$(\mathcal{S}_{L_1} + \mathcal{S}_{L_2} + \mathcal{S}_{L_3} + \mathcal{S}_{L_4})$	0.7848	0.7558	0.5673	0.1543
$(\mathcal{S}_{L_1} + \mathcal{S}_{L_3})$	<b>0.7848</b>	<b>0.7676</b>	<b>0.5753</b>	<b>0.1537</b>

Table 3.4: Effect of various metrics on fusion of deep-features. PP stands for Performance Parameter

Metric	Cosine Similarity		SSIM		PSNR	
PP	PLCC	SRCC	PLCC	SRCC	PLCC	SRCC
$\mathcal{S}_{L_1} + \mathcal{S}_{L_3}$	<b>0.7848</b>	<b>0.7676</b>	0.7566	0.7233	0.7180	0.7065

Table 3.5: Performance comparison of the proposed algorithm with existing algorithms on IRCCyN dataset, sorted in descending order of PLCC values.

<b>IQA Metric</b>	<b>NR/FR</b>	<b>PLCC</b>	<b>SRCC</b>	<b>KRCC</b>	<b>RMSE</b>
Wang's [41]	NR	0.7995	0.7869	-	0.4000
GANs-NRM [52]	NR	0.7940	0.7720	-	0.4100
IDEA [22]	FR	0.7796	0.6652	0.4986	0.3566
<b>Proposed</b>	FR	<b>0.7772</b>	<b>0.7337</b>	<b>0.4698</b>	<b>0.4189</b>
OMIQA [71]	NR	0.7678	0.7036	0.4466	0.4266
APT [57]	NR	0.7307	0.7157	-	0.4546
LOGS [24]	FR	0.7243	0.6511	0.4849	0.3890
NIQSV+ [43]	NR	0.7114	0.6668	0.4679	0.5040

Table 3.6: Statistical Significance (SS) comparison of the proposed algorithm with existing state-of-the-art IQA algorithms for the IETR dataset.

<b>Metric</b>	<b>Proposed</b>	<b>SSPD</b>	<b>APT</b>	<b>DSCB</b>	<b>NIQSV+</b>	<b>LPIPS</b>
<b>Proposed</b>	-	+1	+1	+1	+1	+1
<b>SSPD</b>	-1	-	+1	+1	+1	0
<b>APT</b>	-1	-1	-	-1	0	-1
<b>DSCB</b>	-1	-1	+1	-	+1	-1
<b>NIQSV+</b>	-1	-1	0	-1	-	-1
<b>LPIPS</b>	-1	0	+1	+1	+1	-

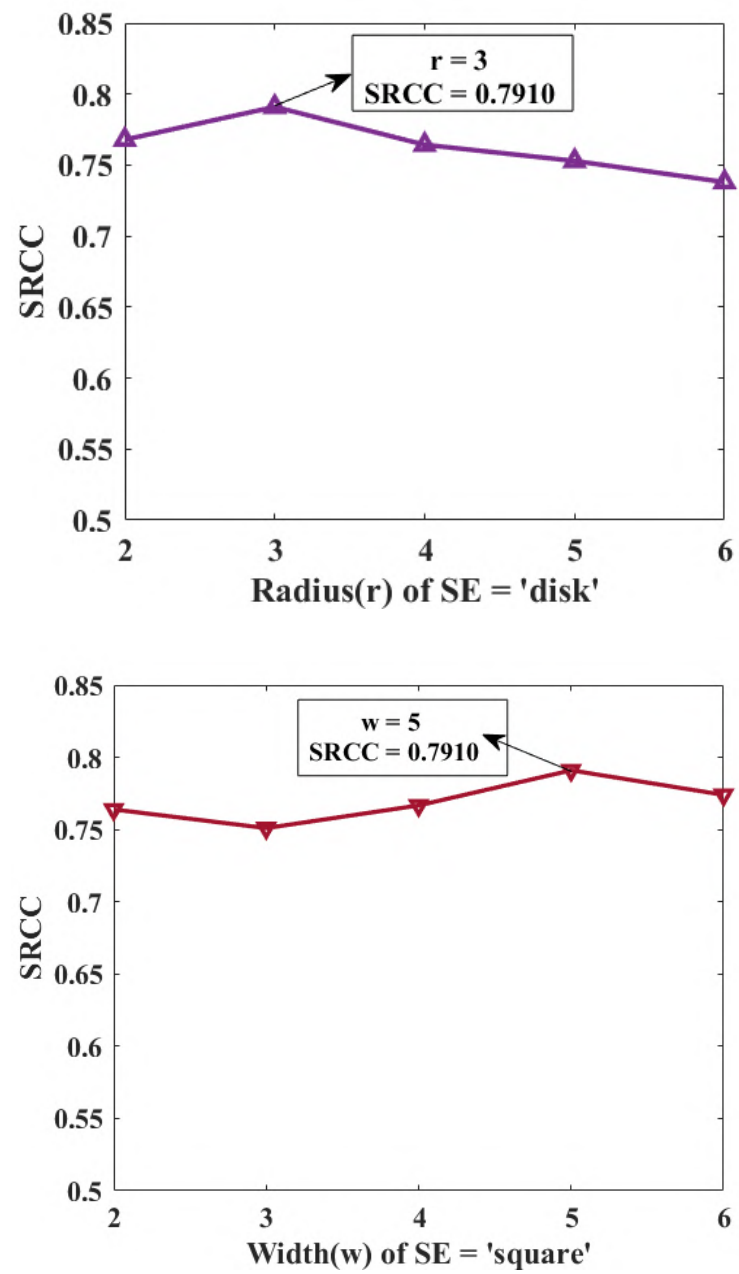


Figure 3.10: Performance variation of the proposed algorithm with change in Structuring Elements (SE).

# Chapter 4

## Full Reference 3D IQA 2

### Context Region Identification based Quality Assessment of 3D Synthesized Views

Depth information is an integral part of the DIBR process for the 3D view synthesis. In this context, a few existing algorithms have used depth information for the quality assessment. Li *et al.* [102] proposed a method for the quality assessment of depth images using the edge characteristics of depth images. In this algorithm, two types of maps are predicted using the depth information, i.e., similarity map and weighting map. They are further pooled using edge-guided pooling to get the final predicted quality score. Shao *et al.* [103] proposed a 3D synthesized video quality assessment algorithm using the local binary patterns as feature representation and dictionary learning. Liu *et al.* [104] proposed a new quality metric for distortions arising due to texture/depth compression in 3D synthesized videos. This objective metric considers two features, i.e., temporal flickering and Spatio-temporal activity. On the same line, Wang *et al.* [105] proposed a new depth perception quality assessment considering stereoscopic and spatial orientation structural features.

In the above-discussed state-of-the-art 3D QA metrics, we observed the following issues: 1. Most of the existing algorithms highly depend upon the tunable parameters and can not be generalized. 2. Although a few existing algorithms use depth information for the quality assessment of views. At the same time, none of the existing algorithms analyze the context information and incorporate this information while estimating the perceptual quality of 3D synthesized views. All existing algorithms ignore the context information, such as whether the region is foreground or not. Considering these drawbacks, in this

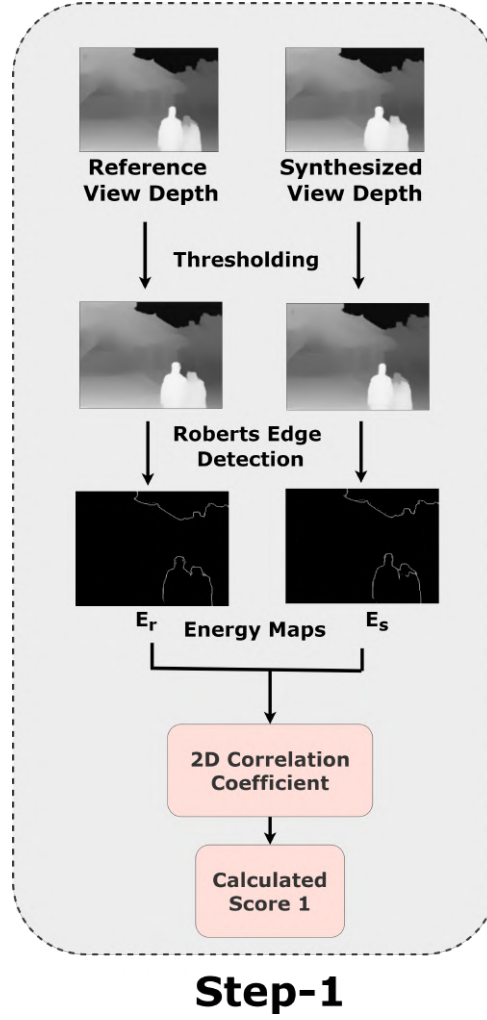


Figure 4.1: Workflow diagram of the proposed algorithm when the context region is foreground.

paper, we propose to predict the perceptual quality of 3D-synthesized views by identifying the context region and integrating it with the quality assessment model. The main contributions of the proposed algorithm are as follows:

1. In general, the geometric distortions which are perceptually important occur near the edges and disoccluded regions and not in the full 3D synthesized views. The proposed algorithm efficiently identifies the possible location of distortions and identifies the distortions only in these regions.
2. As the context region from the foreground significantly generates the distortions near the disoccluded regions. It is vital to analyze if the context region is foreground or not. With this view, the proposed algorithm identifies the context regions based upon the variation in the depth views of reference and distorted 3D synthesized views. The proposed algorithm is the first effort toward integrating the context

information with the quality assessment model.

3. The proposed algorithm is free from any parameter tuning. The proposed metric achieves better performance than the existing algorithms and requires less than a second to predict the perceptual quality of 3D synthesized views.

we have done an extensive literature survey on using depth information for the quality assessment purpose and presented below:

## 4.1 Proposed Methodology

The context region is where the information is taken for the disocclusion process during the 3D view synthesis. The context region can either be taken from the foreground, background, or both. Recent research in 3D view synthesis suggests that geometric distortions significantly affecting the perceptual quality of a 3D view occur if the context region is foreground and does not affect the quality much, otherwise [106]. With this view, it is vital to analyze the context region and use this information for 3D IQA. From empirical study, we have observed a significant variation in the depth of the distorted view concerning the depth of the reference view if the context region is foreground. In Fig. 4.1, two reference views, their synthesized views, and corresponding depths from the IETR dataset [?] are shown to validate this argument. From this figure, it can be observed that if the context region is foreground, there is a significant variation in the depths of the distorted view with respect to the reference view. Hence, the depth of information can help analyze whether the context region is foreground or background. Subsequently, this information can be utilized in estimating the quality score for 3D synthesized views.

