Image Quality Assessment of 3D Synthesized Views

Shishun Tian

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

INSA RENNES

THESE DE DOCTORAT DE

L'INSA RENNES
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : *Signal, Image, Vision*

Par

« Shishun TIAN »

« Image Quality Assessment of 3D Synthesized Views»

«Évaluation de la qualité des images obtenues par synthèse de vues 3D»

Thèse présentée et soutenue à Rennes, le 22 Mars 2019
Unité de recherche : L'Institut d'électronique et de Télécommunications de Rennes, UMR 6164
Thèse N° : 19ISAR 04 / D19 - 04

Rapporteurs avant soutenance :

William PUECH	Professeur à l'Université Montpellier
Marco CAGNAZZO	Professeur à Télécom ParisTech

Composition du Jury :

Attention, en cas d'absence d'un des membres du Jury le jour de la soutenance, la composition ne comprend que les membres présents

Président :
Patrick LE CALLET Professeur à l'Université de Nantes

Membres du jury :

William PUECH	Professeur à l'Université Montpellier
Marco CAGNAZZO	Professeur à Télécom ParisTech
Chaker LARABI	Maître de conférences à l'Université de Poitiers
Frédéric DEVERNAY	Senior Applied Scientist, Amazon

Directeur de thèse :
Luce MORIN Professeure à l'INSA de Rennes
Co-encadrante de thèse :
Lu ZHANG Maître de conférences à l'INSA de Rennes

Invité :
Olivier DEFORGES Professeur à l'INSA de Rennes

Intitulé de la thèse: INSA de Rennes

Titre : Image Quality Assessment of 3D Synthesized Views

Shishun TIAN

En partenariat avec :

ACKNOWLEDGEMENT

I am grateful for the past three and a half years of my Ph.D. study in the laboratory IETR and VAADER team at INSA de Rennes. I have benefited a lot from the excellent research environment in the laboratory. I deeply believe that this thesis will not have been finished without the support and kindness of my supervisors, my colleagues, my friends and my family.

I would like to express my sincere gratitude to my supervisors Lu Zhang, Luce Morin and Olivier Déforges. Their insightful and inspiring guidance, their continuous support and encouragement are indispensable to the accomplishment of this thesis. I learned a lot from their serious attitude and passion for life and work.

I also would like thank Prof. William Puech and Marco Cagnazzo for reviewing my thesis and Prof. Patrick Le Callet, Chaker Larabi and Frédéric Devernay for accepting to be the members of the committee. Thank you for all the comments and questions that help to improve the quality of this thesis.

I thank all my colleagues in the laboratory for their helps and friendliness during my life in France. I will always remember the precious time we have spent together. I must thank my close and dear friends who keep me company and gave me the courage on the road of PhD study.

I'd also like to thank China Scholarship Council (CSC) for funding.

I sincerely thank my parents, my brother and all my family for their endless love and support in my life. I want to especially thank my fiancée Ting for being very understanding and considerable, for the beautiful memories of travelling and for the countless hours we have spent on the TGV between Rennes and Lyon every weekend.

Résumé en français

À mesure que les besoins de l'expérience visuelle augmentent, les applications offrant une perception plus immersive de la scène visuelle 3D ont acquis un intérêt exceptionnel pour le public et sa curiosité au cours des dix dernières années. Des films 3D de haute qualité peuvent être visionnés dans des milliers de cinémas 3D de nouvelle génération à travers le monde. Entre-temps, certaines applications 3D, telles que la télévision en 3D et Free-viewpoint TV (FTV), sont devenues de plus en plus matures sur le plan technique. Par exemple, Canon a annoncé en septembre 2017 son système Free-Viewpoint Video (FVV), qui offre aux utilisateurs un meilleur QOE où ils peuvent visionner des événements sportifs sous différents angles et points de vue.

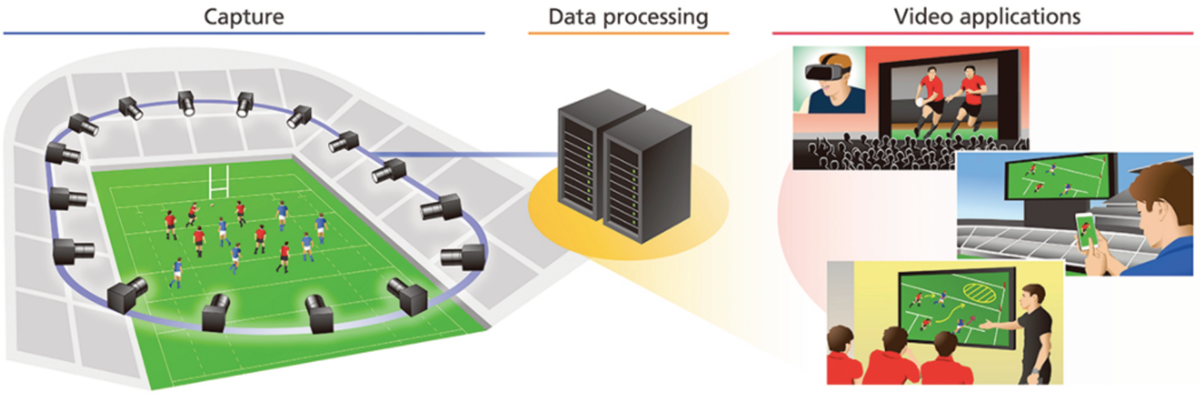


Figure 1: Exemple de Canon's Free Viewpoint Video System.[1]

Pour fournir des informations pour de telles applications 3D, plusieurs vues doivent être compressées, transmises et stockées. Multiview video coding (MVC), l'extension multiview de H.264 / MPEG-4 AVC, a été explorée pour compresser les séquences vidéo 3D. Cependant, lorsque le nombre de vues augmente, le coût en débit est très élevé. Même avec un très grand nombre de vues, il est toujours impossible de couvrir tous les points de vue d'une scène particulière.

Pour pallier cet inconvénient, Multiview-Video-Plus-Depth (MVD) a été proposé pour les représentations vidéo 3D. Il se compose de plusieurs vues de texture de dif-

[1] Source: `http://global.canon/en/news/2017/20170921.html`

férents points de vue et de leurs cartes de profondeur associées. En utilisant le format MVD, seul un nombre limité de vues originales et leur carte de profondeur correspondante doivent être codés et transmis. Les vues virtuelles supplémentaires côté récepteur peuvent être synthétisées à partir des vues décodées et des profondeurs basées sur le rendu en profondeur par image (DIBR). En outre, des méthodes de compression efficaces: le codage vidéo multiview à haute efficacité (MV-HEVC) et le codage vidéo à haute efficacité (3D-HEVC) ont été développées par le groupe de normalisation sur le développement d'une extension de codage vidéo 3D (JCT-3V) pour compresser les données du MVD. Bénéficiant des méthodes de codage efficaces et du DIBR, MVD est devenu l'une des méthodes de présentation 3D les plus populaires.

Bien que le potentiel de DIBR soit élevé, les algorithmes actuels du DIBR peuvent introduire de nouveaux types de distorsions, bien différents de celles provoquées par la compression vidéo. La plupart des normes de codage vidéo reposent sur la transformation en cosinus discrète, qui entraîne des distorsions spécifiques, par exemple, "blur", "blockiness" et "ringing". Ces distorsions sont souvent dispersées sur l'ensemble de l'image, tandis que les artefacts synthétisés par DIBR sont pour la plupart locaux. Les artefacts de vue synthétisés par DIBR proviennent d'une compression avec pertes de profondeur et d'une synthèse de vue, et se produisent généralement dans les zones exclues. Étant donné que la plupart des mesures de qualité objective 2D couramment utilisées sont initialement conçues pour évaluer les distorsions de codage courantes, elles peuvent échouer dans l'évaluation de la qualité des images contenant des distorsions dues à la synthèse et à la compression.

Les contributions présentées dans cette thèse visent à améliorer l'évaluation de la qualité des vues synthétisées par le DIBR. Bien que plusieurs efforts aient été faits pour évaluer la qualité des vues synthétisées par le DIBR, celui-ci ne peut toujours pas aboutir à un résultat satisfaisant. La première contribution de cette thèse est la proposition de deux métriques No-reference (NR) et de deux métriques Full-reference (FR) pour des vues synthétisées par DIBR. Deuxièmement, lors de l'étude des métriques, il n'existait pas de base de données adéquate pour l'évaluation de la qualité des vues synthétisées par DIBR; une nouvelle base de données d'images DIBR est proposée.

Organisation de la thèse

La thèse est organisée comme suit. L'introduction générale de cette thèse est donnée dans ce chapitre, puis le chapitre 1 introduit le principe de la synthèse de vues 3D et analyse ses distorsions particulières. Ensuite, les principales contributions de cette thèse sont organisées dans les trois chapitres suivants. Le chapitre 2 présente deux mesures de qualité sans référence basées sur la morphologie; chapitre 3 présente les deux métriques FR proposées et le chapitre 4 est consacré à notre nouvelle base de données DIBR. Enfin, la conclusion et les perspectives sont présentées au chapitre 5.

Chapitre 1 Vue d'ensemble de la synthèse de vue 3D et de l'analyse de distorsion

Ce chapitre présente les bases de la synthèse de vues 3D. Tout d'abord, la perception de la vision 3D du Human Vision System (HVS) et les formats de représentation de contenu les plus couramment utilisés pour les vidéos 3D sont présentés en détail. Nous discuterons ensuite du principe de la synthèse de vues DIBR et de son influence particulière sur la qualité des vues synthétisées. Pour conclure, quelques métriques objectives de pointe dédiées à l'évaluation de la qualité des vues synthétisées par DIBR sont brièvement examinées.

Chapitre 2 Métriques Full-reference (FR) proposées pour l'évaluation de la qualité d'image

Ces dernières années, plusieurs métriques FR ont été proposées pour évaluer la qualité des vues synthétisées par le DIBR. Cependant, aucune d'entre eux n'a pu obtenir des performances satisfaisantes. Dans ce chapitre, deux métriques de qualité d'image FR axées sur la distorsion et la désocclusion géométriques dans les vues synthétisées par DIBR sont présentées. Tout d'abord, nous observons qu'il existe une grande quantité de distorsions géométriques et de décalage d'objet dans les vues synthétisées par DIBR que les métriques de qualité 2D traditionnelles ne peuvent pas évaluer. Cela est dû au fait que le système HVS est plus sensible aux artefacts locaux que le changement d'objet global. Ainsi, une approche SURF + RANSAC est utilisée pour compenser approximativement le changement d'objet global. Ensuite, dans le premier modèle de qualité FR SC-DM, nous utilisons un masque de désocclusion pour pondérer la différence entre l'image synthétisée et l'image de référence déformée. Il peut être utilisé avec n'importe quelle métrique de qualité basée sur les pixels. D'autre part, dans la deuxième métrique de qualité FR SC-IQA, une méthode de concordance de blocs multi-résolution est proposée pour compenser avec précision le décalage d'objet et pénaliser également la distorsion géométrique locale. En outre, une carte de saillance est également utilisée pour pondérer les distorsions finales. Les résultats expérimentaux montrent que le modèle de qualité SC-DM améliore considérablement les performances du PSNR et du SSIM. La métrique SC-IQA surpasse toutes les métriques de qualité testées de l'état de l'art, y compris les métriques FR, NR et RR.

Chapitre 3 Métriques No-reference (NR) proposées pour l'évaluation de la qualité d'image

Comme indiqué au chapitre 1, dans certaines applications 3D telles que FVV, même avec un très grand nombre de vues, il n'est toujours pas possible de couvrir tous les points de vue arbitraires d'une scène donnée. C'est-à-dire qu'il n'y a pas d'image de référence pour la vue synthétisée dans certaines circonstances. Dans ce cas, les métriques de qualité NR sont nécessaires. Dans ce chapitre, nous proposons deux métriques de qualité d'image complètement NR (NIQSV et NIQSV+) pour les vues synthétisées par DIBR. Ces deux métriques reposent toutes deux sur l'hypothèse que l'image de bonne qualité est supposée présenter des contours nets et réguliers, des valeurs lissées à l'intérieur de l'objet et des discontinuités importantes aux limites de l'objet. De telles images sont insensibles aux opérations morphologiques d'ouverture et de fermeture, tandis que certains artefacts tels que les zones floues autour des bords de l'objet et l'effritement dans les vues synthétisées sont sensibles à de telles opérations morphologiques. Sur la base de cette propriété, NIQSV utilise une paire d'opérations d'ouverture et de fermeture pour mesurer ces distorsions. En tant que version étendue de NIQSV, NIQSV+ améliore ses performances en estimant les deux autres types de distorsions: trous noirs et étirement. Ensuite, les résultats expérimentaux de la base de données d'images IRCCyN / IVC DIBR montrent que la métrique NIQSV proposée surpasse les métriques 2D traditionnelles et se rapproche de la performance du FR en mode DIBR dédié. Le NIQSV+ proposé occupe la deuxième place dans toutes les métriques dédiées au DIBR, y compris FR et NR, et il n'existe aucune différence significative entre les métriques dédiées à la vue synthétisées par DIBR.

Chapitre 4 Base de référence pour métriques d'évaluation de la qualité de la vue synthétisée DIBR

Ce chapitre présente une nouvelle base de données d'images synthétisées par DIBR, qui se concentre sur les distorsions produites uniquement par différentes méthodes de synthèse de DIBR. Tout d'abord, nous donnons un aperçu des bases de données existantes du DIBR, puis nous présentons la contribution principale de ce travail. Par rapport aux bases de données existantes, nous testons des algorithmes plus nombreux et plus récents, notamment la synthèse entre vues et la synthèse à vue unique, ce qui exclut les distorsions «anciennes».

Chapitre 5 Conclusion et perspectives

Ce chapitre conclut l'apport de cette thèse: deux métriques de qualité NR, deux métriques de qualité FR dédiées aux vues synthétisées par DIBR et une nouvelle base de données d'images synthétisées DIBR accessible au public.

TABLE OF CONTENTS

INTRODUCTION

Objective and contributions of the thesis

The past decade has witnessed the fast increasement of the 3D CINEMA market size, cf. Fig. 2 . However, this stereoscopic video can only provide two viewpoint videos, the observer can not get a stereoscopic perception at another viewpoint. Nowadays, the customers are desiring the applications which can provide more immersive perception.

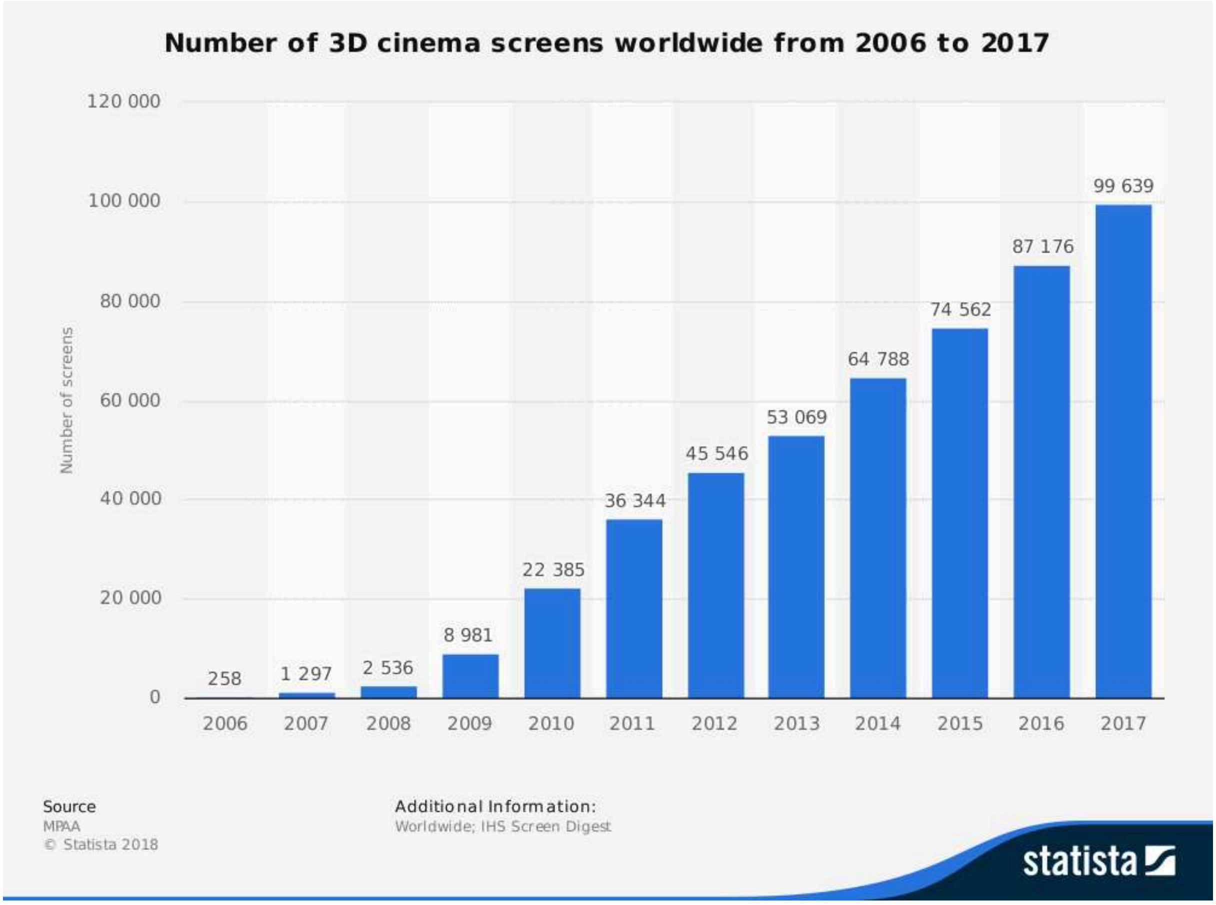


Figure 2: Number of 3D cinema screens worldwide from 2006 to 2017.[1]

[1] Source: https://www.statista.com/statistics/271863/number-of-3d-cinema-screens-worldwide/

Based on stereoscopic videos, several immersive applications are proposed in recent years. Such as Multi-view Video (MVV), Free-viewpoint Video (FVV) [91], and Virtual Reality (VR). Especially, FVV allows the users to view a 3D scene by freely changing the viewpoints. For example, Canon announced on September 2017 its Free Viewpoint Video System that gives the users a better Quality of Experience (QoE) where they can view sporting events from various different angles and viewpoints, cf. Fig. 3. However, containing much more views, these applications require a huge size of data. At the same time, it is also practically impossible to acquire images at all the viewpoints of a particular 3D scene, which is instead captured by multiple cameras at different viewpoints. Thus, some of the views have to be synthesized.