As discussed earlier, the location of the context region (foreground or background) significantly affects the overall perceptual quality of 3D synthesized views. Thus, we propose two different steps for both situations. The first step of the algorithm is based on the correlation between the depth energy maps of the reference and distorted views. Also, we propose an algorithm based upon the Discrete Cosine transform (DCTs) of the distorted local region for the second situation. Finally, to get the consolidated quality score, we propose simply multiplying these two scores in Step 1 and Step 2. The complete workflow diagram of the proposed model is shown in Fig. 4.4. The step-wise detailed methodology of the proposed algorithm is explained as follows:

### 4.1.1 Quality Score when context region is the foreground (Step 1)

The process of 3D synthesis brings some annoying distortions to the synthesized scenes, and these distortions mainly occur around the foreground objects. Recent studies suggest that while filling the missing pixels information during 3D-synthesis, using the texture information from the background compared to the foreground [106] produces perceptually better 3D views. In this context, it is conspicuous to use the depth information of the 3D scenes in the process of 3D synthesis as also proposed in [106]. With this view, it is also essential to incorporate depth information while estimating the perceptual quality of these 3D views. Similarly, to predict the quality of these 3D synthesized views, we have obtained the depth maps of the reference and the 3D synthesized views using the single image depth estimation algorithm [107]. However, the estimated depth images are generally noisy, so we propose reducing noise using a simple thresholding method. Also, if the context region is foreground, there is a significant variation in the reference and distorted views (as shown in Fig. 4.1 and Table 2.3). This variation can be easily highlighted using the energy maps of the depths of both reference and distorted views. The two kernels (Robert's edge detection) used for estimating the energy maps are as follows:

$$\begin{bmatrix} +1 & 0 \\ 0 & -1 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & +1 \\ -1 & 0 \end{bmatrix}$$

Let  $D_r(u, v)$  and  $D_s(u, v)$  be a pixel in the reference depth and the synthesized depth views, respectively.  $G_r^x(u, v)$  and  $G_r^y(u, v)$  be the pixel in the reference, and the synthesized depth images formed by convolving with the first kernel and similarly  $G_s^x(u, v)$  and  $G_s^y(u, v)$  be a pixel in the depth images formed by convolving with the second kernel. Then gradient or the energy map of the reference and the synthesized view (i.e.  $E_r(u, v)$  and  $E_s(u, v)$ ) can then be defined as:

$$\begin{aligned} E_r(u, v) &= \sqrt{(G_r^x)^2 + (G_r^y)^2}, \quad \text{and} \\ E_s(u, v) &= \sqrt{(G_s^x)^2 + (G_s^y)^2}, \quad \text{respectively} \end{aligned} \tag{4.1}$$



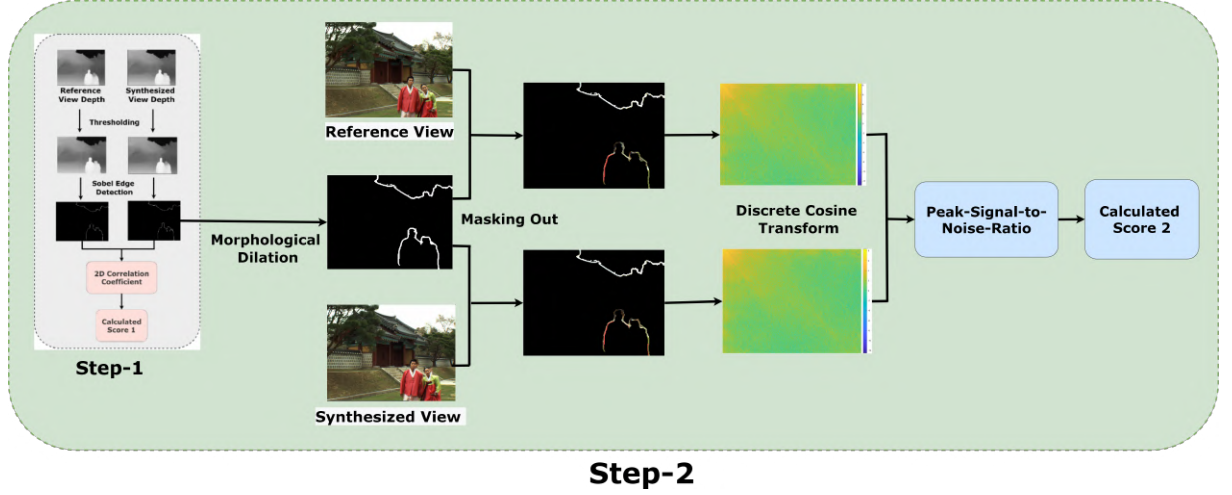


Figure 4.2: Workflow Diagram of Step 2 of the proposed algorithm when context region is background. This diagram shows the images visually after applying operations described in Step 2, such as morphological dilation, masking out, and discrete cosine transform.

Further, to identify the variation in the depth energy maps, which also act as the perceptual quality score when the context region is from the foreground, we propose calculating Pearson's correlation coefficient between the energy maps of the reference and distorted depths. Hence, the quality score for foreground context ( $Q_{FG}$ ) can be given as:

$$Q_{FG} = \frac{\sum \sum (E_r - \overline{E_r})(E_s - \overline{E_s})}{\sqrt{(\sum \sum (E_r - \overline{E_r})^2)(\sum \sum (E_s - \overline{E_s})^2)}} \quad (4.2)$$

Here,  $\overline{E_r}$  and  $\overline{E_s}$  represents the mean of energy maps  $E_r$  and  $E_s$ , respectively. From equation (2), it can be observed that a higher correlation coefficient value suggests that there is not much variation in the depth map of the distorted view with respect to the depth map of the reference view. So, a higher correlation coefficient value suggests two observations: 1. The perceptual quality of the 3D synthesized view is good, or 2. during the 3D synthesis, the context of the disoccluded region is taken from the background, and they are not generating perceptually significant distortions around the objects [106].

At this point, it is crucial to understand how  $Q_{FG}$  could quantify the quality of views based on whether the texture information was from the foreground or background. For this purpose, Table 2.2 shows the change in depths and their corresponding energy maps with different synthesized views. For these images, it can be concluded that the energy map of SV-1 (the case of context region from the foreground) is less similar to the energy map of the reference view as compared to the SV-2 (the case of context region from the background). In addition to the visual analysis, the correlation coefficient value ( $Q_{FG}$ )

for SV-1 is 0.6509, which is lesser than for SV-2, which is 0.8231. With this view, a lesser correlation between energy maps of reference and synthesized views suggests that these views are not similar and have poorer perceptual quality. Although the proposed algorithm is pretty simple, it exploits the fundamental properties of the 3D synthesized views. Table 2.5 shows that  $Q_{FG}$  individually has the performance of 0.7369 PLCC and 0.7324 SRCC, which is better than the state-of-the-art algorithms.

#### 4.1.2 Quality score when the context region is background (Step 2)

The calculated quality score ( $Q_{FG}$ ) gives a fair idea about the distortions caused by the context region as foreground by using the correlation between depth energy maps. However, when the context region is background, substantial distortions are not present around the edges [106]. Subsequently, there is not much variation in the depth of the distorted view with respect to the reference view. Table 2.3 shows the views from the IETR dataset [?] and corresponding depths to validate these arguments. There is not much variation in the depth of images in the first and second rows in Table 2.3 if the context region is the background. With this view, depth information can not be directly utilized for the quality assessment of 3D synthesized views when the context region is background. Further, in 3D synthesized views, distortions are generally present in the vicinity of objects, and their edges [22]. In algorithm IDEA [22], authors have used instance segmentation to identify the possible locations of distortions. The same task of possible location of distortion identification can be done using depth maps. So, we have proposed a new algorithm to identify the location of the distortions using depth maps in Step 2 and to predict the quality score based on the Discrete Cosine Transform (DCT) of these locations. The detailed workflow of the proposed algorithm (Step 2) is shown in Fig. 4.3.

Generally, in 3D synthesized views, most distortions are present near the edges of objects, and it is required to extract the possible location of these distortions. In Step 1, the object boundaries are estimated using depth information. We extracted the distorted region of the synthesized view and the corresponding region in the reference view by utilizing their depth information. We propose using the dilation operation on the energy map ( $E_s$ ) to get a mask image and further extracting the distorted region in the 3D views

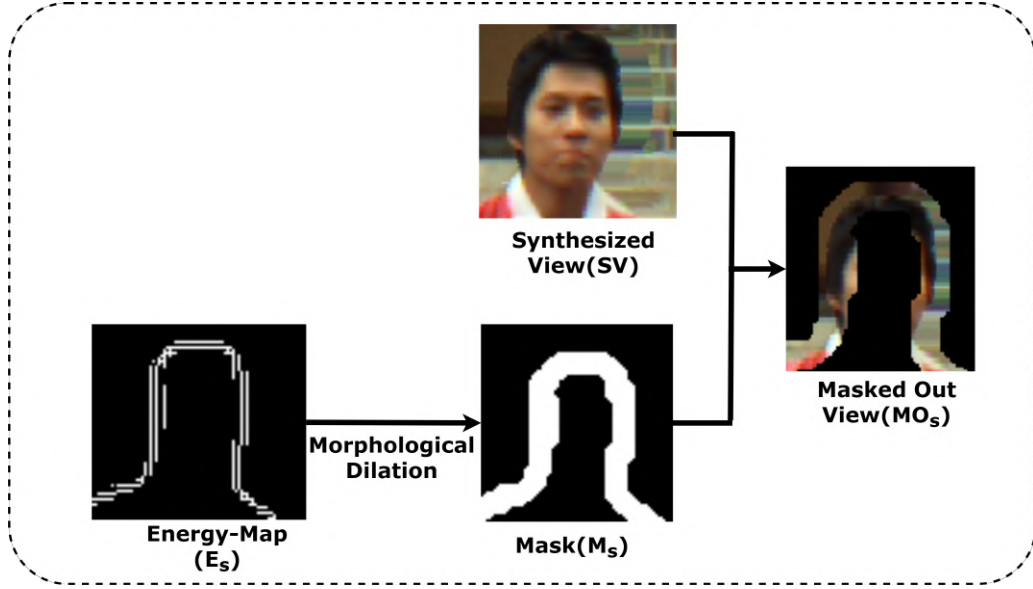


Figure 4.3: Masking out operation flow in detail.