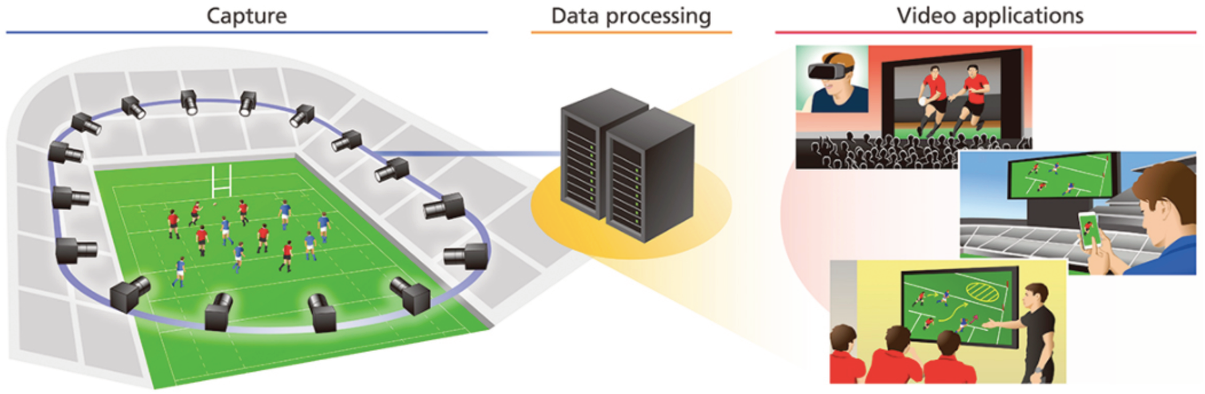
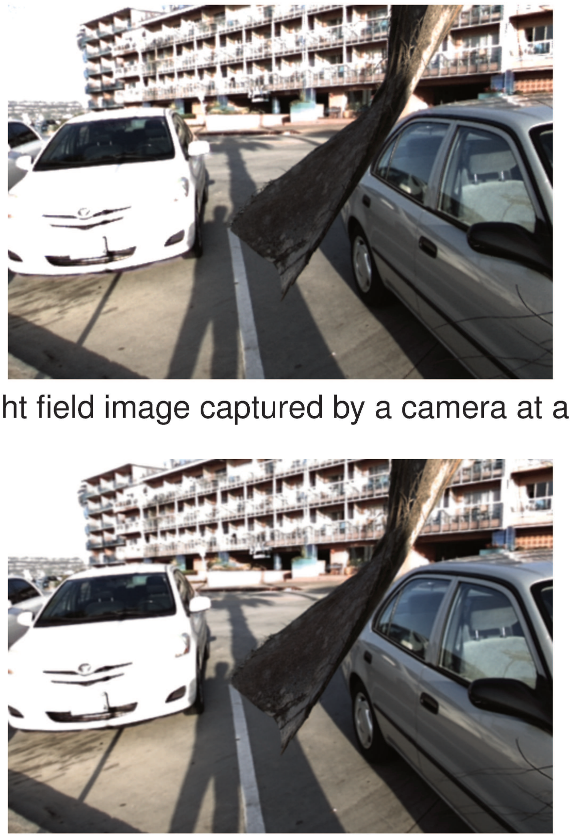


Figure 3: Example of Canon’s Free Viewpoint Video System.[2]

As one widely accepted data representation format for 3D scenes, the format Multiview Video plus Depth (MVD) [57] consists of multiple texture images and their corresponding depth maps at some particular viewpoints. The other views are then synthesized through the Depth-Image-Based-Rendering (DIBR) technique [23]. The idea of DIBR is to synthesize the virtual views by using the texture and depth information at another viewpoint. Firstly, the image points at the original viewpoint are reprojected into the 3D world by using its associated depth data. Then, these 3D points are projected into the image plane at the virtual target viewpoint. With the MVD format and the DIBR technology, only a limited number of original views and their corresponding depth maps are needed to be captured, stored and transmitted.

Recently, great efforts have been put into invertigating DIBR technology[47] [15] [48] [82] [56]. DIBR is not only useful in FVV [69], but also a promising solution for

[2] Source: `http://global.canon/en/news/2017/20170921.html`



(a) A light field image captured by a camera at a corner

(b) A light field image synthesized by the DIBR

Figure 4: Example of light field image rendered by DIBR in [40]

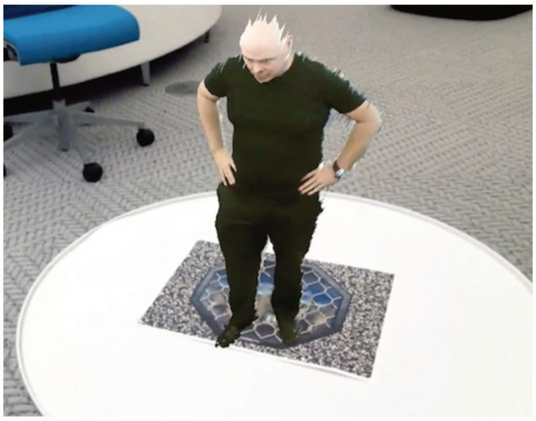


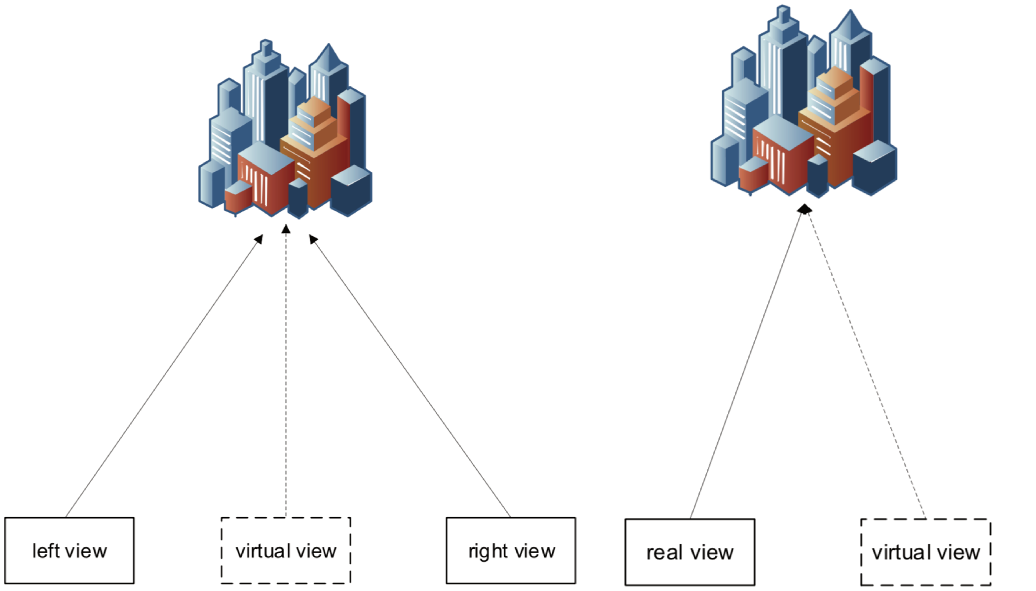
Figure 5: Augmented Reality Screen Capture.[3]

synthesizing virtual views in many other recent popular immersive multimedia applications, such as VR [70], Augmented Reality (AR) [2] and Light Field (LF) multi-view videos [73], etc. For example, DIBR has already been used in a light field compression scheme where only very sparse samples (four views at the corners) of light field views are transmitted while the others are synthesized (cf. Fig. 4). This new scheme significantly outperformed High Efficiency Video Coding (HEVC) inter coding for the tested LF images [40]. Another example concerns 360-degree and volumetric videos: two developing areas pointing to how video will evolve as VR/AR technology becomes the mainstream [95]. Current 360-degree videos allow viewers to look around in all directions, but only at the shooting location: they do not take into account the translation (changes in position) of the head. To achieve more immersive QoE, some companies propose to use DIBR to synthesize the non-captured views when users move from the physical camera's position, as proposed in Technicolor's volumetric video streaming demonstration [34]. One similar approach[3] is proposed (see Fig. 5), where typical DIBR artifacts appear (around the contours) if users try to look at an object on the floor hidden behind the person. In the social and embodiment VR media applications, where a VR media designed for 360-degree videos mixed with real-time objects for multiple users, an eye-contact technique based on the DIBR [34] can provide the users the viewpoint according to their eye positions, which gives the users a better interactive QoE.

As shown in Fig. 6 , there are two main kinds of DIBR view synthesis algorithms: the single view based synthesis and the interview synthesis: for the single view based synthesis, we use the one base view to synthesis another; for the interview synthesis, we use two base views to render the middle one. As it is well known that during the video compression or transmission, we should consider the quality of decompressed or decoded videos. When it comes to view synthesis, how about the quality of synthesized views?

Although DIBR has a great potential, current DIBR algorithms may introduce some new types of distortions which are quite different from those caused by image compression. Most compression methods can cause specific distortions [114], eg. blur [119], blockiness [109] and ringing [24]. These distortions are often scattered over the whole image, while the DIBR-synthesized artifacts (caused by distorted depth map and imperfect view synthesis method) mostly occur in the disoccluded areas. Since most of the

[3] Source: `https://developer.att.com/blog/shape-future-of-video`



(a) interview synthesis (b) single view based synthesis

Figure 6: DIBR view synthesis

commonly used 2D objective quality metrics are initially designed to assess common compression distortions, they may fail in assessing the quality of DIBR-synthesized images [9, 30].

The contributions presented in this thesis aim at the improvement of quality assessment of DIBR-synthesized views. Although several efforts have been made towards the quality assessment of DIBR-synthesized views, it still can not achieve a satisfactory result. The first contribution of this thesis is the proposition of two NR metrics and two FR metrics for DIBR-synthesized views. Secondly, during the study of metrics, there is no existing adequate database for the quality assessment of DIBR-synthesized views, a new DIBR image database is thus proposed.

Organization of the thesis

The thesis is organized as follows. The general introduction of this thesis is given in this chapter, then Chapter 1 introduces the principle of 3D view synthesis and analyzes its special distortions. Next, the major contribution of this thesis are organized in the following three chapters. Chapter 2 presents the proposed two FR metrics; chapter 3 presents two morphological based no-reference quality metrics and chapter 4 is dedicated to our new DIBR database. Finally, the conclusion and perspectives and presented in Chapter 5.

Chapter 1 Overview of 3D view synthesis and distortion analysis

This chapter presents the basis of 3D view synthesis. Firstly, the 3D vision perception of Human Vision System (HVS) and the most commonly used content representation formats of 3D videos are introduced in detail. Then, we discuss the principle of the DIBR view synthesis and its particular influence on the quality of synthesized views. After that, some state-of-the-art objective metrics dedicated to assess the quality of DIBR-synthesized views are briefly investigated.

Chapter 2 Proposed FR image quality assessment metrics

In recent years, several FR metrics have been proposed to assess the quality of DIBR-synthesized views. However, none of them could get a satisfactory performance. In this chapter, two FR image quality metrics focusing on the geometric distortion and dis-occlusion in the DIBR-synthesized views are presented. Firstly, we observe that there exist a large amount of object shift and geometric distortions in the DIBR-synthesized views which the traditional 2D quality metrics may fail to assess. This is due to the fact that the HVS is more sensitive to local artifacts compared to the global object shift. Thus, a Speeded Up Robust Features (SURF) + Random sample consensus (RANSAC) approach is used to compensate the global object shift roughly. Then, in the first FR quality model SC-DM, we use a dis-occlusion mask to weight the difference between the synthesized image and the warped reference image. It can be used with any pixel based quality metric. On the other hand, in the second FR quality metric SC-IQA, a multi-resolution block matching method is proposed to precisely compensate the object shift and penalize the local geometric distortion as well. In addition, a saliency map is also used to weight the final distortions. The experimental results show that the SC-DM quality model improves the performance of Peak signal-to-noise ratio (PSNR) and Structural similarity index (SSIM) significantly, the SC-IQA metric outperforms all the tested state-of-the-art quality metrics, including the FR, NR and Reduced-reference (RR) metrics.

Chapter 3 Proposed NR image quality assessment metrics

As introduced in Chapter 1, in some 3D applications, such as FVV, even with a very large number of views, it is still not possible to cover all the arbitrary viewpoints of a particular scene. That is to say, there is no reference image for the synthesized view in some circumstance. In this case, the NR quality metrics are in great need. In this chapter, we propose two completely NR image quality metrics (NIQSV and NIQSV+) for DIBR-synthesized views. These two metrics are both based on the assumption that the image with good quality is supposed to present sharp and regular edges, smooth values inside the object and large discontinuities at the object borders. Such images are insensitive to opening and closing morphological operations while some artifacts such as blurry regions around the object edges and crumbling in the synthesized views are sensitive to such morphological operations. Based on this property, NIQSV uses a pair of opening and closing operations to measure these distortions. As an extended version of NIQSV, NIQSV+ improves its performance by estimating the other two types of distortions: black holes and stretching. Then, the experimental results on the IRCCyN/IVC DIBR image database show that the proposed NIQSV metric outperforms the traditional 2D metrics and approaches the performance of DIBR dedicated FR. The proposed NIQSV+ achieves the second place in all the DIBR dedicated metrics including FR and NR at the time of experiment, and there is no significant difference between the DIBR-synthesized view dedicated metrics.

Chapter 4 A benchmark of DIBR-Synthesized View Quality Assessment Metrics

This chapter presents a new DIBR-synthesized image database focusing on the distortions only produced by different DIBR synthesis methods. Firstly, we give an overview of the existing DIBR databases, and then the main contribution of this work is presented. Compared to the existing databases, we test more and newer algorithms including inter-view synthesis and single view based synthesis, which excludes the “old fashioned” distortions.

Chapter 5 Conclusion and Perspectives

This chapter concludes the contribution of this thesis: two NR quality metrics, two FR quality metrics dedicated to DIBR-synthesized views and a new publicly accessible DIBR synthesized image database. Besides, some directions for future work are also given in this chapter.

CHAPTER 1

OVERVIEW OF 3D VIEW SYNTHESIS AND DISTORTION ANALYSIS

The 3D vision perception allows the human to perceive the world in three dimensions. In this chapter, we firstly introduce the principle of 3D vision perception and the most widely used 3D content representation; then the DIBR view synthesis method and its particular distortions are analyzed; next we give an overview of the existing state-of-the-art DIBR-synthesized view dedicated quality assessment metrics; this chapter is concluded in the last section.

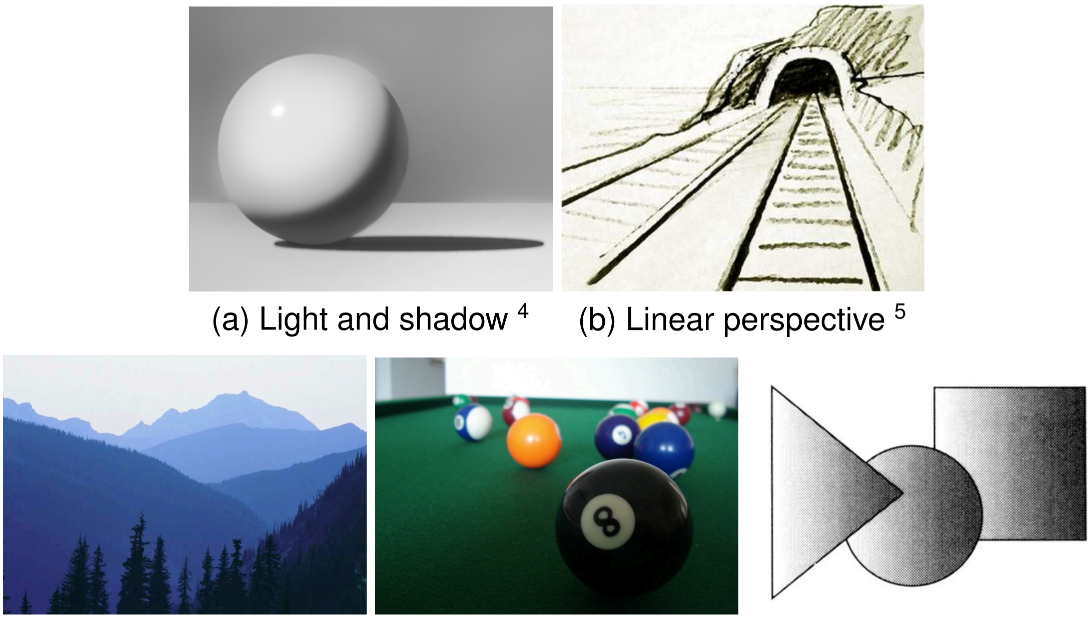
1.1 3D vision perception

3D vision comes from the perception of depth. Depth perception arises from a variety of depth cues: the monocular cues and the binocular ones. The monocular cues can be extracted from a single two-dimensional image and be observed with just one eye, while the binocular cues are based on the receipt of sensory information in three dimensions from both eyes.

1.1.1 Monocular cues

By using only one image, we can still perceive the depth information. Here lists some monocular cues that we can use to help perceive the depth:

- Light and shadow: the objects near the light source are more brightly illuminated than those who are far away from the light source. Thus, the way light falls on objects and the amount of shading present can be an important monocular cue to tell the depth. cf. Fig 1.1 (a).



(a) Light and shadow[4] (b) Linear perspective[5]

(c) Aerial perspective[6] (d) Relative size[7] (e) Occlusion[8]

Figure 1.1: Examples of depth information from monocular cues.

- Linear perspective: it refers to the fact that the parallel lines appears to converge in the distance. cf. Fig 1.1 (b).

- Aerial perspective: due to the light scattering of atmosphere, the object which are farther away always have lower luminance contrast and color saturation, they seem to be blurred or slightly hazy. cf. Fig 1.1 (c).

- Relative size: if there are two objects with the same known size, relative size cues can provide information about the relative depth of the two objects. The depth difference can be perceived based on the fact that the larger objects appear closer and the smaller objects appear farther away. cf. Fig 1.1 (d).

- Occlusion: it occurs when there is overlapping of objects. The occluded objects is considered farther away. cf. Fig 1.1 (e).

[4] Source: `http://www.paintdrawpaint.com/2016/01/drawing-basics-basics-of-light-and.html`

[5] Source: `http://draw23.com/perspective`

[6] Source: `https://www.pinterest.com/pin/280841726744268329/`

[7] Source: `http://poshaylapsych15.blogspot.com/2014/11/monocular-cue-relative-size.html`

[8] Source: `https://www.psywww.com/intropsych/ch04-senses/depth-perception.html`

- Motion parallax: it is a dynamic depth cue provided by motion. If we focus our eye on a certain object and then move our our head to the right side, the objects which are closer to us will move to the left while the objects farther away will move to the right. It gives the relative information about the distance to an object and expresses how close an object is from the certain one. It can come from moving objects or moving observers.

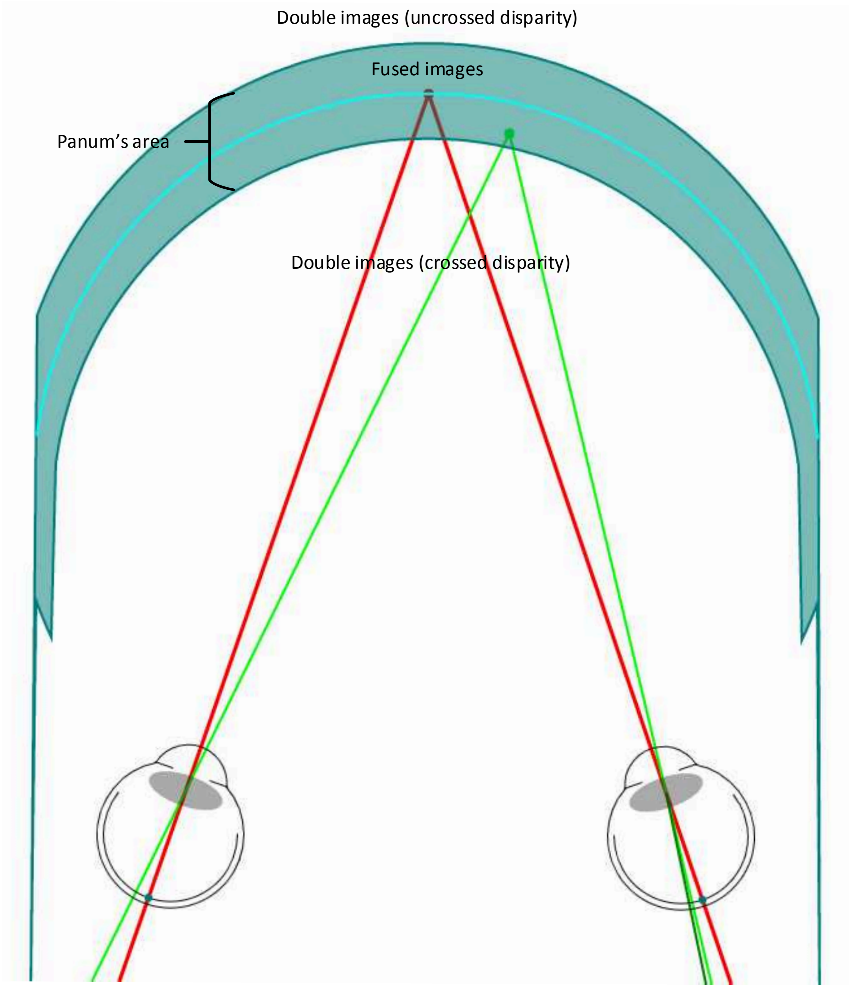

Double images (uncrossed disparity)
Fused images
Panum's area
Double images (crossed disparity)

Figure 1.2: Panum's Area of Fusion. [9]

[9] Source: `https://isle.hanover.edu/Ch07DepthSize/Ch07Panum.html`

1.1.2 Binocular cues

Generally, we perceive depth because of the two slightly different retinal images from the two eyes, which is known as binocular disparity. The binocular disparity is present because the human eyes are horizontally separated by 6.3 cm (on average) which provides each eye with a different viewpoint of the real scene.[33] The retinal points from one eye's view are matched to corresponding points in the other eye's view, the point to point disparity variation provides the information of the relative distances of objects, the depth structure of objects, as well as the depth structure of surroundings. This perception of depth is referred to as "stereoscopic depth". [31]

Our nervous system fuses the two retinal images to a single image. As shown in Fig 1.2, the points which have no disparity construct a line which is called “horopter” (when the eyes converge on the object). The points whose disparities are within a certain interval can be fused to a single experienced image. This small region around the “horopter” is named Panum's fusional area. The points lying out of this Panum's fusional area will not be fused and double images will be caused. The points which are closer or farther produce double images of crossed or uncrossed disparity respectively. Although the complex mechanisms of human vision system are not clearly understood, the use of monocular and binocular cues already enable creators and artists to impress the public by illusion of depth. In the next section, we will introduce some widely used 3D content representation formats.

1.2 3D content representation

Several applications have been developed to provide the observer immersive perception of 3D visual scene, such as 3D-TV, MVV and FVV. In these applications, at least two images from small different viewpoints are required for the stereoscopic visualization. Especially, for MVV and FVV, the images from more viewpoints are needed.

To provide the input for such 3D applications, various 3D content representation formats have been proposed for different applications.

The stereoscopic video could be the simplest type of 3D data format. It contains a pair of conventional 2D videos for the left and the right eye respectively. One of the main drawback of this format is that the baseline distance depends on the video capture configuration, the parallax is limited. It can only be used for stereoscopic display, for the

FVV or MVV, it is not suitable.

In order to improve the stereoscopic video format, many 3D content representation types have been proposed. Multi-view video format can be recognized as a simple extension of stereoscopic video since it consists of more views than the stereoscopic ones. However, the data size increases greatly as the number of views increases. Another approach is the video-plus-depth format, which is widely known as 2D+Z. This data presentation consists of only one 2D video sequence and its associated depth map, and the stereo pair can be synthesized based on Depth-Image-Based-Rendering (DIBR) [23]. Owing to the imperfect of DIBR technique, the baseline range of the synthesized stereopair videos are quiet limited.

To overcome these disadvantages, Multiview-Video-Plus-Depth (MVD) [57] has been proposed. It consists of multiple texture views from different viewpoints and their associated depth maps. Using the MVD format, only a limited number of original views and their corresponding depth map need to be encoded and transmitted. The additional virtual views on the receiver side can be synthesized from the decoded views and depths based on DIBR technique. In addition, efficient compression methods: Multiview High-Efficiency-Video-Coding (MV-HEVC) [93] and 3D High-Efficiency-Video-Coding (3D-HEVC) [93] have been developed by the Joint Collaborative Team on 3D Video Coding Extension Development (JCT-3V) to compress the MVD data. Benefiting from the efficient coding methods and DIBR, MVD has become one of the most popular 3D presentation methods.

1.3 3D view synthesis principles

Depth-image-based-rendering (DIBR) is one of the most popular methods to generate virtual views. As shown in Fig. 1.3, DIBR uses the captured texture images and depth maps to generate the novel view as if there were a virtual camera.

The DIBR is a process of generating novel views of a scene from original texture images and associated depth information. Firstly, the original texture image is re-projected into 3D world aided by the associated per-pixel depth data; then these 3D space points are projected into the image plane of a new virtual view position. This concatenation of 2D-to-3D re-projection and the subsequent 3D-to-2D projection is usually called 3D image warping in the Computer Graphics (CG) literature. As shown in Fig. 1.4, the DIBR view synthesis can be divided into two parts: 3D warping and hole

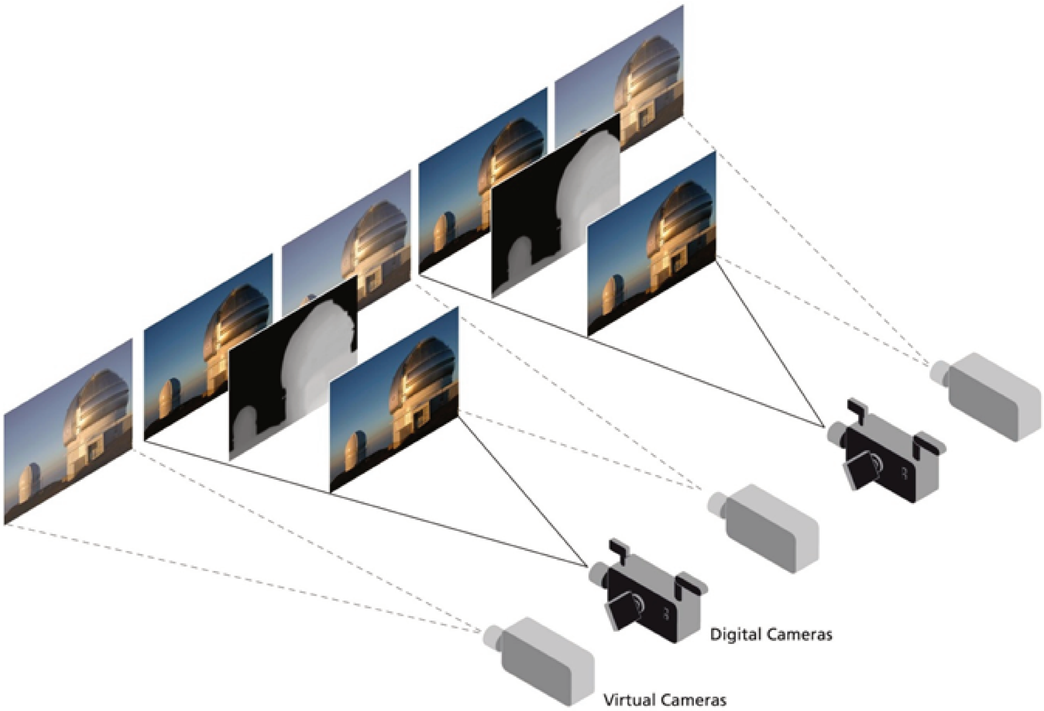


Figure 1.3: Example of view synthesis

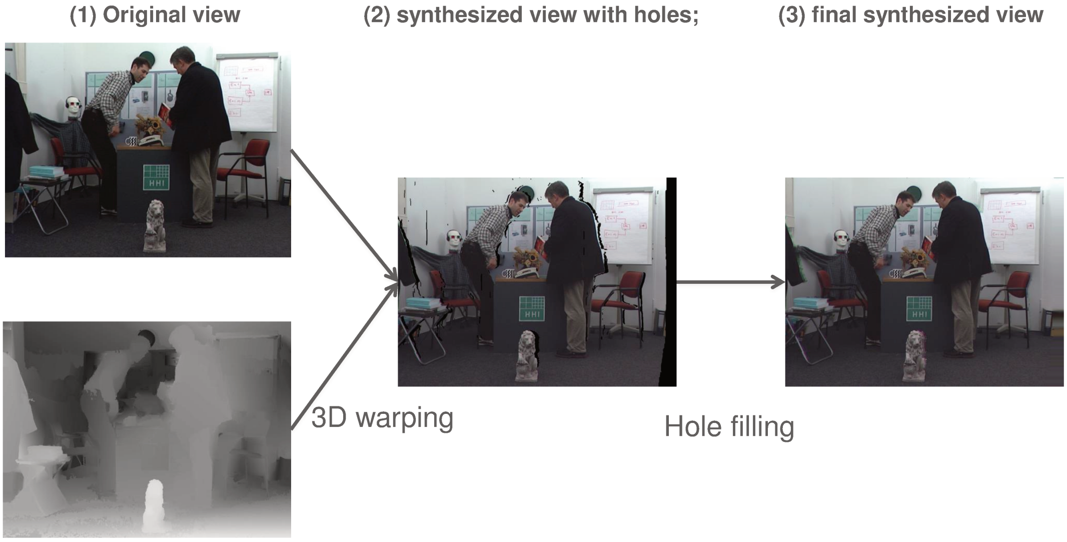


Figure 1.4: Procedure of DIBR

filling. During the 3D warping procedure, the pixels in the original view are warped to there corresponding positions in the target view. Owing to the changing of viewpoint, some objects which are invisible in the original view may become visible in the target one, which is called dis-occlusion and causes black holes in the synthesized view. Then, the second step is to fill the black holes, as shown in Fig. 1.4. The holes could be filled by typical image inpainting algorithms. Most of the image inpainting algorithms use the pixels around the "black holes" to search the similar regions in the same image, and then use this similar region to fill the "black holes". The distortions caused by the DIBR view synthesis method will be discussed in the next section.

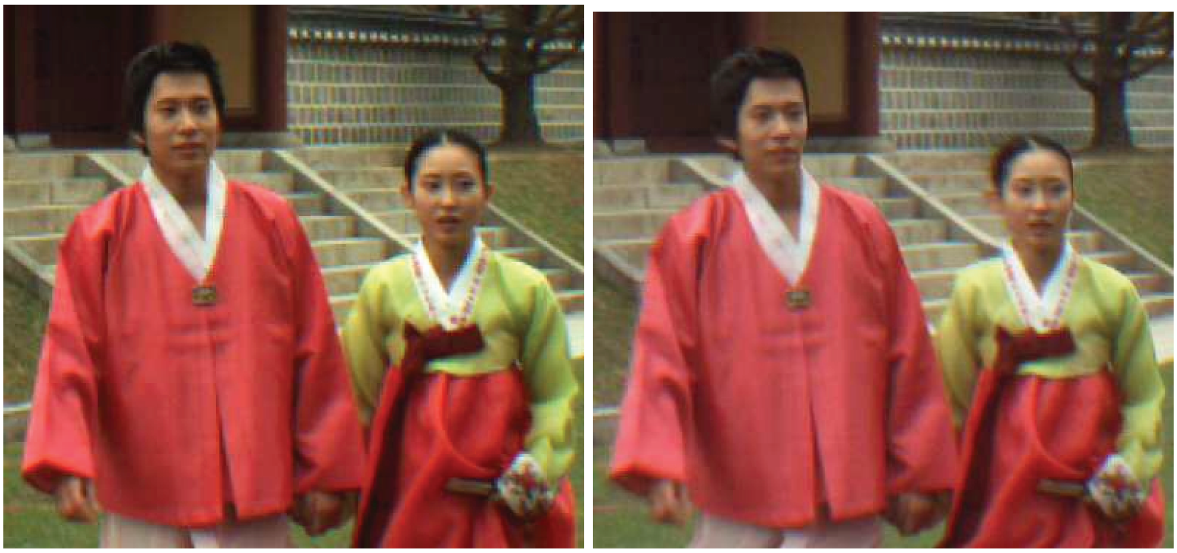
1.4 View synthesis distortion analysis

Unlike the distortions induced by video compression, the distortions of DIBR-synthesized views mainly result from inaccurate depth map and view synthesis algorithms. They mainly happen in the dis-occluded regions which are non-visible in the previous view but become visible in the target view.

Owing to the lack of texture information of the dis-occluded regions, many inpainting methods have been developed to reduce these synthesis artifacts. But these processes may generate some new types of distortions and these distortions from image inpainting are specific and depending on the algorithm.

Another source of distortion may come from the depth map. Firstly, during the 3D warping process, a large number of errors can also be induced by the numerical rounding operations of pixel positions since the corresponding pixel positions in the target viewpoint may be not an integer. In addition, the lossy compression of the depth map may also lead to artifacts, for example, the blocking effects, blurry or quantization errors in the depth map can induce the pixels in the original viewpoint to be rendered to the wrong positions in the target viewpoint, which will lead to object shifting or crumbling the synthesized texture image. These distortions can be summarized as follows.

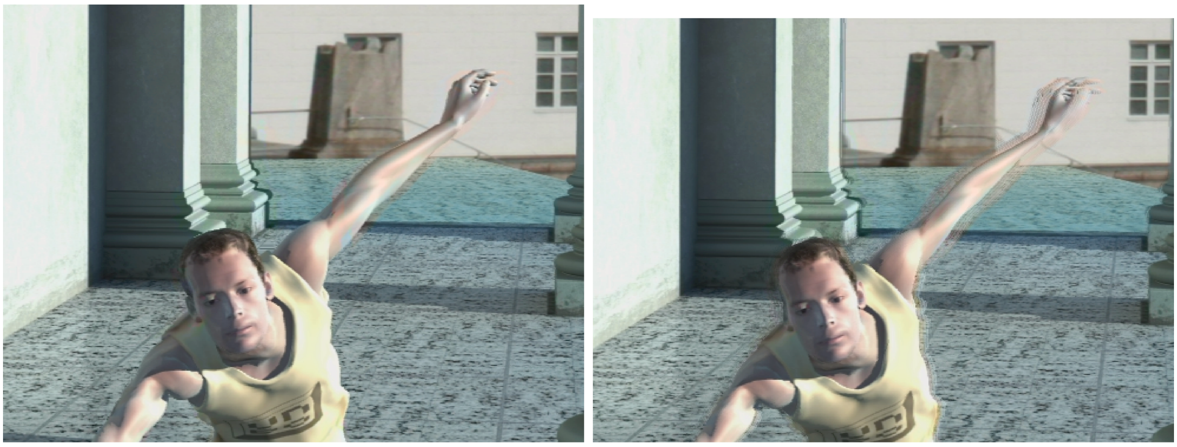
- Object shifting: object regions can be slightly shifted or resized in the synthesized view due to incorrect depth information produced by high-frequency noise or depth pre-processings including low-passing filters and depth encoding methods to smooth the object borders. As shown in Fig. 1.5, the right borders of the character's faces are slightly modified.



(a) Reference view

(b) Synthesized view

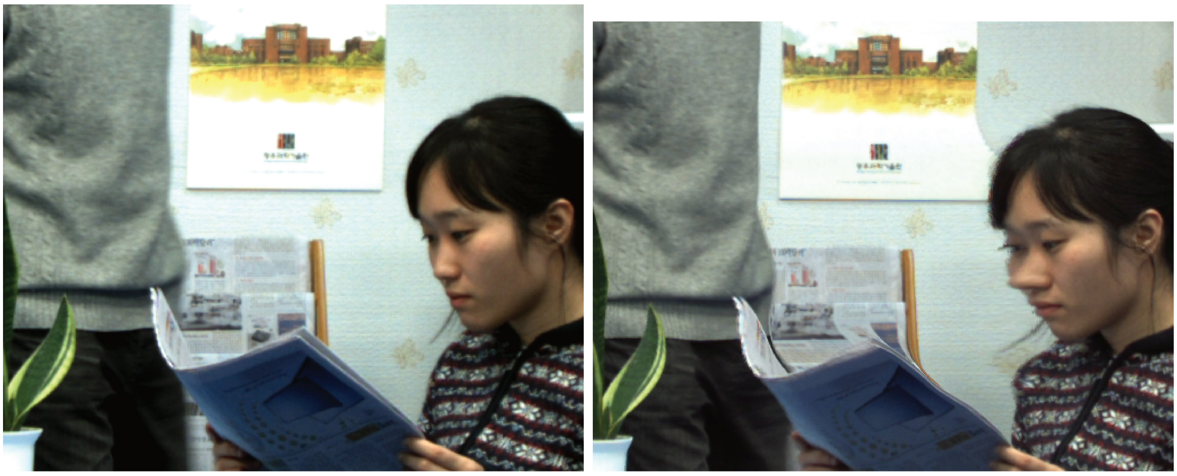
Figure 1.5: Object shifting caused by depth low-passing filter, the right borders of the character's faces are slightly modified.