(pictorial representation is shown in Fig. 4.4). The dilation operation is applied to the energy map of the synthesized view ( $E_s$ ) to obtain the desired mask ( $M$ ) at a pixel  $(x, y)$  is given as,

$$M_s(x, y) = E_s(x, y) \oplus S = \max_{(i,j) \in S} \{E_s(x - i, y - j)\} \quad (4.3)$$

Here  $S$  is the structuring element of the shape ‘disk’ with a radius of 2. The slight variation in  $S$ , such as shape and size, does not significantly affect the proposed algorithm’s performance. This mask is further used to extract the distorted portion from the reference view and the synthesized view termed Masking Out, i.e.,  $MO_r$  and  $MO_s$ , respectively, as:

$$MO_r = M_s \times RV \quad (4.4)$$

$$MO_s = M_s \times SV \quad (4.5)$$

The Reference and Synthesized Views are  $RV$  and  $SV$ , respectively. These masked-out references and synthesized views can identify the possible distortions and are further employed in this quality score measurement process.

The Human Visual System (HVS) is highly sensitive to the high-frequency components of images compared to the low-frequency. Even slight distortions in high-frequency distortions are perceivable by the human visual system [108]. In this context, we propose to extract the high-frequency components of the Masked Out views ( $MO_s$  and  $MO_r$ ) using

the Discrete Cosine Transform (DCT). The DCT [109] domain of images has proved to be a good alternative for quality prediction [110, 111] of natural images. We imitate this idea of using the DCT domain to extract the high-frequency components in our proposed 3D-IQA algorithm. The process of transformation of reference and synthesized masked out images, i.e.,  $MO_r$  and  $MO_s$  into the reference and synthesized transforms  $T_r$  and  $T_s$ , respectively, is defined as:

$$T_{pq} = \frac{1}{\alpha_p \alpha_q} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} MO_{mn} \cos \frac{\pi(2m+1)p}{2M} \cos \frac{\pi(2n+1)q}{2N} \quad (4.6)$$

$$\text{where, } 0 \leq p \leq M-1, 0 \leq q \leq N-1 \quad (4.7)$$

$$\alpha_p = \begin{cases} \frac{1}{\sqrt{M}}, & p = 0 \\ \sqrt{\frac{2}{M}}, & 1 \leq p \leq M-1 \end{cases} \quad (4.8)$$

$$\alpha_q = \begin{cases} \frac{1}{\sqrt{N}}, & q = 0 \\ \sqrt{\frac{2}{N}}, & 1 \leq q \leq N-1 \end{cases} \quad (4.9)$$

Where M and N are image O's row and column sizes, respectively.

After extracting the DCT coefficients, we only utilized the DCT coefficients present in the lower triangle for the quality prediction. For perceptual quality estimation, we propose to identify the variation in DCT coefficients of the masked distorted image to the masked out reference view, and the variation is estimated as:

$$Q_{BG} = \log \frac{255^2}{\sqrt{\sum (T_r - T_s)^2}} \quad (4.10)$$

Where  $T_r$  and  $T_s$  are the DCT coefficients of the reference and synthesized masked-out views, respectively.

### 4.1.3 Final Perceptual Quality Score Pooling

We proposed to fuse the above evaluated two quality scores ( $Q_{FG}$  and  $Q_{BG}$ ) into one final score ( $Q$ ) by simply combining both the scores using a multiplication operation.

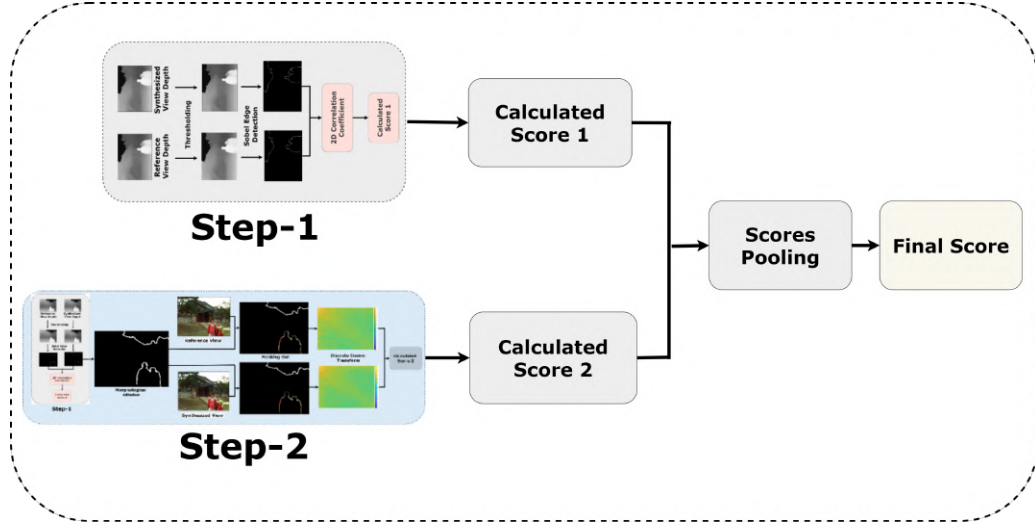


Figure 4.4: Workflow Diagram of the proposed model.

$$Q = Q_{BG} \times Q_{FG} \quad (4.11)$$

It is interesting to observe from the above equation that the proposed algorithm does not depend on any parameters. The proposed algorithm controls the contribution of foreground and background information based on the two scores calculated individually in the above sections. Both scores  $Q_{FG}$  and  $Q_{BG}$  have a proportional relationship with the perceptual quality. Hence, the final predicted score is also directly proportional to the quality. A higher value of  $Q$  indicates better image perceptual quality of 3D synthesized views and vice-versa.

## 4.2 Experimental Results

### 4.2.1 Dataset and evaluation criteria for performance comparison

We examined the proposed algorithm on two publicly available 3D IQA datasets with the established benchmark, i.e., the IETR-DIBR dataset [?] and IVY dataset [19].

#### 4.2.1.1 IETR dataset

The dataset consists of synthesized views created using 7 different Depth Image Based Rendering (DIBR) algorithms i.e.  $D1$ - $D7$  [88–94], along with their reference views. This

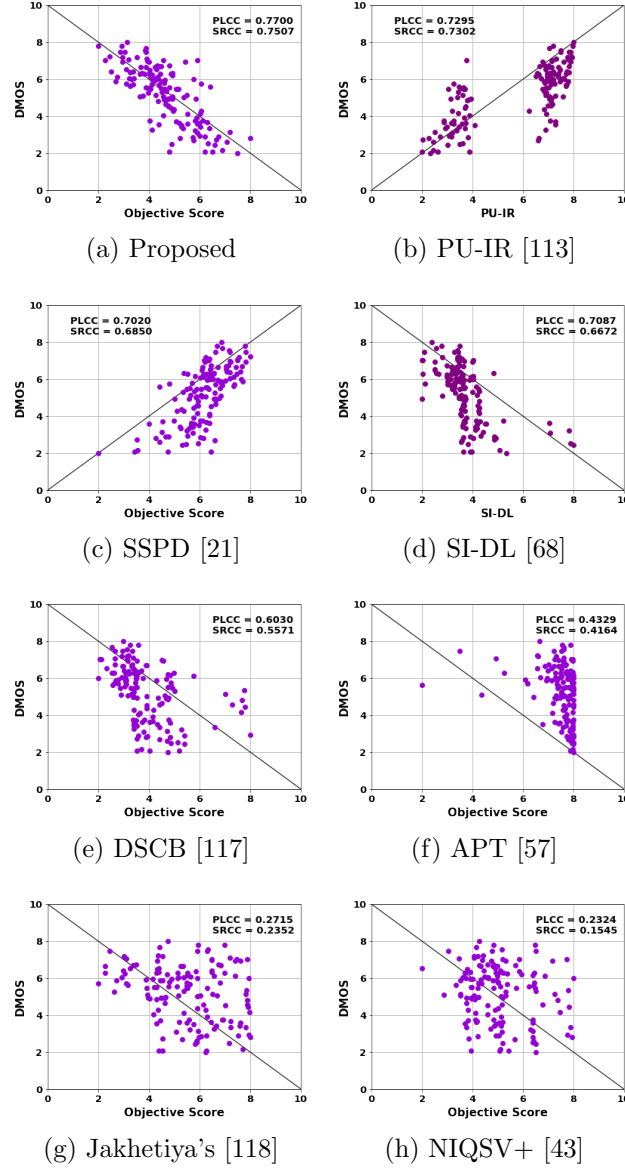


Figure 4.5: Scatter Plot between DMOS Values and Objective Scores of different IQA methods.

way, the dataset consists of 140 3D and ten original views. This dataset is distinguished from all the existing 3D IQA datasets in the literature as it suggests that black-holes type artifacts are obsolete and not included as one of the types of distortion.

#### 4.2.1.2 IVY dataset

This data consists of 84 views synthesized from 7 reference views using four different DIBR algorithms *D8-D11* [88, 90, 93, 121]. The DMOS values are generated using the double stimulus continuous quality scale. To get the final quality score, our experiment averages the quality of two views, the left- and right views.

We employed four widely used evaluation criteria for performance evaluation compared

to existing IQA algorithms, i.e., PLCC, KRCC, SRCC, and RMSE. Any IQA metric with a higher value of PLCC, SRCC, KRCC, and a lower value of RMSE is considered a better metric. Similar to the APT [57], a five-parameter non-linear mapping is used to map the calculated objective scores to subjective scores as,

$$g(x) = \gamma_1 \left( \frac{1}{2} - \frac{1}{1 + e^{\gamma_2(x - \gamma_3)}} \right) + \gamma_4 x + \gamma_5 \quad (4.12)$$

where  $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ , and  $\gamma_5$  are the five parameters to be fitted.  $g(x)$  is the mapped subjective score,  $x$  is the objective score.

### 4.2.2 Performance Analysis

We compared the proposed algorithm with 25 existing IQA algorithms, most specifically designed for the quality assessment of 3D synthesized views. Of these 25 algorithms, 13 are FR-IQAs, while the remaining 12 are NR-IQAs. The proposed algorithm achieved 0.7707, 0.7572, 0.5700, and 0.1580 values of PLCC, SRCC, KRCC, and RMSE, respectively. The proposed algorithm achieves 5.32 % and 7.60 % gain in terms of PLCC and SRCC, respectively, as compared to the best performing existing FR 3D-IQA algorithm, i.e., MLFA [112]. Further, the proposed algorithm achieved a total gain of 6.15 % and 9.67 % in terms of PLCC and SRCC, respectively, as compared to Yan's metric [116], which is an NR 3D-IQA metric. In addition to 3D Synthesized Views, we compared the proposed algorithm with IQA algorithms proposed for Natural Images. The LPIPS [23] algorithm works exceptionally well for natural images. However, for 3D synthesized views, it can only achieve 0.6659 and 0.6144 values of PLCC and SRCC, respectively. A detailed comparison of the performance analysis can be seen in Table III.