(a) Reference image

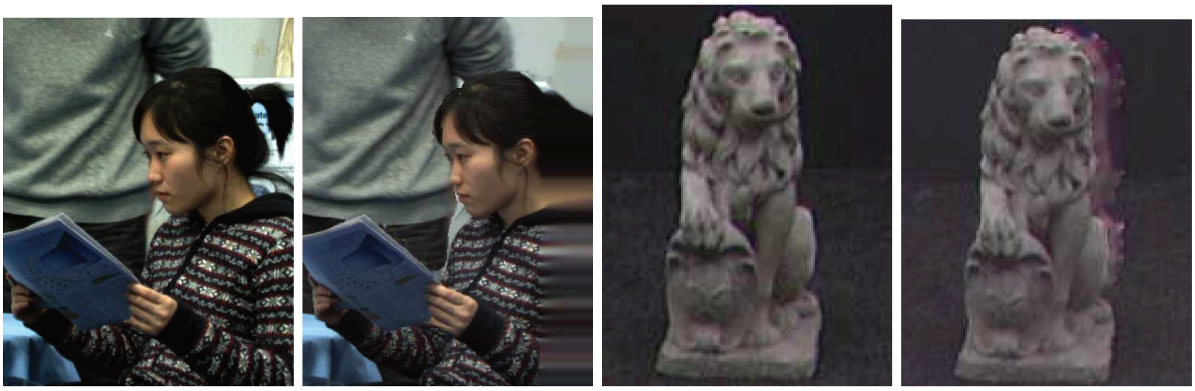
(b) Synthesized image

Figure 1.6: Example of ghosting effect distortions caused by Gaussian noised depth map. There exists a ghosting effect around the “hand”.



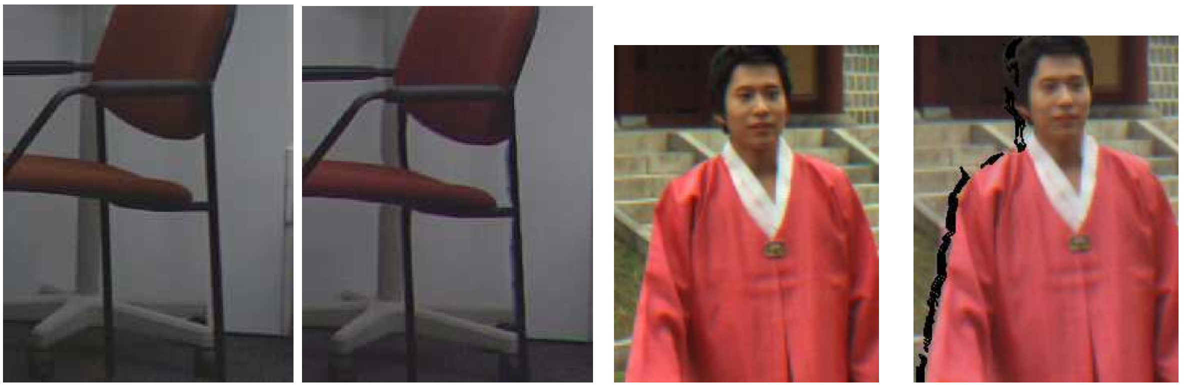
(a) Reference image (b) Synthesized image

Figure 1.7: Example of object warping distortion caused by imperfect image inpainting method, the “newspaper” and the “girl’s nose” are extremely stretched, and their shapes are greatly changed.



(a) Reference view (b) Synthesized view (c) Reference view (d) Synthesized view

Figure 1.8: Synthesis distortions caused by imperfect image inpainting method. (a) and (b) stretching; (c) and (d) blurry regions. The “girl’s ” hair and clothes in image (b); and the right border of the sculpture in (d).



(a) Reference view (b) Synthesized view (c) Reference view (d) Synthesized view

Figure 1.9: Synthesis distortions, (a) and (b) crumbling; (c) and (d) black holes. The Crumbling in the "chair arms" in image (b) and black hole around the "man's arm".

- Ghosting effect: Fig. 1.6 also shows an example of ghosting effect. In this image, the depth map is distorted by a Gaussian noise. It could be observed that the object ("hand" in the image) of the two base views are rendered to the wrong pixel positions, causing a ghost in the synthesized view.

- Object warping: Fig. 1.7 gives an example of object warping caused by imperfect image impainting method. In this image, the Tela's image impainting method which in introduced in [94] is used. It could be observed that the "newspaper" and the "girl's nose" are extremely stretched, and their shapes are greatly changed.

- Slight geometric distortion: a large number of slight geometric distortions can be induced by the numerical rounding operations of pixel positions and depth inaccuracy. This may not be noticeable to the human eye, but it could be overestimated by pixel-based metrics.

- Stretching: stretching errors mainly happen at the left or right side of the image where in-painting methods may fail to reconstruct. This type of distortion is shown in Fig. 1.8 (b).

- Blurry regions: some blurry regions may be produced by in-painting methods used to fill the disoccluded regions. They can be noticed at the foreground and background transitions. As shown in Fig 1.8 (d), blurry regions can be perceptible around the sculpture.

- Crumbling: the object edge may seem distorted in the synthesized view. This is mainly cased by quantization artifacts in depth data around strong discontinuities which appear like erosion as shown in Fig. 1.9 (c). It typically occurs when applying wavelet-based compression on depth maps.

- Flickering: pixels could be projected into an erroneous location due to the random errors happening in depth sequence. These pixel positions suffer slight changes. This is a temporal artifact which occurs in synthesized video sequences.

- Black holes: when the disoccluded areas are left unfilled after the 3D warping, they induce a strong visual artifact in the synthesized view, as shown in Fig. 1.9 (d).

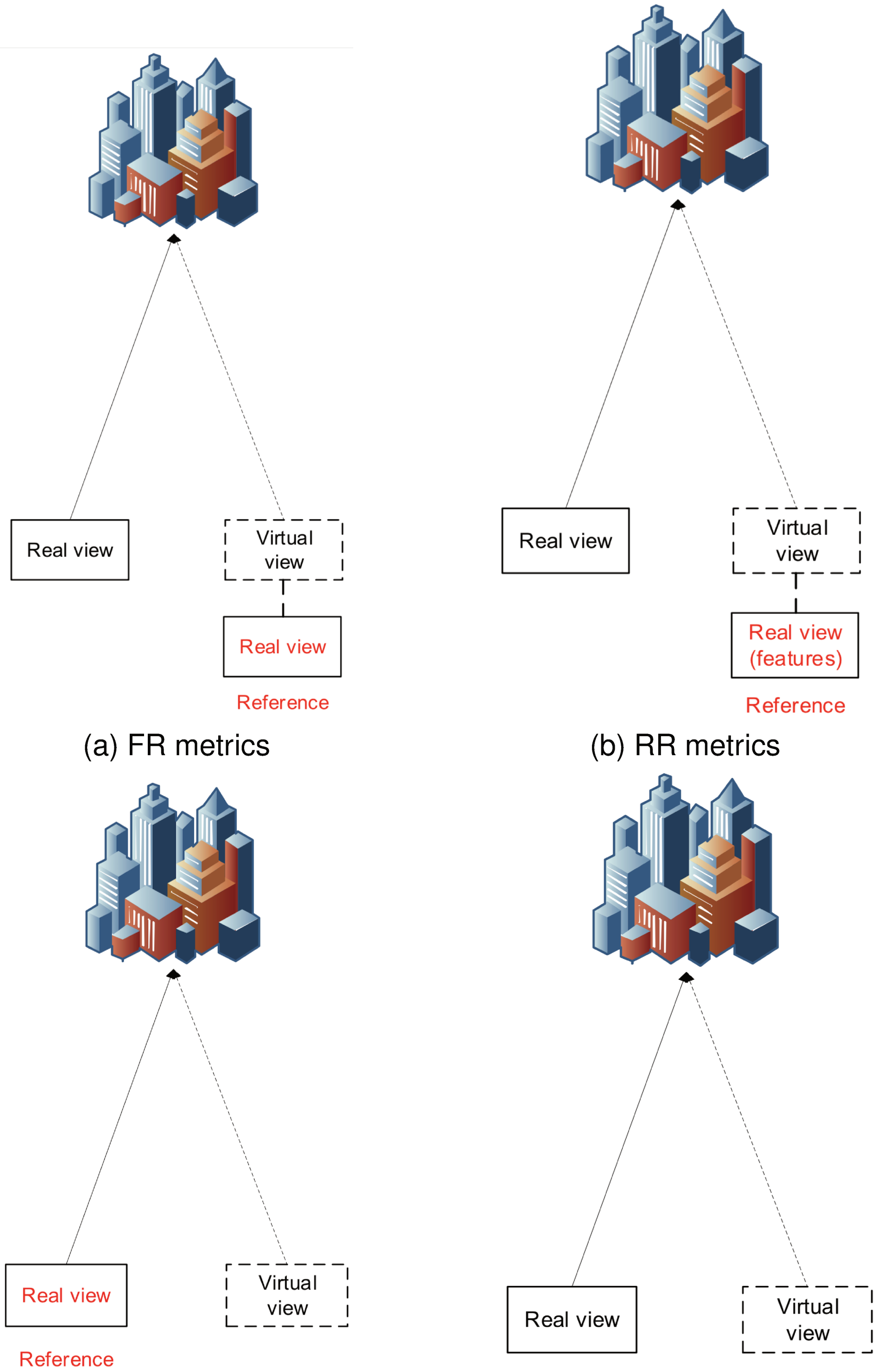
As mentioned above, since the synthesis artifacts are far different from the compression artifacts, they are geometric distortions : the position of the pixel is modified, whereas classic metrics measure modification of pixel value, but assume no position modification. The traditional 2D IQA metrics do not work well on these types of artifacts.

1.5 Overview of the state-of-the-art quality assessment methods of DIBR-synthesized views

The objective Image quality assessment (IQA) metric can be divided into three categories relying on the amount of reference information used in the metric: FR metrics assess the quality of the distorted image by using the original undistorted image; RR metrics use reduced information extracted from the original undistorted image; NR or blind metrics can evaluate the quality of the distorted images without access to the original undistorted image.

In recent years, several methods have been proposed to assess the quality of DIBR-synthesized views, including FR, NR, RR and side view based FR metrics. Especially, the side view based FR metrics use the image at the original viewpoint as reference image. As shown in Fig. 1.10 .

Table 1.1 classify the metric based on the approaches. Most of the metrics (VSQA, MP-PSNR, MW-PSNR, EM-IQA and CT-IQA) evaluate the quality of synthesized views by considering the contour or gradient degradation between the synthesized and the



(a) FR metrics

(b) RR metrics

(c) Side view based FR metrics

(d) NR metrics

Figure 1.10: Categories of quality assessment metrics for DIBR-synthesized views.

reference images which is one of the most annoying characterization of geometric distortion. While some metrics (DSQM, 3DSwIM) calculate the quality score by comparing the extracted perceptual features between the synthesized and the reference images. Especially, the APT metric uses a local image description model to reconstruct the image, and the use the reconstructed error to assess the quality of the synthesized views. These metric are introduced as follows.

Table 1.1: Overview of the existing metrics. The features in the first column indicate frequency domain feature (FF), contour/gradient (C/G), JND, Multi-scale decomposition (MSD), local image description (LID), depth estimation (DE), dis-occlusion Region (DR), Reblurring (RB), ML (Machine Learning).

Approach Metric		FF	C/G	JND	MSD	LID	DE	DR	RB	ML
FR	VSQA	-	✓	-	-	-	-	-	-	-
	3DSwIM	✓	-	-	-	-	-	-	-	-
	MW-PSNR	-	✓	-	✓	-	-	-	-	-
	MP-PSNR	-	✓	-	✓	-	-	-	-	-
	CT-IQA -	✓	-	-	-	-	-	-	-	
	EM-IQA	-	✓	-	-	-	-	-	-	-
	PSPTNR	-	-	✓	-	-	-	-	-	-
	VQA-SIAT	-	✓	-	-	-	-	-	-	-
	3VQM	-	-	-	-	-	✓	-	-	-
RR	MP-PSNRr	-	✓	-	✓	-	-	-	-	-
	MW-PSNRr	-	✓	-	✓	-	-	-	-	-
SV FR	LOGS	-	-	-	-	-	-	✓	✓	-
	DSQM	✓	-	-	-	-	-	-	-	-
NR	APT	-	-	-	-	✓	-	-	-	-
	CSC-NRM	-	-	-	-	-	-	-	-	✓

The FR quality metrics:

- **VSQA** (View Synthesis Quality Assessment):

  The main feature of the VSQA [16] proposed by Conze et al. is to apply three weighting maps on the SSIM distortion map [112]. The purpose of these three weighting maps is to characterize the image complexity in terms of textures, diversity of gradient orientations and presence of high contrast. These three weighting maps are calculated from three visibility maps.

Firstly, the texture-based visibility map is obtained as the mean of the gradient magnitude of a certain window $N_t \times N_t$. As shown in Eq. 1.1

$$V_t(i,j) = \frac{1}{N_t^2} \times \sum_{l=i-[\frac{N_t}{2}]}^{i+[\frac{N_t}{2}]} \sum_{k=j-[\frac{N_t}{2}]}^{j+[\frac{N_t}{2}]} w_{l,k} grad[l,k] \tag{1.1}$$

where, $V_t(i,j)$ indicates the texture-based visibility map, $grad$[] means the corresponding gradient map with a $Sobel$ operator, $w_{l,k}$ is the gaussian weighting function and N_t is the window size.

Except the gradient magnitude, the gradient orientation is also taken into consideration to form an orientation-based visibility map, due to the fact that the HVS is sensitive to local orientation features. The orientation-based visibility map is calculated as follows:

$$V_o(i,j) = \min_q [\frac{1}{N_o^2} \times \sum_{l=i-[\frac{N_o}{2}]}^{i+[\frac{N_o}{2}]} \sum_{k=j-[\frac{N_o}{2}]}^{j+[\frac{N_o}{2}]} w_{l,k}[Lum(l,k) - Lum(i,j)]]] \tag{1.2}$$

where Lum represents the luminance value, θ is the gradient orientation, which is obtained as Eq. 1.3:

$$\theta(i,j) = tan^{-1}(\frac{G_y(i,j)}{G_x(i,j)}) + \frac{\pi}{2} \tag{1.3}$$

where G_x and G_y indicates the horizontal and vertical gradient respectively.

The third visibility map corresponds to the image contrast, since the distortions at the pixels with significant luminance difference to their neighborhood are much more annoying. This contrast-based visibility map is computed by:

$$V_c(i,j) = \min_q [\frac{1}{N_c^2} \times \sum_{l=i-[\frac{N_c}{2}]}^{i+[\frac{N_c}{2}]} \sum_{k=j-[\frac{N_c}{2}]}^{j+[\frac{N_c}{2}]} w_{l,k} min[(\theta(l,k) - \theta_q)^2, (\theta(l,k) + \pi - \theta_q)^2]] \tag{1.4}$$

In order to get the overall quality score, a threshold is used on the distortion map to count the number of remained pixels after thresholding. The final results gives the quality score. The threshold is set according to the following equation.

$$th = min_{VSQA} + p \times \frac{max_{VSQA} - min_{VSQA}}{100} \quad (1.5)$$

where p is a positive parameter, max_{VSQA} and min_{VSQA} represent the maximum and minimum distortion values respectively.

The final weighting maps $W_x(i,j)$ $x \in (t,o,c)$ are calculated by rescaling the value of the associated texture-based visibility maps to $[0,2]$. cf. Eq. 1.6

$$W_x(i,j) = \frac{2}{max(V_x) - min(V_x)} \times V_x(i,j) - \frac{min(V_x)}{max(V_x) - min(V_x)}, x \in (t,o,c) \quad (1.6)$$

where t, o, c represent the texture, orientation and contrast respectively. This VSQA metrtic is reported that this method approaches a gain of 17.8% over SSIM in correlation with subjective measurements.

- **3DSwIM** (3D Synthesized view Image Quality Metric):

  In the 3DSwIM [4], the quality score is obtained by comparing the statistical features of the reference and the synthesized view in the wavelet transformed domain. The Kolmogorov–Smirnov distance between the histograms of the reference and distorted images in wavelet domain is measured as the quality index, as shown in Eq. 1.7.

$$d_b = max(|F_{o_b} - F_{s_b}|) \quad (1.7)$$

where F_{o_b} and F_{s_b} represent the distribution function of the real view and the synthesized view respectively, d_b indicates the distortion of each image block.

Since the synthesis distortions mainly occur in the horizontal direction, only horizontal features are used. In addition, a registration step is performed as a preprocessing to align the content of the synthesized view and the reference one.