To show that both the proposed algorithms (Step 1 and Step 2) and simple fusion scheme are working efficiently, ablation study results are presented in Table 2.5. Score 1 is the Correlation Coefficient between depth energy maps, which attains 0.7369 and 0.7324 values of PLCC and SRCC, respectively. Score 2 is PSNR between DCTs of masked out views, which attains 0.4446 and 0.4248 values of PLCC and SRCC, respectively. Finally, these scores are fused using the simple multiplication operation. This ablation study validates the proposed algorithms' performance and fusion scheme.

To validate the generality of the proposed algorithm, we have also checked the per-

formance on the IVY dataset. The proposed algorithm attains 0.6726, 0.6547, 0.4775, and 10.5412 values of PLCC, SRCC, KRCC, and RMSE, respectively. The proposed algorithm performs better for this dataset than the existing 3D IQA algorithms, except for the SSPD algorithm. At the same time, the performance of the proposed algorithm is significantly better than the SSPD algorithm on the much more comprehensive dataset IETR. Further, the proposed algorithm is free from any tunable parameters, which is not the case with SSPD. Additionally, the time taken to predict the perceptual quality score by the proposed algorithm is less than a second for a 3D synthesized view. In contrast, SSPD takes around 28 seconds for the same. These results suggest that the proposed algorithm performs better than the SSPD algorithm.

In Fig. 4.6, we have compared the scatter plots of the scores from objective metrics versus Differential-Mean-Opinion-Scores (DMOS) for seven different IQA metrics from the literature to compare these IQAs intuitively. The seven existing 3D IQA metrics compared are: SSPD [21], PU-IR [113], SI-DL [68], DSCB [117], APT [57], Jakhetya's [118], and NIQSV+ [43]. From the said figure, it can be analyzed that the proposed objective scores converge better than other metrics.

### 4.2.3 Parameters Sensitivity Analysis

As discussed earlier, the proposed algorithm does not depend upon any parameters; subsequently, parameter tuning is not required. To validate this argument, we have also shown the performance of the proposed algorithm when structuring elements and the edge detection operators are varied in Fig. 4.6 and Table 2.5. These figures show that the proposed algorithm does not depend on any parameters.

We have also checked the dependency of the proposed algorithm on various edge detection algorithms such as Roberts, Sobel, Prewitt, Canny, and Holistically-Nested Edge Detection (HED) [122], and the results are shown in Table 2.7. This table shows that the proposed algorithm performs similarly when simple horizontal and vertical gradient-based edge detection operators (such as Roberts, Sobel, and Prewitt) are used. Further, when Canny and HED algorithms are used, the performance of the proposed algorithm is poor. One of the primary reasons behind the poor performance is that these algorithms cannot efficiently detect the foreground edges in depth maps and can subsequently not accurately identify the context region.



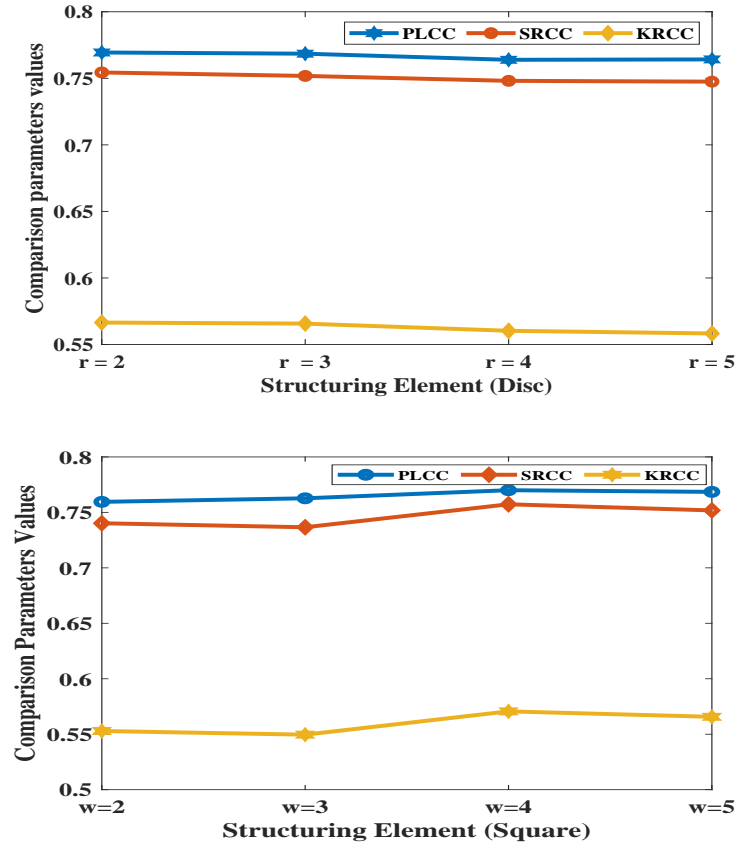


Figure 4.6: Performance dependency of the proposed algorithm with variation in Structuring Elements. Here, ‘r’ is the radius, and ‘w’ is the width in terms of pixels.

#### 4.2.4 Statistical Significance

We adopted a Statistical Significance (SS) analysis method (F-Test [101]) to compare the proposed method to other IQA algorithms. F-test is derived from the variance hypothesis between the objective scores of the proposed metric and the existing IQA algorithms. The score of the F-test,  $SS$ , is given by,

$$SS = \frac{m_{\mu_1}^2}{m_{\mu_2}^2} \quad (4.13)$$

here,  $m_{\mu_1}, m_{\mu_2}$  are the Root-Mean-Square-Error values of the two objective metrics ( $\mu_1, \mu_2$ ) being tested, respectively. With a confidence interval of 90 %, the proposed method ( $\mu_2$ ) is statistically superior to all the compared IQA algorithms.

### 4.2.5 Existing 3D IQA algorithms as a plug-in for the proposed algorithm.

The proposed idea of developing a context-aware 3D IQA algorithm is novel and can be incorporated with the existing algorithms to enhance their performances. Since no existing 3D IQA algorithm incorporates the context information in their algorithm, we examined whether the proposed algorithm can serve as a plug-in to improve the performances of the existing algorithms for 3D IQA. We define the new, improved 3D IQA metric ( $Q_{new}$ ) as,

$$Q_{new} = Q^a \pm Q_e^b \quad (4.14)$$

$Q$  and  $Q_e$  are the predicted quality scores using the proposed and existing IQA algorithms. The parameter  $a$  and  $b$  are the weighting parameters used to balance the difference in their scales and the relationship diversity. The operator  $+$  or  $-$  depends on whether both algorithms have a proportional or inversely proportional relationship. In Table 2.8, we have shown the results when the proposed algorithm is fused with the existing, BIQI [46], DSCB [117], LPIPS [23], and NIQSV+ [43] algorithms. From this table, it is clear that the proposed algorithm can work as a plug-in to improve the performance of the existing algorithms.

## 4.3 Conclusions and future work

In this work, we have used the fact that context information plays an essential role in the perceptual quality of 3D synthesized views. For example, when the context region is foreground, the synthesized views may have significant distortions near the object boundaries and vice-versa. Hence, in the proposed algorithm, we have identified the context region and proposed two new algorithms based on whether the context region is from the foreground or background. Interestingly, when the context is foreground, there is a significant variation in the depth of the reference and synthesized views. Thus, we have proposed to use depth information to identify the context region. Finally, these quality scores are fused via the simple multiplication of both scores obtained when the context region is foreground or not. The proposed algorithm achieves 0.7707 and 0.7572 values of PLCC

and SRCC, respectively, for the IETR dataset. Also, the proposed algorithm is free from parameter tuning and only requires less than a second to predict the perceptual quality of 3D synthesized views. The proposed algorithm requires complete information about the reference view, and in the future, we will use the same principle for creating the NR IQA algorithm. One possible way of creating NR IQA is by creating pseudo-reference views. The depth information can help to identify the possible distortions, and the remaining view can behave as the reference view.

Table 4.1: The literature survey shows how depth information has been used earlier for the quality assessment of 3D synthesized images.

S. No	Research Paper	Selected Views
1.	"Predicting-the-Quality-of-View-Synthesis-With-Color-Depth-Image-Fusion,"-L.-Li,-Y.-Huang,-J.-Wu,-K.-Gu,-and-Y.-Fang,-IEEE-Transactions-on-Circuits-and-Systems-for-Video-Technology,-vol.-31,-no.-7,-pp.-2509-2521,-July-2021-	-Pre-DIBR-Image-Quality-Assessment-Algorithm. The experimental results are only available for the VSRS method. Based on color-depth image fusion in the frequency domain (wavelet transform). Limited performance on the IETR dataset and performance is included in the revised manuscript.
2.	Depth-Image-Quality-Assessment-for-View-Synthesis-Based-on-Weighted-Edge-Similarity-Leida-Li,-Xi-Chen,-Yu-Zhou,-Jinjian-Wu,-Guangming-Shi;-Proceedings-of-the-IEEE/CVF-Conference-on-Computer-Vision-and-Pattern-Recognition-(CVPR)-Workshops,-2019,-pp.-17-25.	This work is for the quality assessment of depth images, not for the 3D synthesized images. It only considers the distortions in 3D synthesized images arising from poor depth images but does not include distortions due to improper rendering. Our algorithm works on predicted depths based on input views and depths. Poor performance on the IETR dataset and performance is included in the revised manuscript.
3.	"Depth-Perception-Assessment-of-3D-Videos-Based-on-Stereoscopic-and-Spatial-Orientation-Structural-Features,"-W.-Wang-et-al,-IEEE-Transactions-on-Circuits-and-Systems-for-Video-Technology,-2022.	This work is mainly for the quality assessment of depth videos, not for the 3D synthesized images. Only considers the distortions in depth images and does not consider the distortions in 3D synthesized images due to the improper rendering. Results are not available on the IR-CCyN(Video), IRCCyN(Images), IETR, and IVY datasets.

4.	"No-Reference-Quality-Prediction-for-DIBR-Synthesized-Images-Using-Statistics-of-Fused-Color-Depth-Images,"-Y.-Huang,-X.-Meng-and-L.-Li,-IEEE-Conference-on-Multimedia-Information-Processing-and-Retrieval-(MIPR),-2020,-pp.-135-138.-	CODIF (S. No 1) is a further extension of this method.
5.	"Quality-assessment-of-3D-synthesized-views-with-depth-map-distortion,"-C.-Tsai-and-H.-Hang,-2013-Visual-Communications-and-Image-Processing-(VCIP),-2013,-pp.-1-6.	This paper tries to do the shift compensation between the reference and distorted images. The assumption is that this shift arises due to the poor quality of depth images. After shift compensation, SSIM is applied to estimate the quality score. This algorithm does not use depth images for quality assessment.
6.	"Subjective-and-Objective-Video-Quality-Assessment-of-3D-Synthesized-Views-With-Texture/Depth Compression-Distortion,"-X.-Liu,-Y.-Zhang,-S.-Hu,-S.-Kwong,-C.-.-J.-Kuo-and-Q.-Peng,-IEEE-Transactions-on-Image-Processing,-vol.-24,-no.-12,-pp.-4847-4861,-Dec.-2015	This work mainly consider distortions due to compression of depth images and 3D videos. For 3D views containing depth distortions, and have not used depth information in obtaining the quality score.

Table 4.2: effect of context region as foreground and background on depth and energy map of the 3D-synthesized views. Here, Synthesized View-1 (SV-1) and Synthesized View-2 (SV-2) are the case of context region from foreground and background, respectively. CC stands for Correlation Coefficient










<b>View Type</b>	<b>Patch</b>	<b>Depth</b>	<b>energy map</b>	<b>CC</b>
<b>Reference View</b>				-
<b>SV-1</b>				<b>0.6509</b>
<b>SV-2</b>				<b>0.8231</b>

Table 4.3: Analysis of the effect of context region as foreground (FG) and background (BG) on the depth maps.





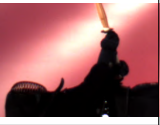



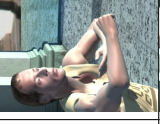
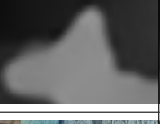




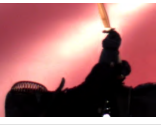




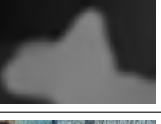

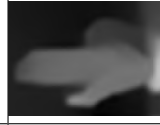


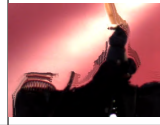




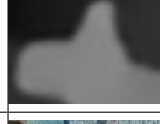
	Patch- 1	Depth- 1	Patch- 2	Depth- 2	Patch- 3	Depth- 3	Patch- 4	Depth- 4	Patch- 5	Depth- 5
Reference View										
context re- gion from BG										
Context re- gion from FG										

Table 4.4: comparison for the performance of the proposed algorithm with existing IQA algorithms on the IETR dataset. Unavailability of data is indicated using the ‘-’ symbol. ‘ $\uparrow$ ’ indicates a higher value is better while ‘ $\downarrow$ ’ indicates a lower value is better.

	S. No.	Metrics	Oriented for	Trained on IETR Dataset?	Trainable parameters?	PLCC $\uparrow$	SRCC $\uparrow$	KRCC $\uparrow$	RMSE $\downarrow$
Full-Reference	1.	<b>Proposed</b>	<b>3D Synthesized Views</b>	No	No	<b>0.7707</b>	<b>0.7572</b>	<b>0.5700</b>	<b>0.1580</b>
	2.	MLFA [112]	3D Synthesized Views	No	No	0.7378	0.7036	-	0.1899
	3.	PU-IR [113]	3D Synthesized Views	No	No	0.7295	0.7302	0.5332	0.1696
	4.	SSPD [21]	3D Synthesized Views	No	Yes	0.7020	0.6850	-	0.1790
	5.	LPIPS [23]	Natural Images	No	No	0.6659	0.6144	0.4386	0.1850
	6.	SC-IQA [73]	3D Synthesized Views	No	No	0.6620	0.5960	-	0.1850
	7.	LOGS [24]	3D Synthesized Views	No	Yes	0.6280	0.6160	-	0.1930
	8.	MP-PSNR [26]	3D Synthesized Views	No	No	0.6190	0.5809	0.3802	0.1947
	9.	PSNR	Natural Images	No	No	0.6012	0.5809	0.4024	0.1985
	10.	MW-PSNR [27]	3D Synthesized Views	No	No	0.5389	0.4875	0.3364	0.2088
	11.	Cheon’s [114]	Natural Images	No	No	0.4644	0.4416	0.3151	0.2882
	12.	Li’s* [102]	Depth Images	No	Yes	0.4584	0.4304	0.3009	0.4155
	13.	Lao’s [115]	Natural Images	No	No	0.4266	0.4458	0.3040	0.3062
	14.	SSIM [29]	Natural Images	No	No	0.4016	0.2395	0.2647	0.2275
No-Reference	1.	CODIF [39]	3D Synthesized Views	No	Yes	0.7260	0.6904	0.5033	0.1063
	2.	SI-DL [68]	3D Synthesized Views	Yes	Yes	0.7087	0.6672	0.4726	0.1749
	3.	Yan’s [116]	3D Synthesized Views	Yes	Yes	0.6881	0.6261	0.4660	0.1750
	4.	GANs-NRM [40]	3D Synthesized Views	No	No	0.6460	0.5710	-	0.1980
	5.	DSCB [117]	3D Synthesized Views	No	Yes	0.6030	0.5571	0.3677	0.1978
	6.	Wang’s [53]	3D Synthesized Views	No	Yes	0.4338	0.4254	-	0.2244
	7.	APT [57]	3D Synthesized Views	No	Yes	0.4329	0.4164	0.2830	0.2235
	8.	Wang’s [105]	3D Synthesized Videos	No	No	0.4230	0.4259	-	0.2243
	9.	Jakhetiya’s [118]	3D Synthesized Views	No	Yes	0.2715	0.2352	0.1607	0.2386
	10.	OMIQA [119]	3D Synthesized Views	No	Yes	0.2705	0.2331	0.1593	0.2387
	11.	NIQSV+ [43]	3D Synthesized Views	No	Yes	0.2324	0.1545	0.1083	0.2411
	12.	Yue [120]	3D Synthesized Views	No	Yes	0.1146	0.0860	-	0.2463



Table 4.5: Ablation study of the proposed algorithm.

	IETR Dataset		IVY Dataset	
Stage	PLCC↑	SRCC↑	PLCC↑	SRCC↑
Score 1 ( $Q_{FG}$ )	0.7369	0.7324	0.6142	0.5974
Score 2 ( $Q_{BG}$ )	0.4446	0.4248	0.4833	0.4950
<b>Final Score (Pooling)</b>	<b>0.7707</b>	<b>0.7572</b>	<b>0.6726</b>	<b>0.6547</b>

Table 4.6: Comparison of the proposed algorithm with existing algorithms on the IVY dataset for performance. Unavailability of data is indicated using the ‘-’ symbol.

IQA Metric	PLCC↑	SRCC↑	KRCC↑	RMSE↓
SSPD [21]	0.6892	0.6814	0.4872	10.3210
<b>Proposed</b>	<b>0.6726</b>	<b>0.6547</b>	<b>0.4775</b>	<b>10.5412</b>
LOGS [24]	0.6442	0.6385	0.4509	18.8549
IDEA [22]	0.6311	0.6132	0.4405	19.0379
MP-PSNR [26]	0.6114	0.5954	0.4217	19.0379
SI-DL [68]	0.5459	0.5396	-	11.9349
MW-PSNR [27]	0.5240	0.5051	0.3528	20.9969
APT [57]	0.5240	0.4748	0.3389	20.9961
NIQSV+ [43]	0.2191	0.2990	0.2037	24.0530

Table 4.7: Comparison of the proposed algorithm when different edge detection methods are used for edge detection.

Edge Detection	PLCC↑	SRCC↑	KRCC↑	RMSE↓
<i>Roberts</i>	0.7707	0.7572	0.5700	0.1580
<i>Sobel</i>	0.7720	0.7528	0.5659	0.1576
<i>Prewitt</i>	0.7652	0.7515	0.5659	0.1596
<i>Canny</i>	0.5851	0.5198	0.3616	0.2010
<i>HED</i> [122]	0.5320	0.5496	0.3934	0.2114

Table 4.8: The proposed algorithm as a plug-in to improve the performance of the existing algorithms on the IETR dataset.

Stage	PLCC $\uparrow$	SRCC $\uparrow$	KRCC $\uparrow$	RMSE $\downarrow$
Proposed	0.7707	0.7572	0.5700	0.1580
BIQI [46]	0.4327	0.4321	0.2898	0.2223
<b>Proposed with BIQI</b>	<b>0.8131</b>	<b>0.8032</b>	<b>0.6125</b>	<b>0.1443</b>
Gain (in %age)	5.59	6.99	8.31	9.63
DSCB [117]	0.6030	0.5571	0.3677	0.1978
<b>Proposed with DSCB</b>	<b>0.8068</b>	<b>0.7925</b>	<b>0.5982</b>	<b>0.1465</b>
Gain (in %age)	4.77	5.54	5.78	7.98
LPIPS [23]	0.6659	0.6144	0.4386	0.1850
<b>Proposed with LPIPS</b>	<b>0.7817</b>	<b>0.7583</b>	<b>0.5723</b>	<b>0.1546</b>
Gain (in %age)	1.15	0.98	1.14	2.26
NIQSV+ [43]	0.2324	0.1545	0.1083	0.2411
<b>Proposed with NIQSV+</b>	<b>0.7726</b>	<b>0.7538</b>	<b>0.5612</b>	<b>0.1574</b>
Gain (in %age)	0.33	0.41	0.76	0.50

Table 4.9: Comparison of the proposed algorithm with different edge detection methods.