Then, the overall image distortion is extracted by normalizing the distortion of all the blocks, cf. Eq. 1.8.

$$d = \frac{1}{D_0} \sum_{b=1}^{B} d_b \quad (1.8)$$

where D_0 is a normalization constant. The image quality score is obtained by the following formula:

$$s = \frac{1}{1+d} \tag{1.9}$$

In addition a skin detection is used to weight the blocks containing human body parts. So the final image quality score is computed as follows:

$$s = \frac{1}{1 + d = \frac{1}{D_0}\sum_{b=1}^{B} w_{skin} max(|F_{o_b} - F_{s_b}|)} \tag{1.10}$$

- **MP-PSNR and MW-PSNR** (Morphological Pyramid PSNR and Morphological Wavelet PSNR):

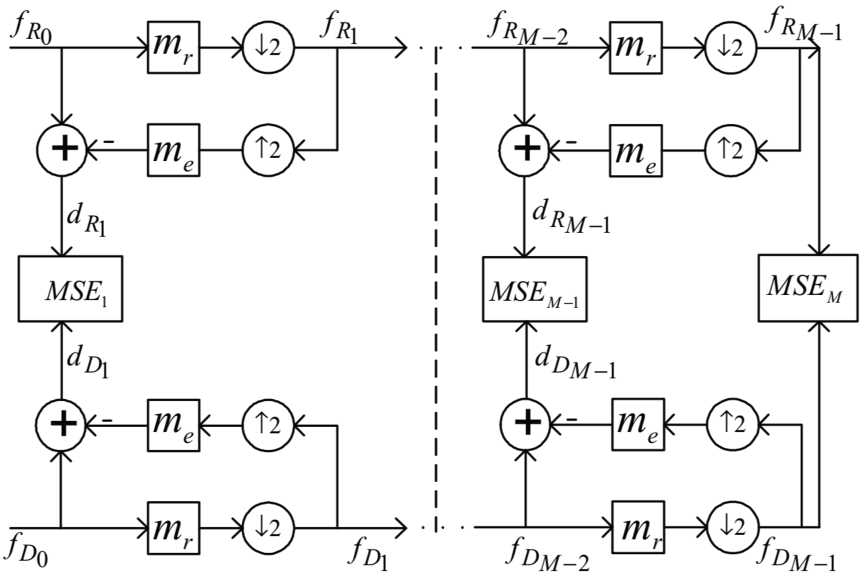


Figure 1.11: Scheme of MP-PSNR, m_r and m_e represent the reduce and expand filter respectively. [9]

Sandic-Stankovic et al. proposed the MP-PSNR [79] based on multi-scaled pyramid decomposition using morphological filters. The basic erosion and dilation operation used in MP-PSNR are calculated as maximum and minimum in the

[9] Source: figure from [79]

neighbourhood define by the structure element, as shown in the following equation:

$$D : dilation_{SE}(f)(x) = max_{y \in SE} f(x - y) \tag{1.11}$$

$$E : erosion_{SE}(f)(x) = min_{y \in SE} f(x + y) \tag{1.12}$$

where f is a grayscale image and SE is binary structure element.

Then, they use the Mean Square Error (MSE) between the reference and synthesized image in all pyramids' sub-bands to quantifier the distortion. As shown in Fig. 1.11,during the decomposition, the dilation is used as reduce operation and the erosion is used as expand operation. Finally, the overall quality is calculated by averaging the MSE in all the sub-bands and transform it to PSNR.

$$MSE_j = \frac{1}{N_j \times K_j} \sum_{k=1}^{K_j} \sum_{n=1}^{N_j} (x_j(k,n) - y_j(k,n))^2 \tag{1.13}$$

where x_j and y_j denote the reference and distorted image at scale j with size $K_j \times N_j$. Multi-scale mean squared error MP-MSE is calculated as weighted product of MSE of all pyramid images with equal weights:

$$MP - MSE = \prod_{j=1}^{M} [MSE_j]^{\beta_j} \tag{1.14}$$

where β_j indicates a weight value parameter. Finally, the MP-PSNR score is calculated as:

$$MP - PSNR = 10 \times log_{10}(\frac{R^2}{MP - MSE}) \tag{1.15}$$

where R is the maximum dynamic range of the image.

The MW-PSNR has been proposed and applied on free viewpoint videos [74] and still images [78] by the same authors. The idea of MW-PSNR is similar to MP-PSNR, the difference is that the MW-PSNR uses morphological wavelet filters for decomposition instead of dilation and erosion. Then a multi-scale wavelet mean square error (MW-MSE) is calculated as the average MSE of all sub-bands and finally the MW-PSNR is calculated from it.

- **CT-IQA** (Context Tree based Image Quality Assessment):

  A variable-length context tree based image quality assessment [51] proposed by

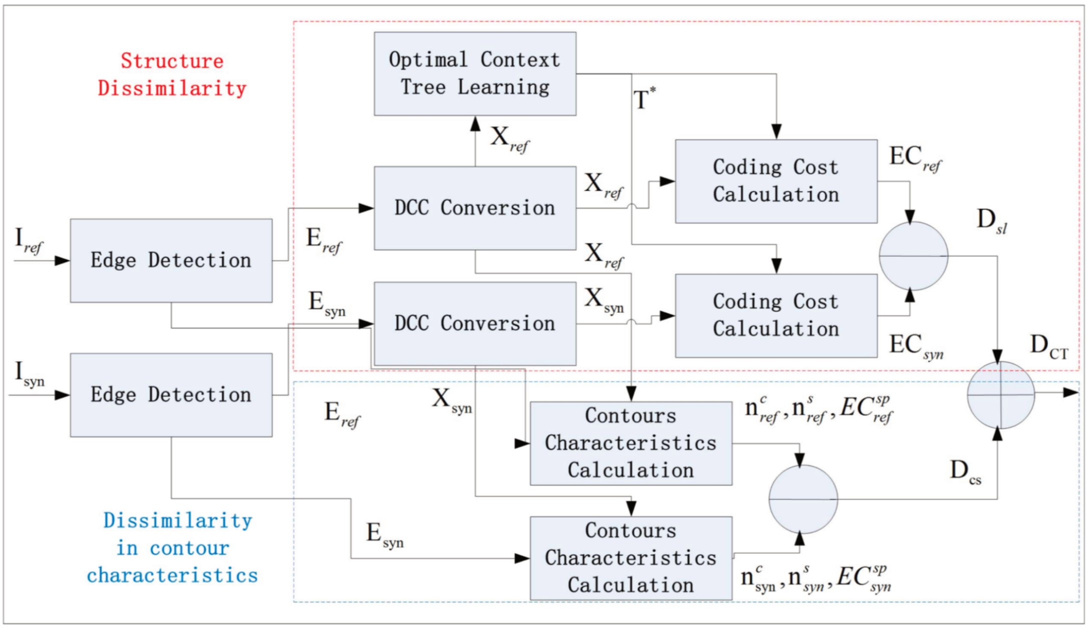


Figure 1.12: The chatflow of CT-IQA. [10]

Ling et al. aims to quantify the overall structure dissimilarity and dissimilarities in various contour characteristics by using context tree based contour coding. As shown in Fig. 1.12, firstly, the contour of reference and the synthesized images are converted to differential chain code (DCC), an optimal context tree is learned from the DCC in the reference image. Then, the overall structural dissimilarity is calculated by subtracting the encoding cost of DCC in synthesized image and the reference image:

$$D_{sl} = |\frac{EC_{ref}}{X_{ref}} - \frac{EC_{syn}}{X_{syn}}| \tag{1.16}$$

where EC_{ref} and EC_{syn} represent the encoding cost of the reference and synthesized views respectively, X_{ref} and X_{syn} indicate the set of DCC strings in the reference and synthesized views respectively.

In addition, the overall dissimilarity in contour characteristics is also obtained by measuring the difference of total contour number, total contour start information and total number of symbols between the reference and synthesized image:

$$D_{cs} = |sum(n^c_{ref}, n^s_{ref}, EC^{sp}_{ref}) - sum(n^c_{syn}, n^s_{syn}, EC^{sp}_{syn}) \tag{1.17}$$

[10] Source: figure from [51]

where the n_{ref}^c, n_{ref}^s represent the number of contours in the reference and synthesized views.

The final quality score is calculated by combining the above two measurements:

$$D_{CT} = \alpha \times D_{ls} + \beta \times D_{cs} \tag{1.18}$$

where α and β are two weighing parameters which are set to 0.9 and 0.1 respectively.

- **EM-IQA** (Elastic Metric based Image Quality Assessment):

  In [50], Ling et al. also proposed an elastic metric based image quality assessment metric by quantifying the deformation of curves in the local distortion regions. It firstly select the distortion region based on interest point matching, then extract the contour of both synthesized and reference image. Finally, the distortions in the synthesized image are measured by the distance between the extracted contours.

  The contour is firstly defined as:

$$c : D \rightarrow (x, y) \in \mathbb{R}^n, \tag{1.19}$$

  where (x, y) represents the coordinate of each point in the curve, $D = [0, 1]$. Then it can be further represented by a square-roor velocity (SRV) function defined by $q : D \rightarrow \in \mathbb{R}^n$:

$$q(t) \equiv F(\dot{c}(t)) = \dot{c}(t)/\sqrt{\|\dot{c}(t)\|} \tag{1.20}$$

  where $\|.\|$ indicates the Euclidean 2-norm in $\mathbb{R}^n$ and $\dot{c} = \frac{dc}{dt}$. The curve is obtained by:

$$c(t) = \int_0^t q(s)\|q(s)\|ds \tag{1.21}$$

  In order to quantify the deformation of the curves of synthesized views, the distortion of synthesized view is measured by measuring the distance of the curves in the synthesized and reference images:

$$D_{EM} = \int_D < \frac{1}{2}e^{\frac{1}{2}\phi}u_1\theta + e^{\frac{1}{2}\phi}\upsilon_1, \frac{1}{2}e^{\frac{1}{2}\phi}u_2\theta + e^{\frac{1}{2}\phi}\upsilon_2 >= \int_D (\frac{1}{4}e^{\theta}u_1u_2 + e^{\theta} < \upsilon_1, \upsilon_2 >)dt \tag{1.22}$$

where ϕ and θ are defined as follows:

$$\phi(t) = ln(\|\dot{c}(t)\|) \tag{1.23}$$

$$\theta = \dot{c}(t)/\|\dot{c}(t)\| \tag{1.24}$$

Finally, the overall quality score is calculated by summing out the distortions of all the curves:

$$S_{EM} = \sum D_{EM}(c^i_{ori}, c^j_{syn}) \tag{1.25}$$

where $(c^i_{ori}, c^j_{syn}) \in (C_{ori}, C_{syn})$

- **PSPTNR** (Peak Signal to Perceptible Temporal Noise Ratio):

  Zhao et al. [117] PSPTNR metric to measure the perceptual temporal noise of the synthesized sequence. The temporal noise is defined as the the difference between inter-frame change in the processed sequence and that in the reference sequence:

  $$TN_{i,n} = ((P_{i,n} - P_{i,n-1}) - (R_{i,n} - R_{i,n} - 1)^2, \tag{1.26}$$

  where TN indicates the temporal noise, P and R represent the distorted and reference sequence respectively.

  Then the per-pixel Perceptible Temporal Noise (PTN) is obtained by filtering it with a Just Noticeable Distortion (JND) model.

  $$PTN_{i,n} = \begin{cases} TN_{i,n} & abs(P_{i,n} - P_{i,n-1} \geq JND_{S-T,i,n} \& i \in staticarea \\ 0 & else \end{cases} \tag{1.27}$$

  where $abs()$ is the absolute function, JND is the used Just Noticeable Distortion model.

  The final perceptible temporal noise is calculated by measuring the filtered noise in the region of higher motion in the synthesized view:

  $$PSPTNR_n = 10 \times log_{10}(\frac{K \times 255^2}{\sum_{i=1}^{K} PTN_{i,n}}) \tag{1.28}$$

  $$PSPTNR = \frac{1}{N-1} \sum_{n=2}^{N} PSPTNR_n \tag{1.29}$$

- **VQA-SIAT** (Video Quality Assessment metric of Shenzhen Institute of Advanced Technology):

  In [52]. Liu et al. proposed a synthesized Video Quality Assessment (VQA) framework to measure flickering which is the most annoying temporal distortion of synthesized sequences. Firstly, A temporal gradient vector is defined as follows:

  $$\vec{\bigtriangledown} I_{x,y,i}^{temporal} = I(x,y,i) - I(x',y',i-1), \tag{1.30}$$

  where (x',y') is the coordinate in frame i corresponding to (x,y) along the motion trajectory in previous frame $i-1$.

  A Spatio-Temporal tube (S-T tube) and a Quality Assessment Group of Pictures (GA-GoP) were generated to measure the annoying variations of pixel luminance which act as flickering distortion in the synthesized sequences. The flickering distortion can be measured along the motion trajectory as below:

  $$DF_{x_i,y_i} = \sqrt{\frac{\sum_{n=i-N+1}^{i+N} \phi(x_n,y_n,n) \times \bigtriangledown(x_n,y_n,n)}{2N}}, \tag{1.31}$$

  where $2N+1$ is the length of QA-GoP. ϕ and $\bigtriangledown$ are used to detect the potential sensible flicker distortion and the strength of the distortion. Which are defined as follows:

  $$\phi(x_n,y_n,n) = \begin{cases} & \vec{\bigtriangledown} I_{x,y,n}^{temporal} \times \vec{\bigtriangledown} \widetilde{I}_{x,y,n}^{temporal} \leq 0 \\ 1 & and\ \vec{\bigtriangledown} \widetilde{I}_{x,y,n}^{temporal} \neq 0 \\ & and\ |I(x,y,n) - \widetilde{I}(x,y,n)| > \mu \\ 0 & else \end{cases} \tag{1.32}$$

  $$\bigtriangledown\ (x,y,n) = \left(\frac{\vec{\bigtriangledown} I_{x,y,n}^{temporal} - \vec{\bigtriangledown} \widetilde{I}_{x,y,n}^{temporal}}{|\vec{\bigtriangledown} I_{x,y,n}^{temporal} + C|}\right)^2 \tag{1.33}$$

  where I and $\widetilde{I}$ represent the reference and synthesized view images. When the temporal gradient direction is different, $\vec{\bigtriangledown} I_{x,y,n}^{temporal} \times \vec{\bigtriangledown} \widetilde{I}_{x,y,n}^{temporal} \leq 0$ and $\vec{\bigtriangledown} \widetilde{I}_{x,y,n}^{temporal} \neq 0$, there may be a potential flicker distortion. The $\vec{\bigtriangledown} I_{x,y,n}^{temporal} - \vec{\bigtriangledown} \widetilde{I}_{x,y,n}^{temporal}$ indicates the magnitude of the temporal gradient distortion. Then, the flicker distortion in the S-T tube, GoP and sequence are calculated:

  $$DF^{tube} = \frac{\sum_{x=1}^{w} \sum_{y=1}^{h} DF_{x,y}}{w \times h}, \tag{1.34}$$

$$DF^{GoP} = \frac{1}{N_W} \sum_{k \in W} DF_k^{tube}, \tag{1.35}$$

$$DF^{Seq} = \frac{1}{K} \sum_{m=0}^{K} DF_m^{GoP} \tag{1.36}$$

where w and h represent the wide and height of the image respectively. W denotes the worst 1% DF^{tube} in the QA-GoP, K is the number of QA-GoPs in the sequence.

Besides, the distortions induced by video compression were measured by detecting the activity in GA-GoP and S-T tube. Firstly, the spatial gradient value of the pixel can be computed as:

$$\vec{\bigtriangledown} I_{x,y,n}^{spatial} = \sqrt{|\vec{\bigtriangledown} I_{x,y,n}^{spatial\_h}|^2 + |\vec{\bigtriangledown} I_{x,y,n}^{spatial\_v}|^2} \tag{1.37}$$

where $\vec{\bigtriangledown} I_{x,y,n}^{spatial\_h}$ and $\vec{\bigtriangledown} I_{x,y,n}^{spatial\_v}$ represent the horizontal and vertical gradient vector respectively. Then, the spatial-temporal activity is computed by measuring the mean and standard deviation value of the spatial gradient:

$$\overline{\bigtriangledown I_{tube}^{spatial}} = \frac{\sum_{n=i-N}^{i+N} \sum_{y=y_n}^{y_n+h} \sum_{x=x_n}^{x_n+w} \vec{\bigtriangledown} I_{x,y,n}^{spatial\_h}}{w \times h \times (2N+1)} \tag{1.38}$$

$$\sigma_{tube} = \sqrt{\frac{\sum_{n=i-N}^{i+N} \sum_{y=y_n}^{y_n+h} \sum_{x=x_n}^{x_n+w} (\bigtriangledown I_{tube}^{spatial} - \overline{\bigtriangledown I_{tube}^{spatial}})^2}{w \times h \times (2N+1)}} \tag{1.39}$$

$$\Gamma_{tube} = \begin{cases} \sigma_{tube} & \sigma_{tube} > \epsilon \\ \epsilon & else \end{cases} \tag{1.40}$$

where ϵ is a threshold. After that, the overall distortion of spatial-temporal activity of S-T tube, GoP and sequence are computed as:

$$DA^{tube} = |log_{10}(\frac{\widetilde{\Gamma}_{tube}}{\Gamma_{tube}})| \tag{1.41}$$

$$DA^{GoP} = \frac{1}{N_W} \sum_{k \in W} DA_k^{tube} \tag{1.42}$$

$$DA^{Seq} = \frac{1}{K} \sum_{m=0} DA_m^{GoP} \tag{1.43}$$

where, similarly, W denotes the set with the worst 1% DA^{tube} in the GoP.

The overall quality scores were obtained by integrating these two features:

$$D = DA \times log_{10}(1 + DF). \tag{1.44}$$

The VQA-SIAT metric takes flickering, the most annoying temporal distortion of DIBR-synthesized videos and the distortion of compression into consideration. However, the other geometric distortion which could happen in the synthesized view have been ignored.

- **3VQM** (3D Vision- based Quality Measure):

  Solh et al. proposed a full reference metric 3VQM [84] to evaluate synthesized view distortions by deriving an "ideal" depth map from the virtual synthesized view and the reference view at a different viewpoint. "Ideal depth" is the depth map that would generate the distortion-free image given the same reference image and DIBR parameters. The calculation of "ideal" depth map can be divided into three steps. Firstly, the horizontal coordinate vector $\bar{X}_v$ of the synthesized view is calculated from the reference view $\bar{X}_r$ by using the 3D warping function:

  $$\bar{X}_v = \bar{X}_r + s\frac{F \times B}{\bar{Z}} + h, \tag{1.45}$$

  where F, B and Z represent the focal length, baseline distance and depth value respectively. s is a direction symbol, $s = -1$ when the synthesized view is to the left while $s = 1$ when the synthesized view is to the right.