Edge Detection	PLCC	SRCC	KRCC	RMSE
<i>Roberts</i>	0.7707	0.7572	0.5700	0.1580
<i>Sobel</i>	0.7720	0.7528	0.5659	0.1576
<i>Prewitt</i>	0.7652	0.7515	0.5659	0.1596
<i>Canny</i> [1]	0.5851	0.5198	0.3616	0.2010
<i>HED</i> [2]	0.5320	0.5496	0.3934	0.2114

# Chapter 5

## Conclusions and Future Work

### 5.1 Future Work



Figure 5.1: (a). A synthesized view. (b). The failure (green arrows) of a random patch (red window) in a 3D synthesized view. Synthesized Using: [4]

Free Viewpoint Video (FVV), 3D-Television, 360°video, and Virtual Reality (VR) are some of the applications of 3D-synthesis, famous because of their realistic and interactive experience [4, 123]. Unfortunately, the rendered 3D views, even using contemporary methods, cannot generate the perfect novel 3D view [4]. These methods cannot perform efficiently on complex surfaces and produce some artifacts, as shown in Fig. 5.1. The artifacts in the 3D synthesized views differ from conventional artifacts in regular natural images. With the advancement of efficient algorithms for generating 3D synthesized views, it is required to have an image quality assessment (IQA) algorithm which can automatically judge the perceptual quality of generated 3D synthesized views that match

Metric	Oriented for	PLCC	SRCC	RMSE
<b>LOGS</b>	3D Views	0.6350	0.6021	0.8400
<b>PSNR</b>	Natural Images	0.2869	0.1772	1.0417
<b>SSIM</b>	Natural Images	0.1610	0.1231	1.1735
<b>LPIPS</b>	Natural Images	0.1921	0.0132	1.3874
<b>APT</b>	3D Views	0.1717	0.0013	1.3849

Table 5.1: Performance of state-of-the-art IQA algorithms for the proposed test dataset.

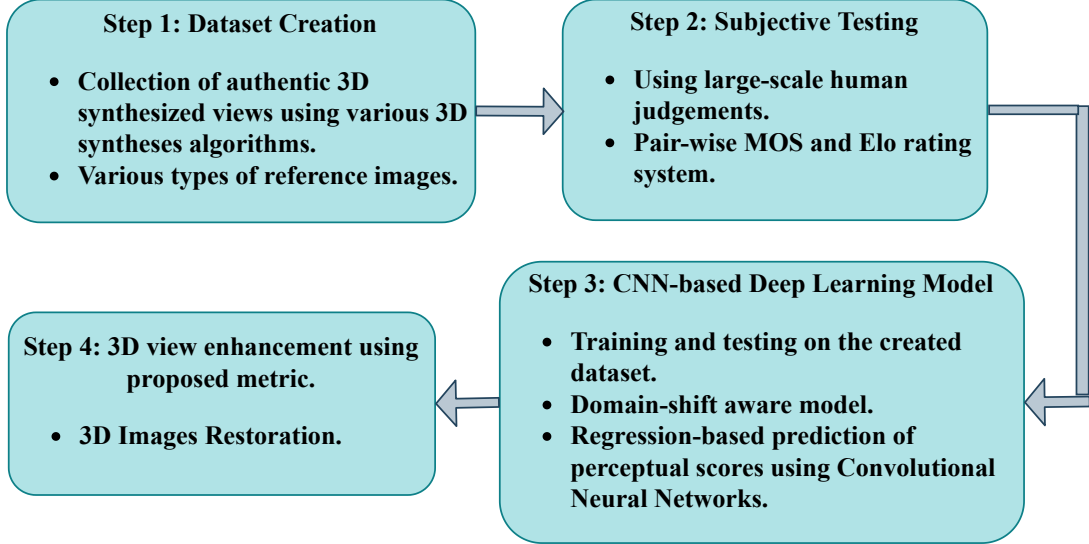


Figure 5.2: Step-wise flow of the proposed future work.

with the human visual system. The 3D IQA algorithms can judge the perceptual quality and are also helpful in the fast development of Image Restoration (IR) (IR includes tasks such as super-resolution (SR), denoising, enhancement, etc.) algorithms. With this view, Is there a need to create a large-scale 3D synthesized IQA database generated using the recently proposed 3D synthesized view generation algorithms?

To the best of our knowledge, there is no large-scale dataset for quality evaluation of 3D synthesized views. Subsequently, no generic IQA algorithms are proposed to judge the quality of 3D synthesized views. The quality evaluation datasets and metrics used by contemporary 3D synthesis methods for various purposes are designed for natural images (for example, to determine the threshold in Section 3.1 in the paper [4], authors used the LPIPS metric [23] which is initially designed for naturally degraded images and not for 3D images).

The process of the creation of the IQA dataset and its subjective testing is hectic. In this context, to validate that the proposed problem is worth pursuing, we created a small test dataset of 60 3D views generated using two recent 3D algorithms (i.e., [4,123]), tested

using five expert subjects. This subjective testing is also validated using Cohen’s Kappa coefficient. The performance of five popular IQA metrics (LOGS [124], Peak Signal to Noise Ratio(PSNR), SSIM [29], LPIPS [23], APT [124] ) oriented for natural as well as 3D images are given in Table 5.1. The error in Table 5.1 suggest that the literature has no proper algorithm for this purpose.

Our preliminary analysis suggests a need for a new perceptual metric designed explicitly for 3D views for 3D image restoration and enhancement. For this purpose, the future steps involved in the proposed process are summarized in Figure 5.2.

## 5.2 Conclusions

In this thesis, we tried to understand the fundamental properties of the 3D synthesized views such as: the shift between the reference and synthesized views and context information. In chapter 2, we propose to use an interesting observation that there is a direct relationship between the number of blocks with stretching artifacts with the perceptual quality of 3D synthesized view. We have proposed to use Deep-learning based approach to detect the blocks with stretching artifacts. In chapter 3, we proposed an efficient and simple approach using morphological operations to reduce the perceptually unimportant information arising due to the shift between the reference and distorted 3D synthesized views. In chapter 4, we have proposed an algorithm based on the context information. Recent progress in the 3D view synthesis domain suggests that if the context region is foreground, substantial distortions are present in the 3D synthesized views and vice-versa. With this view, we first propose to use depth information to identify context information. Also, the same depth information is used to estimate the quality score when the context region is foreground or not. The final quality score is estimated by multiplying the scores obtained when the context region is foreground or not.

# Bibliography

- [1] S. Tian, L. Zhang, L. Morin, and O. Déforges, “Niqsv+: A no-reference synthesized view quality assessment metric,” *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1652–1664, 2018.
- [2] S. Tian, L. Zhang, L. Morin, and O. Deforges, “A benchmark of dibr synthesized view quality assessment metrics on a new database for immersive media applications,” *IEEE Transactions on Multimedia*, vol. 21, no. 5, pp. 1235–1247, 2019.
- [3] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, “Combining local and global measures for dibr-synthesized image quality evaluation,” *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2075–2088, 2019.
- [4] Meng-Li Shih, Shih-Yang Su, Johannes Kopf, and Jia-Bin Huang, “3d photography using context-aware layered depth inpainting,” in *CVPR*, 2020.
- [5] A. Q. de Oliveira, M. Walter, and C. R. Jung, “An artifact-type aware dibr method for view synthesis,” *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1705–1709, 2018.
- [6] T. Masayuki, M. Tehrani, T. Fujii, and T. Yendo, “Free-viewpoint TV,” *IEEE Signal Processing Magazine*, 2011.
- [7] A. M. Andrew, “Virtual reality: Exploring the brave new technologies of artificial experience and interactive worlds from cyberspace to teledildontics,” *Robotica*, vol. 10, no. 3, pp. 278–279, 1992.
- [8] M. Shih, S. Su, J. Kopf, and J. Huang, “3d photography using context-aware layered depth inpainting,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

- [9] C. Fehn, “A 3D-TV approach using depth-image-based rendering (DIBR),” *Visualization, Imaging, and Image Processing*, 2003.
- [10] E. Bosc, R. Pepion, P. Le Callet, M. Koppel, P. Ndjiki-Nya, M. Pressigout, and L. Morin, “Towards a new quality metric for 3-d synthesized view assessment,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, 2011.
- [11] A. Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [12] D. Wang, Y. Zhao, Z. Wang, and H. Chen, “Hole-filling for dibr based on depth and gradient information,” *International Journal of Advanced Robotic Systems*, vol. 12, no. 2, 2015.
- [13] I. Ahn and C. Kim, “A novel depth-based virtual view synthesis method for free viewpoint video,” *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 614–626, 2013.
- [14] G. Luo, Y. Zhu, Z. Li, and L. Zhang, “A hole filling approach based on background reconstruction for view synthesis in 3d video,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1781–1789.
- [15] M. Solh and G. AlRegib, “Hierarchical hole-filling for depth-based view synthesis in ftv and 3d video,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 495–504, 2012.
- [16] O. Stankiewicz, K. Wegner, M. Tanimoto, and M. Domański, “Enhanced view synthesis reference software (vsrs) for free-viewpoint television,” 2013.
- [17] C. Zhu and S. Li, “Depth image based view synthesis: New insights and perspectives on hole generation and filling,” *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 82–93, 2016.
- [18] Emilie Bosc, Romuald Pepion, Patrick Le Callet, Martin Koppel, Patrick Ndjiki-Nya, Muriel Pressigout, and Luce Morin, “Towards a new quality metric for 3-d

- synthesized view assessment,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 7, pp. 1332–1343, 2011.
- [19] Y. J. Jung, H. Kim, and Y. Ro, “Critical binocular asymmetry measure for perceptual quality assessment of synthesized stereo 3d images in view synthesis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 1–1, 01 2015.
- [20] R. Song, H. Ko, and C.-C. J. Kuo, “Mcl-3d: A database for stereoscopic image quality assessment using 2d-image-plus-depth source,” *Journal of Information Science and Engineering*, vol. 31, 03 2014.
- [21] S. Mahmoudpour and P. Schelkens, “Synthesized view quality assessment using feature matching and superpixel difference,” *IEEE Signal Processing Letters*, vol. 27, pp. 1650–1654, 2020.
- [22] L. Li, Y. Zhou, J. Wu, F. Li, and G. Shi, “Quality index for view synthesis by measuring instance degradation and global appearance,” *IEEE Transactions on Multimedia*, vol. 23, pp. 320–332, 2021.
- [23] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *IEEE conference on Computer Vision and Pattern Recognition*, 2018.
- [24] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, “Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness,” *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 914–926, 2017.
- [25] S. Tian, L. Zhang, L. Morin, and O. Deforges, “A full-reference image quality assessment metric for 3-d synthesized views,” *Proc. Image Quality Syst. Perform. Conf., IS&T Electron. Imag., Soc. Imaging Sci. Technol.*, vol. 12, pp. 366–1–366–5, 2018.
- [26] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, “Multi-scale synthesized view assessment based on morphological pyramids,” *Journal of Electrical Engineering*, vol. 67, pp. 1–9, 01 2016.



- [27] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, “Dibr synthesized image quality assessment based on morphological wavelets,” in *International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, pp. 1–6.
- [28] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, “Gradient magnitude similarity deviation: A highly efficient perceptual image quality index,” *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.
- [29] 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, 2004.
- [30] S. Tian, L. Zhang, L. Morin, and O. Déforges, “Sc-iqu: Shift compensation based image quality assessment for dibr-synthesized views,” in *2018 IEEE Visual Communications and Image Processing (VCIP)*, 2018, pp. 1–4.
- [31] Y. Zhou, L. Li, S. Ling, and P. Le Callet, “Quality assessment for view synthesis using low-level and mid-level structural representation,” *Signal Processing: Image Communication*, vol. 74, pp. 309–321, 2019.
- [32] S. Ling, J. Li, P. L. Callet, and Junle Wang, “Perceptual representations of structural information in images: Application to quality assessment of synthesized view in ftv scenario,” in *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 1735–1739.
- [33] Y. Zhang, H. Zhang, M. Yu, S. Kwong, and Y. Ho, “Sparse representation-based video quality assessment for synthesized 3d videos,” *IEEE Transactions on Image Processing*, vol. 29, pp. 509–524, 2020.
- [34] S. Ling and P. L. Callet, “Image quality assessment for dibr synthesized views using elastic metric,” 2017, p. 1157–1163, Proceedings of the 25th ACM International Conference on Multimedia.
- [35] S. Mahmoudpour and P. Schelkens, “Synthesized view quality assessment using feature matching and superpixel difference,” *IEEE Signal Processing Letters*, vol. 27, pp. 1650–1654, 2020.

- [36] L. Li, Y. Zhou, K. Gu, W. Lin, and S. Wang, “Quality assessment of dibr-synthesized images by measuring local geometric distortions and global sharpness,” *IEEE Transactions on Multimedia*, vol. 20, no. 4, pp. 914–926, 2018.
- [37] S. Tian, L. Zhang, L. Morin, and O. Deforges, “A full-reference image quality assessment metric for 3-d synthesized views,” *Proc. Image Quality Syst. Perform. Conf., IST Electron. Imag., Soc. Imaging Sci. Technol.*, vol. 12, pp. 3661–3665, 2018.
- [38] D. Sandić-Stanković, D. Kukolj, and P. Le Callet, “Multi-scale synthesized view assessment based on morphological pyramids,” *Journal of Electrical Engineering*, vol. 67, pp. 1–9, 01 2016.
- [39] L. Li, Y. Huang, J. Wu, K. Gu, and Y. Fang, “Predicting the quality of view synthesis with color-depth image fusion,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 31, no. 7, pp. 2509–2521, 2021.
- [40] S. Ling, J. Li, Z. Che, J. Wang, W. Zhou, and P. L. Callet, “Re-visiting discriminator for blind free-viewpoint image quality assessment,” *IEEE Transactions on Multimedia*, pp. 1–1, 2020.
- [41] G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu, “Blind quality metric of dibr-synthesized images in the discrete wavelet transform domain,” *IEEE Transactions on Image Processing*, vol. 29, pp. 1802–1814, 2020.
- [42] J. Yan, Y. Fang, R. Du, Y. Zeng, and Y. Zuo, “No reference quality assessment for 3d synthesized views by local structure variation and global naturalness change,” *IEEE Transactions on Image Processing*, pp. 1–1, 2020.
- [43] S. Tian, L. Zhang, L. Morin, and O. Déforges, “Niqsv+: A no-reference synthesized view quality assessment metric,” *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1652–1664, 2018.
- [44] V. Jakhetiya, K. Gu, S. Jaiswal, T. Singhal, and Z. Xia, “Kernel ridge regression based quality measure and enhancement of 3d-synthesized images,” *IEEE Transactions on Industrial Electronics*, pp. 1–1, 2020.

- [45] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4695–4708, 2012.
- [46] A.K. Moorthy and A.C. Bovik, “A two-step framework for constructing blind image quality indices,” *IEEE Signal Processing Letters*, vol. 17, no. 5, pp. 513–516, 2010.
- [47] K. Gu, J. Qiao, S. Lee, H. Liu, W. Lin, and P. Le C., “Multiscale natural scene statistical analysis for no-reference quality evaluation of dibr-synthesized views,” *IEEE Transactions on Broadcasting*, vol. 66, no. 1, pp. 127–139, 2020.
- [48] D. Sandić-Stanković, D. D. Kukolj, and P. L. Callet, “Fast blind quality assessment of dibr-synthesized video based on high-high wavelet subband,” *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5524–5536, 2019.
- [49] G. Wang, Z. Wang, K. Gu, and Z. Xia, “Blind quality assessment for 3d-synthesized images by measuring geometric distortions and image complexity,” in *ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 4040–4044.
- [50] S. Ling and P. L. Callet, “How to learn the effect of non-uniform distortion on perceived visual quality? case study using convolutional sparse coding for quality assessment of synthesized views,” in *25th IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 286–290.
- [51] X. Wang, K. Wang, B. Yang, F. W.B. Li, and X. Liang, “Deep blind synthesized image quality assessment with contextual multi-level feature pooling,” in *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 435–439.
- [52] S. Ling, J. Li, Z. Che, J. Wang, W. Zhou, and P. L. Callet, “Re-visiting discriminator for blind free-viewpoint image quality assessment,” *IEEE Transactions on Multimedia*, pp. 1–1, 2020.
- [53] G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu, “Blind quality metric of dibr-synthesized images in the discrete wavelet transform domain,” *IEEE Transactions on Image Processing*, vol. 29, pp. 1802–1814, 2020.

- [54] G. Wang, Z. Wang, K. Gu, K. Jiang, and Z. He, “Reference-free dibr-synthesized video quality metric in spatial and temporal domains,” *IEEE Transactions on Circuits and Systems for Video Technology*, pp. 1–1, 2021.
- [55] A. Q. de Oliveira, M. Walter, and C. R. Jung, “An artifact-type aware dibr method for view synthesis,” *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1705–1709, 2018.
- [56] S. M. Muddala, M. Sjöström, and R. Olsson, “Virtual view synthesis using layered depth image generation and depth-based inpainting for filling disocclusions and translucent disocclusions,” *Journal of Visual Communication and Image Representation*, vol. 38, pp. 351–366, 2016.
- [57] K. Gu, V. Jakhetiya, J. Qiao, X. Li, W. Lin, and D. Thalmann, “Model-based referenceless quality metric of 3d synthesized images using local image description,” *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 394–405, 2018.
- [58] R. Song, H. Ko, and C.-C. Jay Kuo, “Mcl-3d: A database for stereoscopic image quality assessment using 2d-image-plus-depth source,” *J. Inf. Sci. Eng.*, vol. 31, pp. 1593–1611, 2015.
- [59] X. Wang, K. Wang, B. Yang, F. W. B. Li, and X. Liang, “Deep blind synthesized image quality assessment with contextual multi-level feature pooling,” in *IEEE International Conference on Image Processing (ICIP)*, 2019, pp. 435–439.
- [60] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, “Cnn features off-the-shelf: An astounding baseline for recognition,” in *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 512–519.
- [61] J. Johnson, A. Alahi, and L. Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European Conference on Computer Vision*, 2016.
- [62] A. Dosovitskiy and T. Brox, “Generating images with perceptual similarity metrics based on deep networks,” in *International Conference on Neural Information Processing Systems*, 2016, NIPS, p. 658–666.
- [63] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv 1409.1556*, 09 2014.

- [64] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 248–255.
- [65] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.
- [66] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [67] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Signal Processing Letters*, vol. 20, no. 3, pp. 209–212, 2013.
- [68] Sadbhawna, V. Jakhetiya, D. Mumtaaz, and S. Jaiswal, “Distortion specific contrast based no-reference quality assessment of dibr-synthesized views,” in *IEEE Workshop on Multimedia and Signal Processing (MMSP)*, 2020, pp. 1–6.
- [69] D. Kundu, L. K. Choi, A. C. Bovik, and B. L. Evans, “Perceptual quality evaluation of synthetic pictures distorted by compression and transmission,” *Signal Processing: Image Communication*, vol. 61, pp. 54–72, 2018.
- [70] G. Wang, Z. Wang, K. Gu, L. Li, Z. Xia, and L. Wu, “Blind quality metric of dibr-synthesized images in the discrete wavelet transform domain,” *IEEE Transactions on Image Processing*, vol. 29, pp. 1802–1814, 2020.
- [71] V. Jakhetiya, K. Gu, T. Singhal, S. C. Guntuku, Z. Xia, and W. Lin, “A highly efficient blind image quality assessment metric of 3-d synthesized images using outlier detection,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4120–4128, 2019.
- [72] 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, 2006.

- [73] S. Tian, L. Zhang, L. Morin, and O. Déforges, “Sc-iqu: Shift compensation based image quality assessment for dibr-synthesized views,” in *2018 IEEE Visual Communications and Image Processing (VCIP)*, 2018, pp. 1–4.
- [74] V. Jakhetiya, K. Gu, S. Jaiswal, T. Singhal, and Z. Xia, “Kernel ridge regression based quality measure and enhancement of 3d-synthesized images,” *IEEE Transactions on Industrial Electronics*, 2020.
- [75] S. Tian, L. Zhang, W. Zou, X. Li, T. Su, L. Morin, and O. Déforges, “Quality assessment of dibr-synthesized views: An overview,” *Neurocomputing*, vol. 423, pp. 158–178, 2021.
- [76] D. Ghadiyaram and A. C. Bovik, “Massive online crowdsourced study of subjective and objective picture quality,” *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 372–387, 2016.
- [77] Y. Zhang, H. Zhang, M. Yu, S. Kwong, and Y. Ho, “Sparse representation-based video quality assessment for synthesized 3d videos,” *IEEE Transactions on Image Processing*, vol. 29, pp. 509–524, 2020.
- [78] P. Burt and E. Adelson, “The laplacian pyramid as a compact image code,” *IEEE Transactions on communications*, vol. 31, no. 4, pp. 532–540, 1983.
- [79] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, “A survey on deep transfer learning,” in *Artificial Neural Networks and Machine Learning ICANN 2018*. pp. 270–279, Springer International Publishing.
- [80] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?,” in *Advances in Neural Information Processing Systems*. 2014, vol. 27, pp. 3320–3328, Curran Associates, Inc.
- [81] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *European Conference on Computer Vision*, 2016.
- [82] Z. Zhang, Z. Wang, Z. Lin, and H. Qi, “Image super-resolution by neural texture transfer,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7974–7983.