  Similarly, the horizontal coordinate vector $\bar{X}_o$ can be calculated by the "ideal" depth map:

  $$\bar{X}_o = \bar{X}_r + s\frac{F \times B}{\bar{Z}_{ideal}} + h \tag{1.46}$$

  After that, the "ideal" depth map can be computed as follows:

  $$\bar{Z}_{ideal} = \frac{sFB}{(\bar{X}_o - \bar{X}_v) + s\frac{FB}{\bar{Z}}} \tag{1.47}$$

  It has been shown that the sum of squared differences (SSD) of the original video frame and its horizontal translations is linear. Hence, a small horizontal shift can be estimated in terms of intensity. As a result, the “ideal” depth map can be estimated as follows:

$$\bar{Z}_{ideal} = \frac{sFB}{\alpha(\bar{I}_o - \bar{I}_v) + s\frac{FB}{\bar{Z}}} \tag{1.48}$$

where $\bar{I}_o$ and $\bar{I}_v)$ represent the synthesized and the original view respectively.

Then, three distortion measures, spatial outliers, temporal outliers and temporal inconsistency are then integrated into a final quality measurement.

$$SO = STD(\triangle Z) \tag{1.49}$$

$$TO = STD(\triangle Z_{t+1} + \triangle Z_t) \tag{1.50}$$

$$TI = STD(Z_{t+1} + Z_t) \tag{1.51}$$

where SO, TO and TI denotes the spatial outliers, temporal outliers and temporal inconsistencies respectively. $\triangle Z$ is the difference between the “ideal” depth map and the distorted depth map. t is the frame number.

The final $3VQM$ quality score is obtained from pooling the above three measurements:

$$3VQM = K(1 - SO(SO \cap TO)^a)(1 - TI)^b(1 - TO)^c \tag{1.52}$$

where K is a constant.

This metric assume that the horizontal shift of the synthesized view and the original view is small, but when the baseline distance increase, this metric would not work well.

RR quality metrics:

- MP-PSNRr and MW-PSNRr (reduced version of MP-PSNR and MW-PSNR) In [76], the same authors also proposed the reduced version of MP-PSNR, and MW-PSNR. They use only detail images from higher decomposition scales to measure the difference between the synthesized image and the reference image. The reduced version achieved significant improvement than the original FR metrics with lower computational complexity.

The Side view based FR quality metrics:

- **LOGS** (Local Geometric and global Sharpness):

  Li et al. proposed a side view based FR metric for DIBR-synthesized views by measuring local geometric distortions in disoccluded regions and global sharpness (LOGS) [46]. This metric consists of three parts. Firstly, it detects the dis-occlusion region by using SIFT-flow based warping, the absolute difference between the synthesized view I_{syn} and the warped reference view I_{ref}^{w} is computed followed by an additional threshold is used to exclude the small value in the difference map.

  Then the distortion size and strength in the local dis-occlusion regions are combined to obtain the overall local geometric distortion. The distortion size is measured by the number of pixels in the dis-occluded regions; the distortion strength is defined as the mean value of the region in the whole difference map M:

  $$e_i = \frac{1}{s_i} \sum_{(x,y)\in\Omega_i} M(x,y) \tag{1.53}$$

  where e_i is an obtained distortion strength, Ω_i is a dis-occluded region, s_i is a size of dis-occluded regions, M is the difference map.

  After that, the quality of dis-occluded regions is calculated as follows:

  $$Q_R = \frac{\sum_{i=1}^{K} w_i e_i}{\sum_{i=1}^{K} w_i} \tag{1.54}$$

  where w_i represents the weight of which dis-occluded region, which is defined below:

  $$w_i = log_2(1 + \frac{Rank_i}{K}), i = 1, 2, ..., K \tag{1.55}$$

  where $Rank_i$ denotes the ranking index of the ith dis-occluded regions.

  The next part is to measure the global sharpness by using a reblurring-based method. In this part, the synthesized image is firstly blurred by a Gaussian smoothing filter:

  $$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{1.56}$$

  where σ is denotes the standard deviation. The sharpness is measured block by block, both the synthesized image and its reblurred version are divided into

Z blocks. The sharpness of each block is calculated by its textural complexity, which is represented by their variance σ^2. Then, the overall sharpness score is computed by averaging the textural distance of all blocks:

$$Q_S = \frac{\sum_{i=1}^{Z} \sqrt{|\sigma_{1i}^2 - \sigma_{2i}^2|}}{Z} \tag{1.57}$$

where σ_{1i}^2 and σ_{2i}^2 represent the standard deviations of the ith blocks in the synthesized image and the reblurred image.

Finally, the local geometric distortion and the global sharpness are pooled to generate the final quality score:

$$Q = \frac{Q_S^{\alpha}}{Q_R^{\beta} + c} \tag{1.58}$$

where α and β are two parameters used to balance the relative contributions of local dis-occluded regions and global sharpness, and c is a small constant to stable the division.

- **DSQM** (DIBR-Synthesized image Quality Metric):

  Proposed by Farid et al. in [21]. A block matching is firstly used to match the content in the reference image and the synthesized image by using the following normalized cross-correlation:

$$\gamma(x,y) = \frac{\sum_{\mu,\nu} (p(\mu,\nu) \times I_s(x+\mu, y+\nu))}{\sqrt{\sum_{\mu,\nu} p(\mu,\nu)^2 \times \sum_{\mu,\nu} I_\nu(x+\mu, y+\nu)^2}} \tag{1.59}$$

  where I_s is the synthesized image, p indicates the patch in the reference image. μ, ν denote the index of searching block and x, y represent the index of pixel in the reference image.

  Then the difference of Phase congruency (PC) in these two matched blocks is used to measure the quality of the block in the synthesized image, which is defined as follows:

$$PC(x) = \max_{\bar{\phi}(x) \in [0, 2\pi]} \frac{\sum_n A_n cos(\phi_n(x) - \bar{\phi}(x))}{\sum_n A_n} \tag{1.60}$$

  where A_n and $\phi_n(x)$ represent the amplitude and the local phase of the n-th

Fourier component at position x respective. Beside, the phase congruency is equivalent to the ratio of energy and the sum of the Fourier amplitudes:

$$PC(x) = \frac{E(x)}{\sum_n A_n + \epsilon} \tag{1.61}$$

where $E(x) = \sqrt{F^2(x) + H^2(x)}$, $F(x)$ is the signal with its DC component removed, $H(x)$ is the Hilbert transform of $F(x)$, ϵ is a small constant to keep the equation stable.

The quality score of each block is obtained as the absolute difference between the mean values of the phase congruency maps of the matched blocks in the synthesized and reference image:

$$Q_i = |\mu(PC_{si} - PC_r i)| \tag{1.62}$$

where $\mu()$ represents the mean value of the corresponding phase congruency map, the PC_{si} and PC_{ri} indicate the PC map of the matched blocks in the synthesized and reference image. The final image quality is obtained by averaging the quality score of all the blocks:

$$DSQM = \frac{1}{k}\sum_{i=1}^{k} Q_i \tag{1.63}$$

The NR quality metrics:

- **APT** (Autoregression Plus Thresholding):

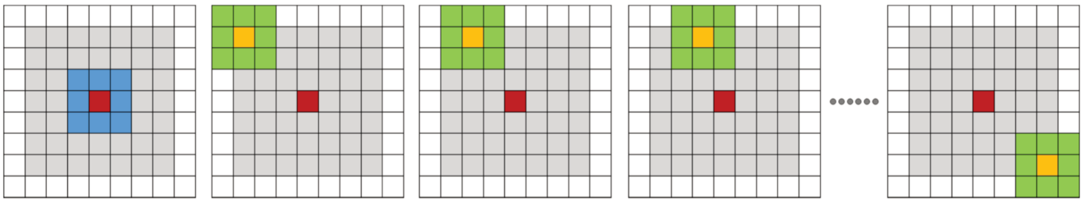


Figure 1.13: Local autoregression.

A no-reference metric proposed by Gu et al. in [29]. It uses an autoregression based local image descriptor to reconstruct the synthesized image. The local aoturegression model is used for image analysis. For a certain image pixel, we denote its location index as i and its value as x_i. the relationship between this pixel and its neighborhood can be expressed as the following formula:

$$x_i = \omega_\theta(x_i)s + d_i \tag{1.64}$$

where $\omega_\theta(x_i)$ denotes a vector which includes its neighborhood pixels in the $\sqrt{\theta+1} \times \sqrt{\theta+1}$ patch. As shown in Fig. 1.13, the red refers to the current pixel to be processed, it and its neighbourhood 8 pixels constitute the local $\sqrt{\theta+1} \times \sqrt{\theta+1}$ patch. In this metric, θ is set to 8.

Then, the least square method is used to estimate the reliable vector of autoregression parameters.

$$\hat{s} = arg \min_s \|x - X_s\|_2 \tag{1.65}$$

where x and X denotes the processed pixels and their surrounding pixels respectively. After that they solve this linear system via least square method and infer the best estimation of the vector of autoregression parameters to be

$$\hat{s} = (X^T X)^{-1} X^T x \tag{1.66}$$

Note that the size of X and x is set to 48, which indicates that the relationship that is built upon the current pixel is assumed to exist for adjacent 48 pixels in the local 7×7 patch.

The reconstructed error between the input synthesized image and the predicted image is used to detect the geometric distortions. In addition, a saliency weighting and a thresholding are added to obtain the final quality measurement. Due to its computational complexity, it owns a high time cost.

- **CSC-NRM** (Convolutional Sparse Coding-based No-Reference Model): In [49], Ling et al. proposed a NR machined learning based metric for DIBR-synthesized views, which focuses on the non-uniform distortions. Firstly, a set of convolutional kernels are learned by using the improved fast convolutional sparse coding (CSC) algorithms. Then, the convolutional sparse coding (CSC) based features of the DIBR-synthesized view images are extracted, from which the final quality score is obtained via support vector regression (SVR).

  The convolutional sparse coding (CSC), which is defined in Eq. 1.67, is used to learn the local interactions via convolution operations.

$$min_{d,z}\frac{1}{2}\|y-\sum_{k=1}^{K}d_k \times z_k\|+\beta\sum_{k=1}^{K}\|z_k\|_1, s.t.\|d_k\|\leq 1 \tag{1.67}$$

where y is the observed samples, z_k indicates the sparse feature maps and d_k is the convolution kernel. K is the number of convolution kernels and β is a positive scalar parameter. The convolutional kernels learning step is done by solving the following optimization problem:

$$min_{\{d_k\}}=\frac{1}{2}\|y-\sum_{k=1}^{K}d_k \times Z_k\|, s.t.\|d_k\|_2^2\leq 1 \tag{1.68}$$

where Z_i denotes the convolution operators with feature maps z_i. During the feature extraction step, all the kernels are set and the features are extracted by minimizing over the feature maps:

$$min_z\frac{1}{2}\|y-D_z\|^2+\beta\|z\|_1 \tag{1.69}$$

where D represents the operators consists of convolutions with K kernels, Z is the feature map vector. Then, the CSC vector is obtained as follows:

$$v_{csc}=(f_{act(1),...,f_{act}(Z^K)}) \tag{1.70}$$

where f_{act} is the activated function which is defined as:

$$f_{act}(Z^K)=\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\mathbf{1}(Z^K(i,j)>\epsilon)}{M\times N} \tag{1.71}$$

where **1**(x) is logical function which equals 1 when x is true, and ϵ is a threshold. Finally, the obtained CSC vectors are used to predict the overall image quality score via support vector regression (SVR).

We list 9 FR, 2 RR, 2 side view based FR and 2 NR metrics above. Among which PSPTNR, VQA-SIAT and 3VQM take the temporal distortion into consideration, they are dedicated in Video Quality Assessment (VQA); the other metrics are Image Quality Assessment (IQA) metrics.

1.6 Performance evaluation of objective quality metrics

The reliability of objective metrics can be evaluated by their correlation with subjective test scores with respect to three aspects of their ability to estimate subjective assessment of video quality: prediction accuracy, prediction monotonicity and prediction consistency. The prediction accuracy is the ability of the model to predict the observers' ratings (subjective scores) with a minimum error on average. The Pearson Linear Correlation Coefficients (PLCC) and Root-Mean-Square-Error (RMSE) are the two commonly used metrics to measure the average error. The higher PLCC value indicates a lower average error; on the contrary, a lower RMSE value refers to a lower average error. Ideally, the objective scores should be totally monotonic in their relationship to the subjective values. This monotonicity can be measured by the Spearman Rank-Order Correlation Coefficients (SROCC) between the objective and subjective scores. A higher SROCC values relates to a better monotonicity. The PLCC, RMSE and SROCC are defined as follows:

$$PLCC(X,Y) = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{Y})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}} \tag{1.72}$$

$$RMSE(X,Y) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(X_i - Y_i)^2} \tag{1.73}$$

$$SROCC(X,Y) = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{1.74}$$

where d_i indicates the difference of ranking of X and Y.

The prediction consistency relates to the objective quality model's ability to provide consistently accurate predictions for all the tested sequences and not fail excessively for a subset of sequences. It can be measured by the number of outlier points (defined as having an error greater than a given threshold such as one confidence interval) as a fraction of the total number of points (Outlier Ratio). Which can be presented as follows:

$$OutlierRatio = \frac{TotalNoOutliers}{N} \tag{1.75}$$

where $TotalNoOutliers$ indicates the total number of outlier points and N is the number of samples. An outlier point is a point which yields the following condition:

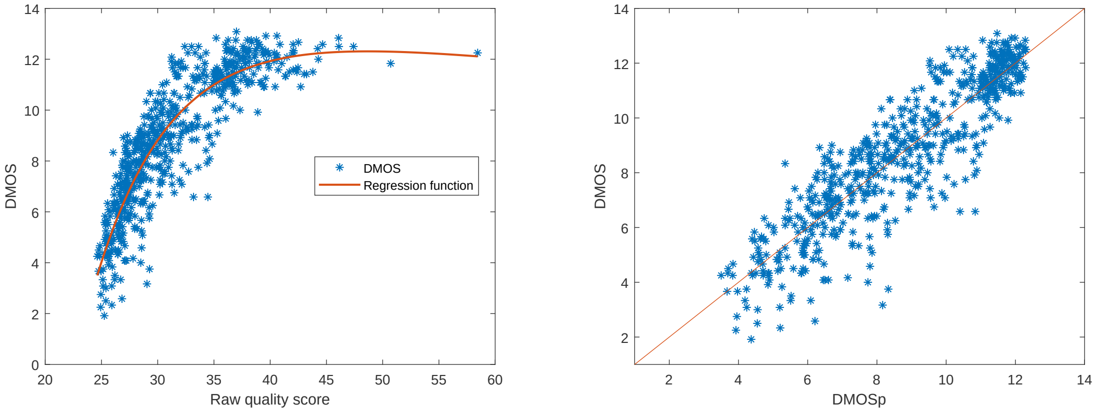
$$|Perror(i)| > K2 \times \frac{\gamma(DMOS(i))}{\sqrt{Nsub}} \tag{1.76}$$

where $K2$ equals 1.96 for 95% confidence interval for a Gaussian distribution, $Nsub$ is the number of viewers. $Perror()$ represents the difference between the objective and subjective scores. A lower Outlier Ratio indicates that the objective quality model holds a better consistency.

Normally, the Differential Mean Opinion Score (DMOS, the MOS difference between the distorted and the reference image) is used as subjective scores. The calculation of DMOS in different databases may be different, it depends on the subjective experiment. For instance, in IRCCyN/IVC DIBR image database [37], [12], the DMOS was calculated following the equation:

$$DMOS = MOS_{syn} - MOS_{ref} + 5 \tag{1.77}$$

where MOS_{syn} and MOS_{ref} represent the MOSs of the synthesized image and the reference image (as introduced in [27]). The $+5$ in this equation is to avoid the negative scores which do not appear in this case.



(a) Before regression

(b) After regression

Figure 1.14: Example relationship between DMOS and objective quality scores

Before calculating PLCC, RMSE, SROCC and Outlier Ratio, the objective scores need to be fitted to the so-called predicted DMOS, which are noted as $DMOS_p$ to remove the nonlinearties due to the subjective rating processing and to facilitate comparison fo the models in a common analysis space, as shown in Fig. 1.14 (a). The Video Quality Expert Group (VQEG) Phase I FR-TV [27] has recommended several nonlinear mapping functions for this fitting step, here two widely used regression functions are listed:

- The 4-parameter cubic polynomial function:

$$DMOS_p = \beta_1 \cdot scores^3 + \beta_2 \cdot score^2 + \beta_3 \cdot score + \beta_4 \tag{1.78}$$

- The 5-parameter logistic function:

$$DMOS_p = \beta_1 \cdot (0.5 - \frac{1}{(1 + exp(\beta_2 \cdot (score - \beta_3)))}) + \beta_4 \cdot score + \beta_5 \tag{1.79}$$

where $score$ is the score obtained by the objective metric and $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5$ are the parameters of these regression functions. They are obtained through regression to minimize the difference between $DMOS_p$ and $DMOS$.

1.7 Conclusion

In this chapter, we reviewed the basic principle of human depth perception and various 3D content formats. Considering their advantages and drawbacks, the MVD format could be the most suitable one for most 3D applications. Next, the principle and the specific distortion of DIBR view synthesis along with the existing state-of-the-art DIBR quality metrics are discussed. Most of the metrics introduced above are FR metrics, however in some 3D applications, there is only a limited number of viewpoints which are captured, coded and transmitted, there is a large number of views which do not have reference views need to be synthesized. And none of these metrics can achieve a satisfactory performance on the existing databases. As a result, the efficient and NR quality assessment tools are in great need. In the chapter 2 and 3, we present two NR, two FR quality metrics.

CHAPTER 2

PROPOSED FULL-REFERENCE IMAGE QUALITY ASSESSMENT METRICS

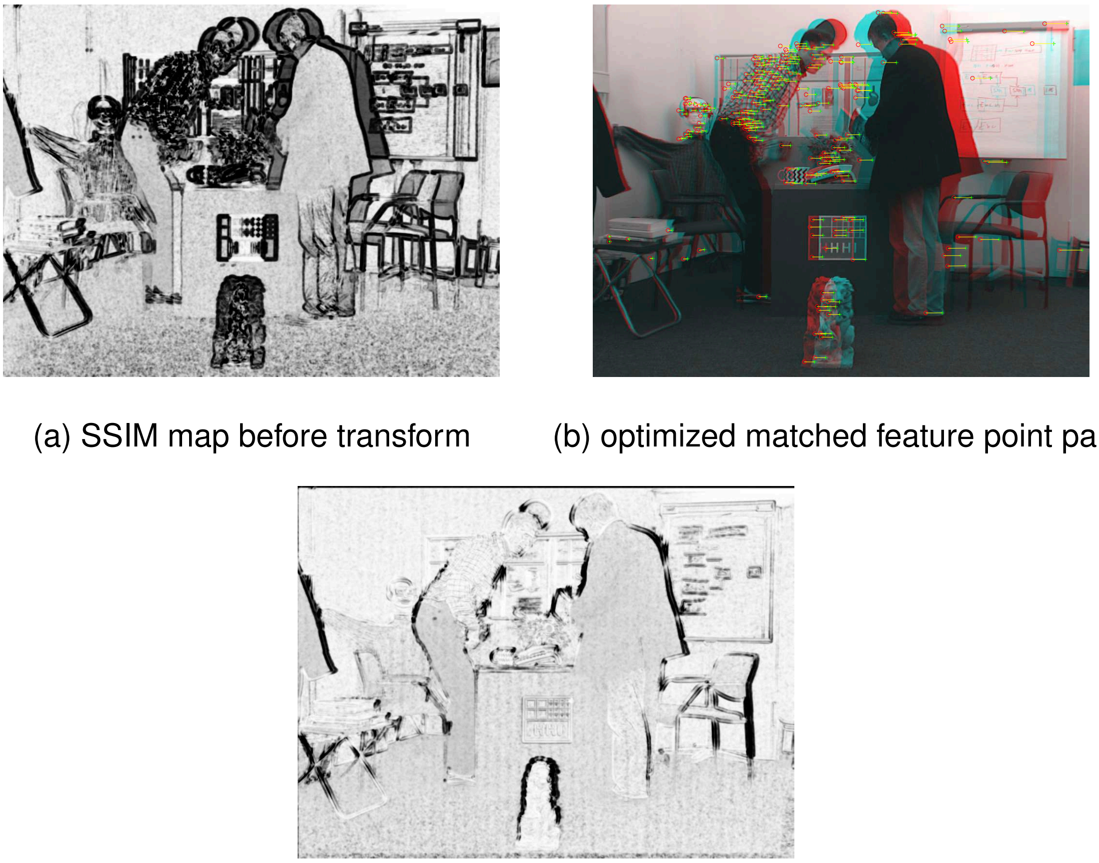
As introduced in the chapter 1, several FR metrics have been proposed to evaluate the quality of DIBR-synthesized views in the past few years, they still can not achieve a satisfactory result on the existing databases. In this chapter, we propose two FR metrics to handle the geometric distortions in the DIBR-synthesized views. First of all, a SURF+RANSAC homography approach is used to roughly compensate the global geometric shift. Then, in the first FR quality model, we use a dis-occlusion mask to weight the final distortions in the synthesized view since the synthesis distortions mainly happen in the dis-occluded areas. While in the second FR quality metric, a multi-resolution block matching method is used to precisely compensate the object shift and penalize the local geometric distortion as well. In addition, a visual saliency map is used as a weighting function. To calculate the final overall quality scores, only the worst blocks are utilized since the biggest distortions have the most effects on the overall perceptual quality.

This study is presented in 4 sections. The first two sections present the FR quality model Shift Compensation and Dis-occlusion based Model (SC-DM) and the FR metric SC-IQA in detail; the experimental results are shown in the third section; finally, the conclusion is drawn in section 4.

2.1 SC-DM: Shift Compensation and Dis-occlusion based Model

According to recent research [61], Human Visual System (HVS) is more sensitive to local artifacts compared to the global object shift. However, the global shift in DIBR-synthesized views can be easily penalized by most pixel-wise quality metrics, eg.

PSNR, SSIM. In this section, we propose a full-reference quality assessment model for 3D synthesized views by firstly compensating the global object shift, and then use a disparity map as a mask to weight the final distortion. This model can be combined with any pixel based FR metrics. In this work, we test it on the commonly used FR metrics PSNR and SSIM. This method can be divided into two parts: global shift compensation and dis-occlusion mask weighting.



(a) SSIM map before transform

(b) optimized matched feature point pairs

(c) SSIM map after transform

Figure 2.1: Example of feature points matching and transform

2.1.1 Global shift compensation

Fig. 2.1 (a) gives an example of the SSIM map between the synthesized image and the reference image in the adopted database [12], it can be observed that there exists great global shift between the synthesized image and the reference image.

In this part, the global geometric shift is compensated roughly by a SURF [5] + RANSAC [25] homography approach. Firstly, SURF feature points in the reference and synthesized images are detected and matched. Then, to be robust, the RANSAC algorithm is used to refine the matching and estimate the homography matrix H. After that, the pixels of the synthesized image are warped to the corresponding positions in the reference image by H. The SSIM map before transform, the matched feature point pairs and the SSIM map after transform are shown in Fig. 2.1.

We can observe that the global shift between the synthesized and the reference images has been roughly compensated since only a limited number of regions gets very low SSIM value (the black regions in Fig. 2.1(c)). Compared to the SSIM map before transform (Fig. 2.1(a)), the SSIM map after transform Fig. 2.1(c) shows that most of the ghost effect in the SSIM map has been removed.

2.1.2 Dis-occlusion Mask

In this section, we use a dis-occlusion mask to weight the difference between the synthesized image and the reference image. As introduced in Chapter 1, the major problem of the DIBR method is the dis-occlusion: regions which are occluded in the captured views become visible in the virtual ones. Due to the lack of original texture information, a synthesized image often contains dis-occlusion holes which significantly degrade the quality. Thus, we utilize a dis-occlusion mask to weight the final distortion. The depth map in the original view-point ($Depth_o$) is used to calculate the dis-occlusion mask. In a rectified configuration, 3D warping process, the horizontal disparity which is the horizontal displacement for each pixel can be obtained by Eq. (2.1):

$$d = \frac{f \times l}{Z} \tag{2.1}$$

where f, l, Z represent the camera focal length, the baseline distance between these two views and the depth value of this pixel respectively.

The depth map in the synthesized view-point ($Depth_s$) given initial value to -1, then the depth map in the original view-point ($Depth_o$) is warped to the synthesized view-point by Eq. (2.2):

$$Depth_s(i + d, :) = Depth_o(i, :); \quad (i + d), i \in [1, W] \tag{2.2}$$

where W is the image width, the colon ":" indicates all subscripts in this array dimension.

The dis-occluded mask dis_mask can then be obtained by extracting all the pixels with value -1 in $Depth_s$, which is shown in Fig. 2.2. This mask is a binary image, the while pixel's value equals "1", while the dark pixel's value equals "0".

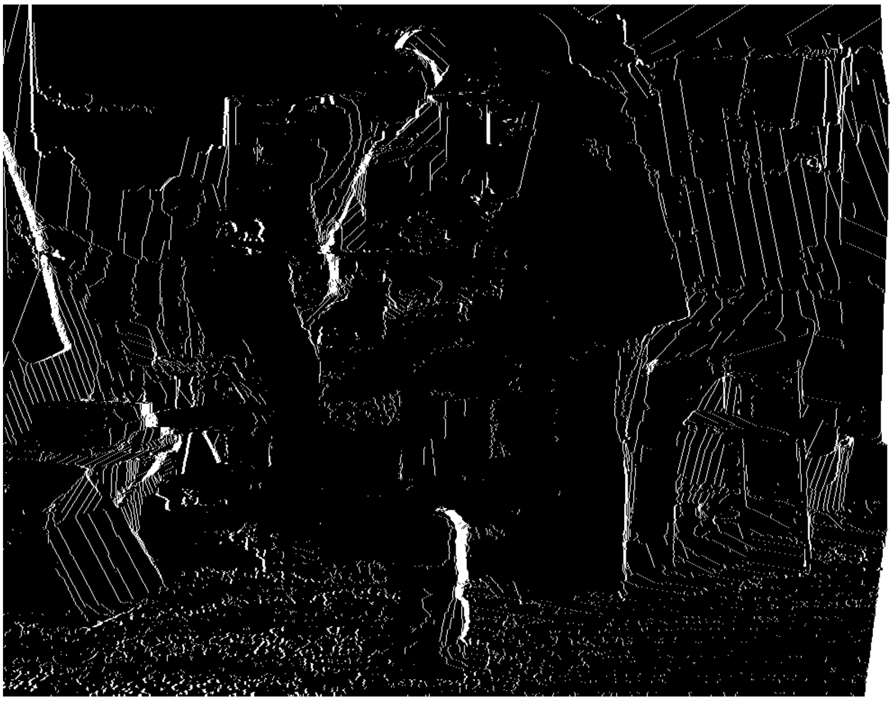


Figure 2.2: Example of dis-occluded mask

2.1.3 Weighted PSNR and Weighted SSIM

Generally speaking, the dis-occlusion mask dis_mask, can be integrated into any existing full-reference metric as a weighting mask since the DIBR view synthesis distortion mainly occur in the dis-occluded regions. We propose and test the weighted $PSNR$ ($PSNR'$) and $SSIM$ ($SSIM'$) as defined in the following equations:

$$MSE' = \frac{\sum_{(i,j)\in I}(I_{syn}(i,j) - I_{ref}(i,j))^2 \cdot dis\_mask(i,j)}{\sum_{(i,j)\in I} dis\_mask(i,j)} \tag{2.3}$$

$$PSNR' = 10 \cdot log_{10}(\frac{255 \times 255}{MSE'}) \tag{2.4}$$

$$SSIM' = \frac{\sum_{(i,j)\in I} SSIM(i,j) \cdot dis\_mask(i,j)}{\sum_{(i,j)\in I} dis\_mask(i,j)} \tag{2.5}$$

where I_{syn} and I_{ref} denote the the compensated synthesized image and the reference image respectively; dis_mask denotes the obtained disocclusion mask; $SSIM$ denotes the $SSIM\ map$ between the compensated synthesized image and the reference image. The experimental results of this model will be presented in section 2.3.

2.2 SC-IQA: Shift compensation based image quality assessment

As mentioned in the previous section, the Human Visual System (HVS) is more sensitive to the local artifacts compared to the global object shift. However, the local object shift within small distance is acceptable to the observer. In the previous model SC-DM, the global shift has not been compensated precisely, in this section, we propose an FR shift compensation based image quality assessment metric (SC-IQA) for DIBR-synthesized views by using a multi-resolution block matching method. Besides, it does not need the depth map. The block diagram is shown in Fig. 2.3, in addition to the SURF + RANSAC homography transform, a multi-resolution block matching is proposed to precisely compensate the object shift and penalize the local artifacts. Besides, a saliency map is used as a weighting function to improve the performance. The final overall quality scores are obtain by measuring the γ% worst blocks since human observers are more sensitive to poor quality regions rather the good ones.

2.2.1 Multi-resolution block matching

In this part, a multi-resolution block matching algorithm is used to precisely compensate the shift and also to detect the large geometric distortions. We will see bellow on an example why a regular block-matching would not be adequate. In the first step, we use a large block $N1 \times N1$ (N1 = 64) for primary matching; then we use a small block $N2 \times N2$ (N2 = 8) for final matching. The matching process can be described by the

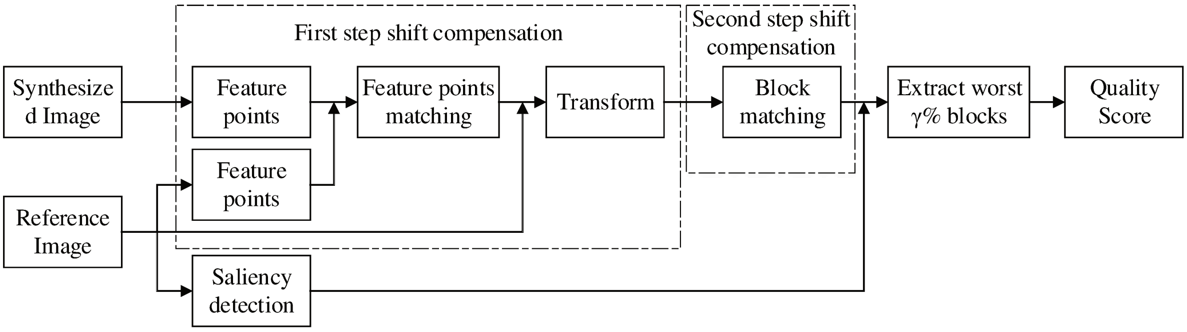


Figure 2.3: Block scheme of the SC-IQA metric

following steps:

1. Divide the synthetized view into a regular grid of $N1 \times N1$ blocks;

2. For each $N1 \times N1$ block, search for the best matching block in the reference view. The best matching block is the one showing the largest following similarity criterion:

$$sim(s,r) = \frac{cov(s,r) + \epsilon}{var(s) + var(r) + \epsilon} \tag{2.6}$$

   where s, r denote the blocks in the synthesized image and the reference image; the operation cov and var denote the co-variance and variance respectively; ϵ is a constant value to stabilize the division with weak denominator.

3. Each $N1 \times N1$ block is divided into smaller $N2 \times N2$ blocks and the process is repeated with a smaller search window.

Since the shift only occurs in the horizontal direction, we only search the blocks in this direction for matching. We assume the biggest shift in the synthesized image to be 30, the search windows of N1 and N2 are restricted to 30 and 5 respectively.

The goal of this multi-resolution block matching algorithm is to compensate the global shift and not compensate local geometric distortions, so that they will be penalized. Now, if we directly use $N2 \times N2$ block for matching, and set the search window to 30 (the biggest shift range in the synthesized image), the computational complexity will be much higher. Besides, as shown in Fig. 2.4, there exists great geometric distortion in the red block ($N1 \times N1$) in Fig. 2.4 (a) compared to its matched block in the reference image (the red block in Fig. 2.4 (b)). If we directly use 8×8 block for matching and set the searching window to 30, the best matched block for the block in Fig. 2.4

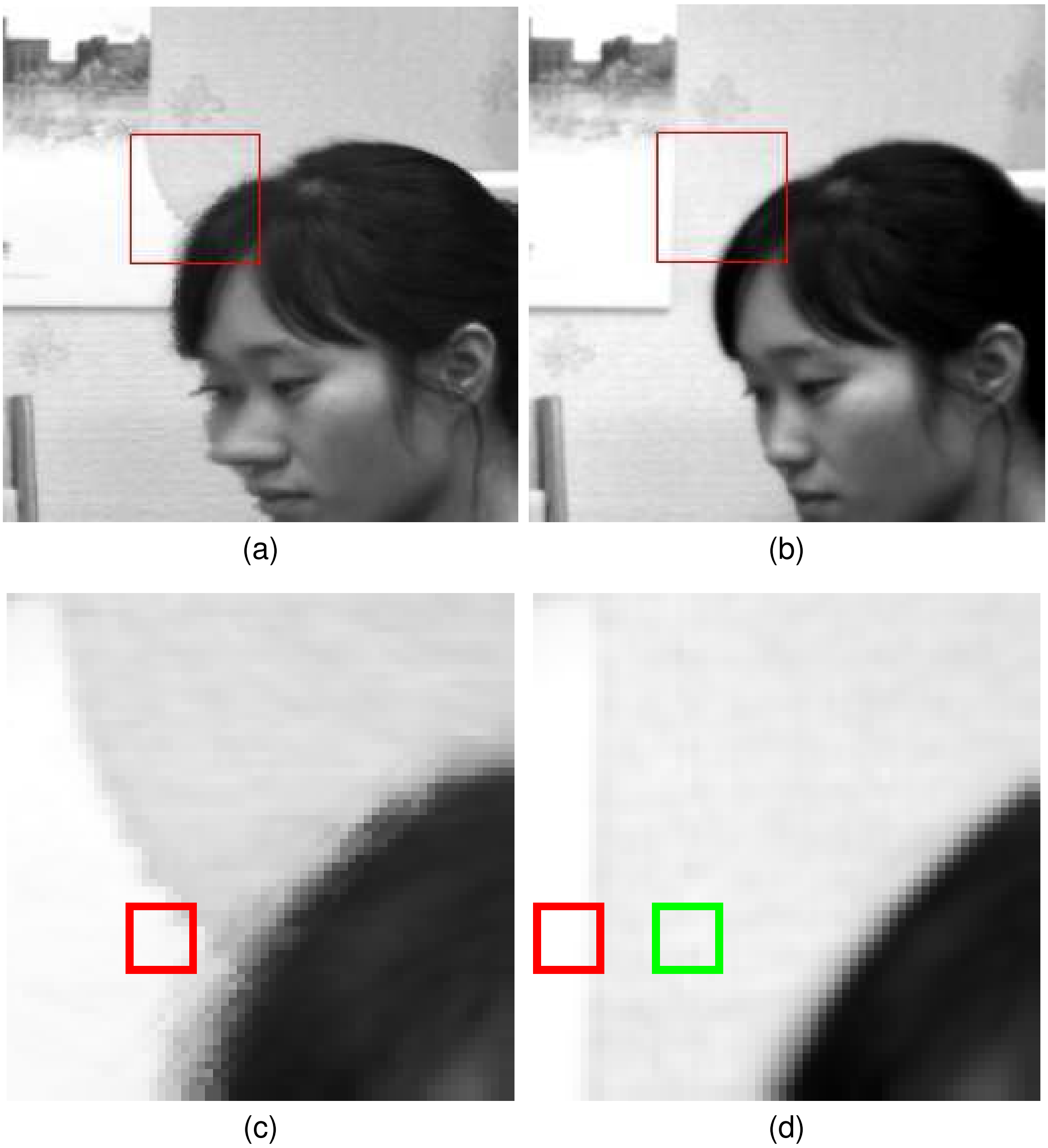


Figure 2.4: Block matching: (a), (b) are the patches in the synthesized and the reference image; (c) block in the synthesized image; (d) matched block in the reference image: for direct 8x8 block-matching (red block), or multiresolution block-matching (green)

(c) is the red block in Fig. 2.4 (d). There exists little difference between these two red blocks, so the geometric distortion will not be penalized. On the contrary, if we use the proposed multi-resolution block matching method, the matched block is the green one, this geometric distortion will be surely penalized. The multi-resolution approach is thus more efficient to find the real physically matching block, and detect wether there is local distorsion within this block..

2.2.2 Saliency weighting

In addition, a saliency detection [39] is also used as a weighting map to improve the performance of the proposed metric. The distortion of each $N2 \times N2$ block is measured by averaging the weighted mean square errors between the blocks of the synthesized and the reference images, as shown in:

$$MSE_B = \frac{\sum_{(i,j)\in B} (syn(i,j) - ref(i,j))^2 \times Sal\_map(i,j)}{\sum_{(i,j)\in B} Sal\_map(i,j)} \tag{2.7}$$

where B means the matched $N2 \times N2$ blocks; (i,j) denotes the pixel in the block; syn and ref represent the blocks in the synthesized image and reference image respectively; Sal_map represents the saliency map in this block.