- [83] S. Wang, L. Zhang, W. Zuo, and B. Zhang, “Class-specific reconstruction transfer learning for visual recognition across domains,” *IEEE Transactions on Image Processing*, vol. 29, pp. 2424–2438, 2020.
- [84] D. Wang, H. Lu, and C. Bo, “Visual tracking via weighted local cosine similarity,” *IEEE Transactions on Cybernetics*, vol. 45, no. 9, pp. 1838–1850, 2015.
- [85] X. Li, J. Wu, Z. Sun, Z. Ma, J. Cao, and J. H. Xue, “Bsnet: Bi-similarity network for few-shot fine-grained image classification,” *IEEE Transactions on Image Processing*, vol. 30, pp. 1318–1331, 2021.
- [86] D. Zhong and J. Zhu, “Centralized large margin cosine loss for open-set deep palmprint recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 6, pp. 1559–1568, 2020.
- [87] S. Eghbali and L. Tahvildari, “Fast cosine similarity search in binary space with angular multi-index hashing,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 2, pp. 329–342, 2019.
- [88] A. Criminisi, P. Perez, and K. Toyama, “Region filling and object removal by exemplar-based image inpainting,” *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1200–1212, 2004.
- [89] D. Wang, Y. Zhao, Z. Wang, and H. Chen, “Hole-filling for dibr based on depth and gradient information,” *International Journal of Advanced Robotic Systems*, vol. 12, no. 2, pp. 12, 2015.
- [90] I. Ahn and C. Kim, “A novel depth-based virtual view synthesis method for free viewpoint video,” *IEEE Transactions on Broadcasting*, vol. 59, no. 4, pp. 614–626, 2013.
- [91] G. Luo, Y. Zhu, Z. Li, and L. Zhang, “A hole filling approach based on background reconstruction for view synthesis in 3d video,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1781–1789.
- [92] M. Solh and G. AlRegib, “Hierarchical hole-filling for depth-based view synthesis in ftv and 3d video,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 6, no. 5, pp. 495–504, 2012.

- [93] O. Stankiewicz, K. Wegner, M. Tanimoto, and M. Domański, “Enhanced view synthesis reference software (vsrs) for free-viewpoint television,” 2013.
- [94] C. Zhu and S. Li, “Depth image based view synthesis: New insights and perspectives on hole generation and filling,” *IEEE Transactions on Broadcasting*, vol. 62, no. 1, pp. 82–93, 2016.
- [95] A. Benoit, P. Le Callet, P. Campisi, and R. Cousseau, “Quality assessment of stereoscopic images,” *Eurasip Journal on Image and Video Processing - EURASIP J Image Video Process*, vol. 2008, 12 2008.
- [96] Sadbahwna, V. Jakhetiya, D. Mumtaz, B.N. Subudhi, and S.C. Guntuku, “Stretching artifacts identification for quality assessment of 3d-synthesized views,” *IEEE Transactions on Image Processing*, 2022.
- [97] J. Yan, Y. Fang, R. Du, Y. Zeng, and Y. Zuo, “No reference quality assessment for 3d synthesized views by local structure variation and global naturalness change,” *IEEE Transactions on Image Processing*, 2020.
- [98] L. Zhang, L. Zhang, X. Mou, and D. Zhang, “Fsim: A feature similarity index for image quality assessment,” *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [99] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Objective quality metric for 3d virtual views,” in *IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3720–3724.
- [100] M. S. Farid, M. Lucenteforte, and M. Grangetto, “Perceptual quality assessment of 3d synthesized images,” in *IEEE International Conference on Multimedia and Expo (ICME)*, 2017, pp. 505–510.
- [101] R. Zhu, F. Zhou, W. Yang, and J. Xue, “On hypothesis testing for comparing image quality assessment metrics [tips tricks],” *IEEE Signal Processing Magazine*, vol. 35, no. 4, pp. 133–136, 2018.
- [102] L. Li, X. Chen, Y. Zhou, J. Wu, and G. Shi, “Depth image quality assessment for view synthesis based on weighted edge similarity,” in *CVPR Workshops*, 2019.



- [103] F. Shao, Q. Yuan, W. Lin, and G. Jiang, “No-reference view synthesis quality prediction for 3-d videos based on color–depth interactions,” *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 659–674, 2018.
- [104] X. Liu, Y. Zhang, S. Hu, S. Kwong, C.-C. . Kuo, and Q. Peng, “Subjective and objective video quality assessment of 3d synthesized views with texture/depth compression distortion,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 4847–4861, 2015.
- [105] G. Wang, Z. Wang, K. Gu, K. Jiang, and Z. He, “Reference-free dibr-synthesized video quality metric in spatial and temporal domains,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 3, pp. 1119–1132, 2022.
- [106] M. L. Shih, S. Y. Su, J. Kopf, and J. B. Huang, “3d photography using context-aware layered depth inpainting,” in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8025–8035.
- [107] Z. Li and N. Snavely, “Megadepth: Learning single-view depth prediction from internet photos,” in *Computer Vision and Pattern Recognition (CVPR)*, 2018.
- [108] V. Jakhetiya, W. Lin, S. Jaiswal, K. Gu, and S. C. Guntuku, “Just noticeable difference for natural images using rms contrast and feed-back mechanism,” *Neurocomputing*, vol. 275, pp. 366–376, 2018.
- [109] N. Ahmed, T. Natarajan, and K. R. Rao, “Discrete cosine transform,” *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, 1974.
- [110] M. Saad, A. Bovik, and C. Charrier, “Dct statistics model-based blind image quality assessment,” in *Proceedings of International Conference on Image Processing (ICIP)*, 09 2011, pp. 3093–3096.
- [111] T. Brandão and M. Queluz, “No-reference image quality assessment based on dct domain statistics,” *Signal Processing*, vol. 88, pp. 822–833, 04 2008.
- [112] C. Ji, Z. Peng, W. Zou, F. Chen, G. Jiang, and M. Yu, “No-reference quality assessment for 3d synthesized images based on visual-entropy-guided multi-layer features analysis,” *Entropy (Basel)*, 2021.

- [113] Sadbhawna, V. Jakhetiya, S. Chaudhary, B. N. Subudhi, W. Lin, and S. C. Guntuku, “Perceptually unimportant information reduction and cosine similarity-based quality assessment of 3d-synthesized images,” *IEEE Transactions on Image Processing*, vol. 31, pp. 2027–2039, 2022.
- [114] M. Cheon, S. Yoon, B. Kang, and J. Lee, “Perceptual image quality assessment with transformers,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2021, pp. 433–442.
- [115] S. Lao, Y. Gong, S. Shi, S. Yang, T. Wu, J. Wang, W. Xia, and Y. Yang, “Attentions help cnns see better: Attention-based hybrid image quality assessment network,” *arXiv preprint arXiv:2204.10485*, 2022.
- [116] J. Yan, Y. Fang, R. Du, Y. Zeng, and Y. Zuo, “No reference quality assessment for 3d synthesized views by local structure variation and global naturalness change,” *IEEE Transactions on Image Processing*, vol. 29, pp. 7443–7453, 2020.
- [117] Sadbhawna, V. Jakhetiya, D. Mumtaz, and S. P. Jaiswal, “Distortion specific contrast based no-reference quality assessment of dibr-synthesized views,” in *IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*, 2020, pp. 1–5.
- [118] V. Jakhetiya, K. Gu, S. P. Jaiswal, T. Singhal, and Z. Xia, “Kernel-ridge regression-based quality measure and enhancement of three-dimensional-synthesized images,” *IEEE Transactions on Industrial Electronics*, vol. 68, no. 1, pp. 423–433, 2020.
- [119] V. Jakhetiya, K. Gu, T. Singhal, S. C. Guntuku, Z. Xia, and W. Lin, “A highly efficient blind image quality assessment metric of 3-d synthesized images using outlier detection,” *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4120–4128, 2019.
- [120] G. Yue, C. Hou, K. Gu, T. Zhou, and G. Zhai, “Combining local and global measures for dibr-synthesized image quality evaluation,” *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 2075–2088, 2019.

- [121] S. S. Yoon, H. Sohn, Y. J. Jung, and Y. M. Ro, “Inter-view consistent hole filling in view extrapolation for multi-view image generation,” in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 2883–2887.
- [122] S. Xie and Z. Tu, “Holistically-nested edge detection,” in *IEEE International Conference on Computer Vision*, 2015.
- [123] Simon Niklaus, Long Mai, Jimei Yang, and Feng Liu, “3d ken burns effect from a single image,” *ACM Transactions on Graphics*, 2019.
- [124] Shishun Tian, Lu Zhang, Wenbin Zou, Xia Li, Ting Su, Luce Morin, and Olivier Deforges, “Quality assessment of dibr-synthesized views: An overview,” *Neurocomputing*, 2021.

# List of Publications

1. Context Region Identification based Quality Assessment of 3D Synthesized Views.  
**Sadbhawna**, V Jakhetiya, BN Subudhi, S Jaiswal, L Li, W Lin  
IEEE Transactions on Multimedia, 2022
2. Perceptually Unimportant Information Reduction and Cosine Similarity-based Quality Assessment of 3D-Synthesized Images.  
**Sadbhawna**, V Jakhetiya, S Chaudhary, BN Subudhi, W Lin, SC Guntuku  
IEEE Transactions on Image Processing, 2022
3. Stretching Artifacts Identification for Quality Assessment of 3D-Synthesized Views  
**Sadbhawna**, V Jakhetiya, D Mumtaz, BN Subudhi, SC Guntuku  
IEEE Transactions on Image Processing, 2022
4. Do We Need a New Large-Scale Quality Assessment Database for Generative Inpainting Based 3D View Synthesis?(Student Abstract)  
**Sadbhawna**, V Jakhetiya, BN Subudhi, H Shakya, D Mumtaz  
2022 Thirty-Sixth AAAI Conference.
5. NTIRE 2022 challenge on perceptual image quality assessment ....., A. Keshari, Komal, **Sadbhawna**, V. Jakhetiya, B. N. Subudhi, ..... Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 2022
6. Detecting Covid-19 and Community Acquired Pneumonia Using Chest CT Scan Images With Deep Learning.  
S Chaudhary, **Sadbhawna**, V Jakhetiya, BN Subudhi, U Baid, SC Guntuku

IEEE ICASSP 2021

7. Distortion specific contrast- based no-reference quality assessment of dibr-synthesized views.

**Sadbhawna**, V Jakhetiya, D Mumtaz, SP Jaiswal

2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)