2.2.3 Quality pooling

Since humans tend to perceive poor regions in an image with more severity than the good ones [61, 52], we only use the blocks with the worst quality to calculate the final quality as shown in Eq. 2.8.

$$MSE_W = \frac{1}{N_W} \sum_{i\in W} MSE_B(i) \tag{2.8}$$

where W represents the set of the worst γ% blocks in the image, N_W is the number of items in the set W. The final quality score is computed as the following equation:

$$Score_{SC-IQA} = 10 \times log_{10}(255 \times 255/MSE_W) \tag{2.9}$$

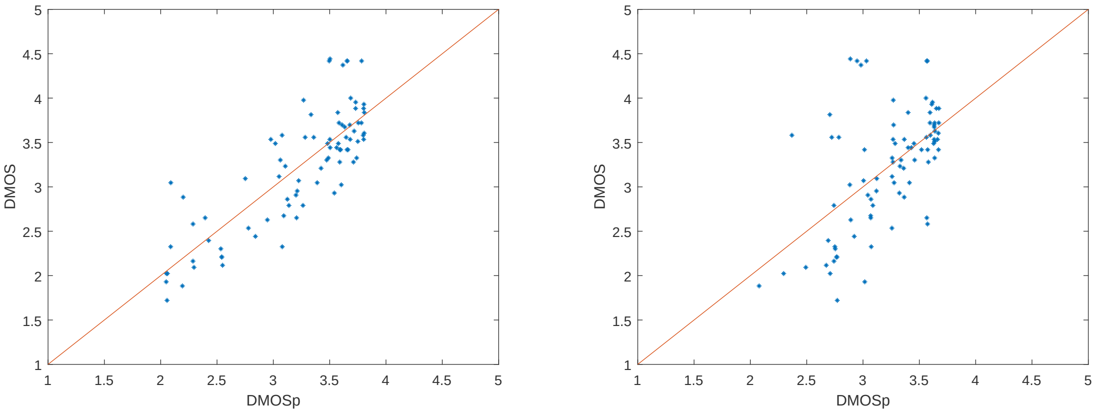
where a higher quality score indicates a better quality.

2.3 Experimental results

This section describes and discusses the validation experiments of the proposed FR quality model and SC-IQA metric on MCL-3D database [87] and IVC database[37].

2.3.1 Performance on IVC database

The performance of SC-IQA is tested on the IRCCyN/IVC DIBR database [37], which consists of 84 synthesized views generated by seven different DIBR view synthesis algorithms and their associated 12 reference views along with the subjective scores - mean opinion score (MOS).



(a) The proposed weighted PSNR

(b) The proposed weighted SSIM

Figure 2.5: Scatter plot of DMOS versus the predicted quality score

Similar to previous Chapter, we use the same nonlinear regression function[27] to map the objective quality scores to the subjective scores. The scatter plot of the predicted scores versus the subjective scores and the regression function are shown in Fig. 2.5 and Fig 2.6.

The obtained PLCC, RMSE and SROCC values are given in table 2.1. It can be noticed that the proposed weighted PSNR ($PSNR'$) and SC-IQA (γ = 1) perform significantly better than the other tested metrics (including 8 FR, 2 RR, 2 sides view based FR and two NR quality metrics, these metrics have been presented in detail in chapter 1) in terms of PLCC and SROCC. The PLCC gain of $PSNR'$ achieves 36.85% compared to the PSNR. The weighted SSIM ($SSIM'$) achieves a gain of PLCC 13.33% compared

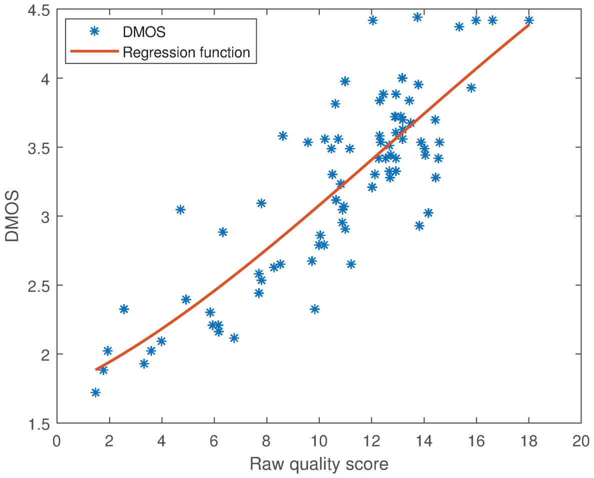


Figure 2.6: Scatter plot of SC-IQA quality score versus DMOS

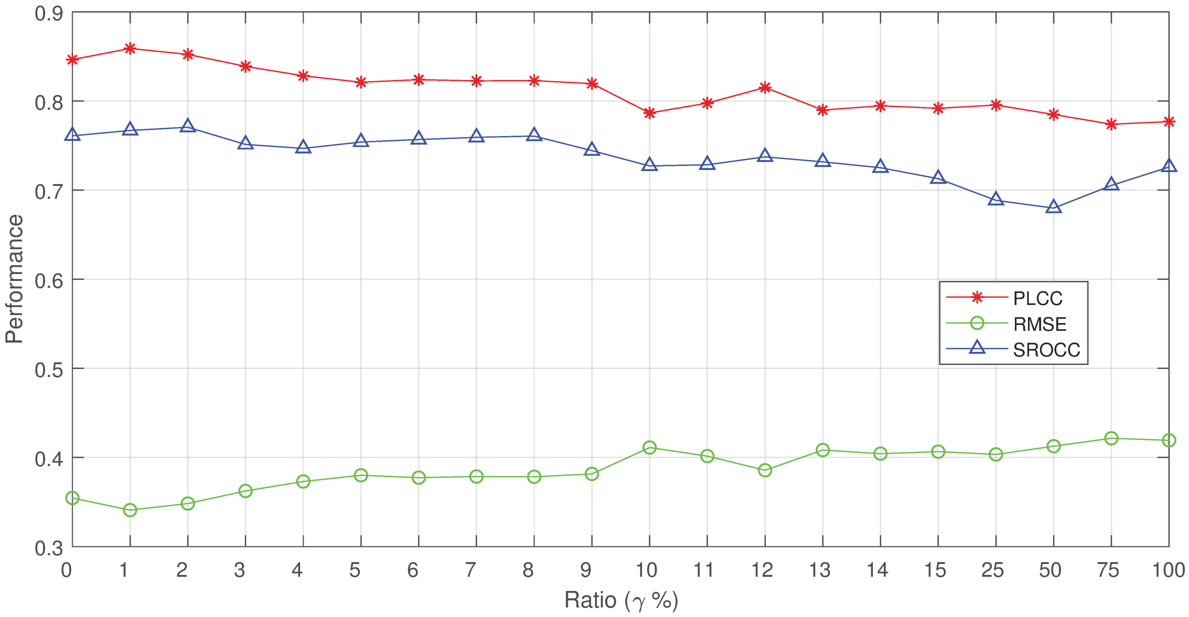


Figure 2.7: Performance dependency of the proposed metric with the changing ratios (γ%)

to the SSIM.

The performance dependency of the proposed metric on ratio γ% is also discussed as shown in Fig. 5. Especially the ratio 0 is associated with the proposed metric without saliency map. According to Fig. 2.7, the use of the saliency map improves the proposed metric slightly; the performance of the proposed metric goes down as the ratio increases, however even the lowest scores are still superior to most of the state-of-the-art metrics in Table I. This shows the robustness of the proposed metric.

Table 2.1: Performance comparison of the proposed method with the state-of-the-art metrics on IVC database

Metric		PLCC	RMSE	SROCC
FR 2D metrics	PSNR	0.4557	0.5927	0.4417
	SSIM	0.4348	0.5996	0.4004
FR 3D metrics	3DSwIM	0.6864	0.4842	0.6125
	VSQA	0.6122	0.5265	0.6032
	MP-PSNR	0.6729	0.4925	0.6272
	MW-PSNR	0.6200	0.5224	0.5739
	CT-IQA	0.6809	0.4877	—
	EM-IQA	0.7430	0.4455	—
	PSNR'(pro)	0.8242	0.3771	0.7889
	SSIM'(pro)	0.5681	0.5479	0.5475
	SC-IQA(γ = 1)	**0.8496**	**0.3511**	**0.7640**
RR 3D metrics	MP-PSNRr	0.6954	0.4784	0.6606
	MW-PSNRr	0.6625	0.4987	0.6232
SV FR metrics	SIQE	0.7650	0.5382	0.4492
	DSQM	0.7430	0.4455	0.7067

2.3.2 Performance on MCL-3D database

Besides, we test the proposed FR metrics on MCL-3D database. This stereoscopic 3D database uses DIBR technology to synthesize the left and the right views from image-plus-depth source. Nine MVD sequences are collected, among which $Kendo$, $Lovebird1$, $Balloons$, $PoznanStreet$ and $PoznanHall2$ are natural images; $Shark$, $Microworld$, $GTFly$ and $Undodancer$ are Computer Graphics images. Many types of distortions are considered in this database, such as Gaussian blur, additive white noise, down-sampling blur, JPEG and JPEG-2000 (JP2K) compression and transmission error. These distortions are applied on either the original texture images or the depth images before the view synthesis. In addition, the distortion caused by imperfect DIBR algorithms are also considered in this database. Four DIBR view synthesis algorithms ([23] [94] [86] plus DIBR without hole filling) were used.

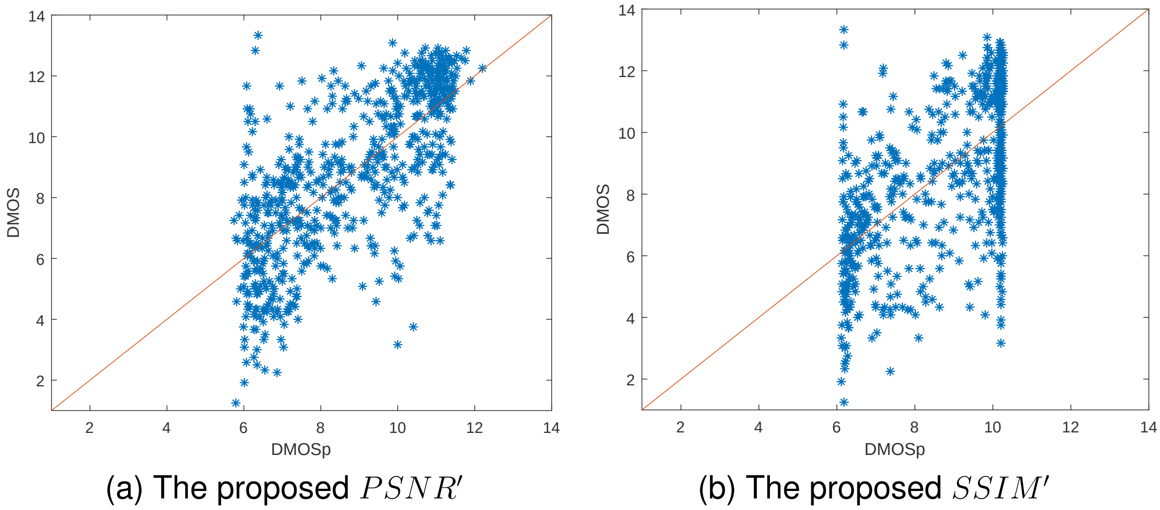
Table 2.2: Performance comparison of the proposed method with the state-of-the-art metrics on MCL-3D database

Metric		PLCC	RMSE	SROCC
FR 2D metrics	PSNR	0.7852	1.6112	0.7915
	SSIM	0.7331	1.7693	0.7470
FR 3D metrics	3DSwIM	0.6519	1.9729	0.5683
	VSQA	0.5078	2.9175	0.5120
	MP-PSNR	0.7831	1.6179	0.7899
	MW-PSNR	0.7654	1.6743	0.7721
	PSNR'(pro)	0.7166	1.8141	0.7197
	SSIM'(pro)	0.6000	2.0814	0.5451
	SC-IQA(γ = 1)	**0.8194**	**1.4913**	**0.8247**
RR 3D metrics	MP-PSNRr	0.7740	1.6474	0.7802
	MW-PSNRr	0.7579	1.7012	0.7665
SV FR metrics	SIQE	0.6734	1.9233	0.6976
	DSQM	0.6995	1.8593	0.6980

Table 2.2 gives the obtained PLCC, RMSE and SROCC coefficient values on the MCL-3D database. It shows that the proposed SC-IQA still performs the best among all the tested quality metrics. However, the proposed weighted PSNR (PSNR') and weighted SSIM (SSIM') perform not as good as they do on IVC database, and even worse than the original PSNR and SSIM. The main reason could be that the PSNR' and SSIM' focus on the distortions caused by DIBR view synthesis, but the main distortions in MCL-3D database are the conventional distortions in texture and depth map. The scatter plot of the predicted scores versus the subjective scores are shown in Fig. Fig. 2.8 and Fig. 2.9. This result shows that the $PSNR'$ and $SSIM'$ fail to assess the low quality images while the and SC-IQA succeed in assessing the high quality images.

2.4 Conclusion

In this chapter, we proposed two full-reference quality metrics for DIBR-Synthesized views. The great advantage of the proposed metrics is their simplicity. The idea of the first FR quality model is to improve the existing simple 2D metrics by addressing two issues: 1) compensating the global significant shift in the synthesized view (by an SURF+RANSAC homography); 2) putting more weight on the distortions occurring in



(a) The proposed $PSNR'$

(b) The proposed $SSIM'$

Figure 2.8: Scatter plot of DMOS versus the predicted quality score on MCL-3D database

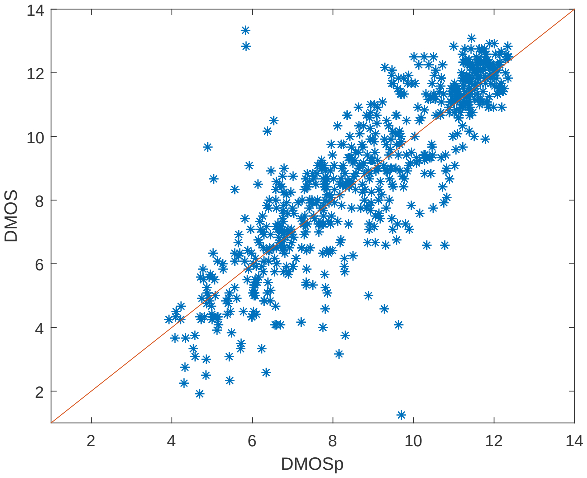


Figure 2.9: Scatter plot of SC-IQA ($\gamma = 1$) quality score versus DMOS on MCL-3D database

the dis-occluded regions (which are estimated using the depth map here). In the second FR quality metric SC-IQA, we focus on the object shift in the DIBR-Synthesized views. The same as the first FR quality model, we use an SURF + RANSAC homograhy approach to compensate the global shift in the synthesized image. Then, a multi-resolution block matching method is used to compensate the object shift and penalize the local geometric distortion as well. In addition, a saliency map is used to weight the final distortions in the synthesized view. Experimental results show that the proposed weighted PSNR ($PSNR'$) greatly improves the performance compared to the original PSNR (gain of 36.85% in terms of PLCC). The weighted SSIM ($SSIM'$) earns also a gain of 13.33% (PLCC) compared to its 2D version. The proposed SC-IQA metric and weighted PNSR ($PSNR'$) significantly outperform the tested state-of-the-art 3D synthesized view dedicated metrics: 3DSwIM, MP-PSNR, MW-PSNR, VSQA, CT-IQA and EM-IQA.

CHAPTER 3

PROPOSED NO-REFERENCE IMAGE QUALITY ASSESSMENT METRICS

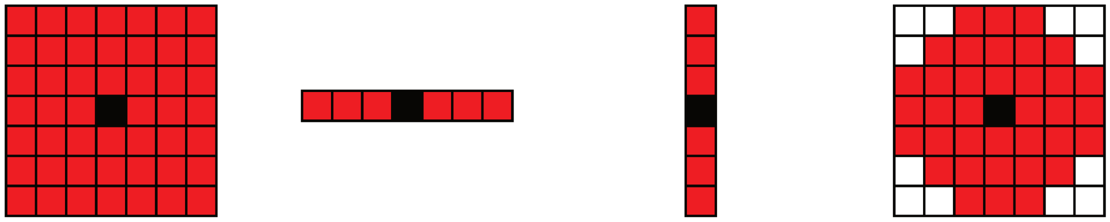
The major limitation of the Full-Reference metrics is that they always need the reference view which may be unavailable in some circumstances. In other words, there is no ground truth for a full comparison with the distorted synthesized view. To this end, in this chapter, we propose two novel No-Reference image quality assessment metrics for DIBR-synthesized views (called NIQSV and NIQSV+). These blind metrics are based on mathematical morphology. They can evaluate the quality of synthesized views by measuring the typical synthesis distortions with access to neither the reference image nor the depth map. (These contributions have been published in [100, 101])

This chapter is organized as follows: firstly, the mathematical morphology in image processing is introduced in Section 2.1; then the proposed metrics NIQSV and NIQSV+ are presented in detail in the second and third sections, followed by the experimental results and discussion in the fourth section; finally the conclusion is draw in the fifth section.

3.1 Introduction of mathematical morphology in image processing

Apart from the convolution based filters, the mathematical morphology in image processing is a collection of non-linear operations related with object shape.[1] It uses a template, which is called Structural element (SE), to probe the image. Some typical SE are shown in Fig 3.1. The red blocks form the origin of SE, while the black blocks present the $origin$ of SE. The $origin$ indicates the processed pixel in the morphological operation, which is generally set to the centroid of SE.

[1] 88.



(a) SE squre, size 7 (b) SE line, size 7 (c) SE line, size 7 (d) SE circle, size 8

Figure 3.1: Examples of structural elements.

3.1.1 Morphological operations on binary image

Mathematical morphology on a binary image (0 and 1 values, with 1 values defining the shape and 0 the background, for instance, or other definitions) can be recognized as set operations. The two basic erosion and dilation operations are defined in Eq. 3.1 and Eq. 3.2.

$$Erosion : I \ominus SE = \{z \in \epsilon^2 | SE_z \subseteq I\} \tag{3.1}$$

$$Dilation : I \oplus SE = \bigcup_{se \in SE} I_{se} \tag{3.2}$$

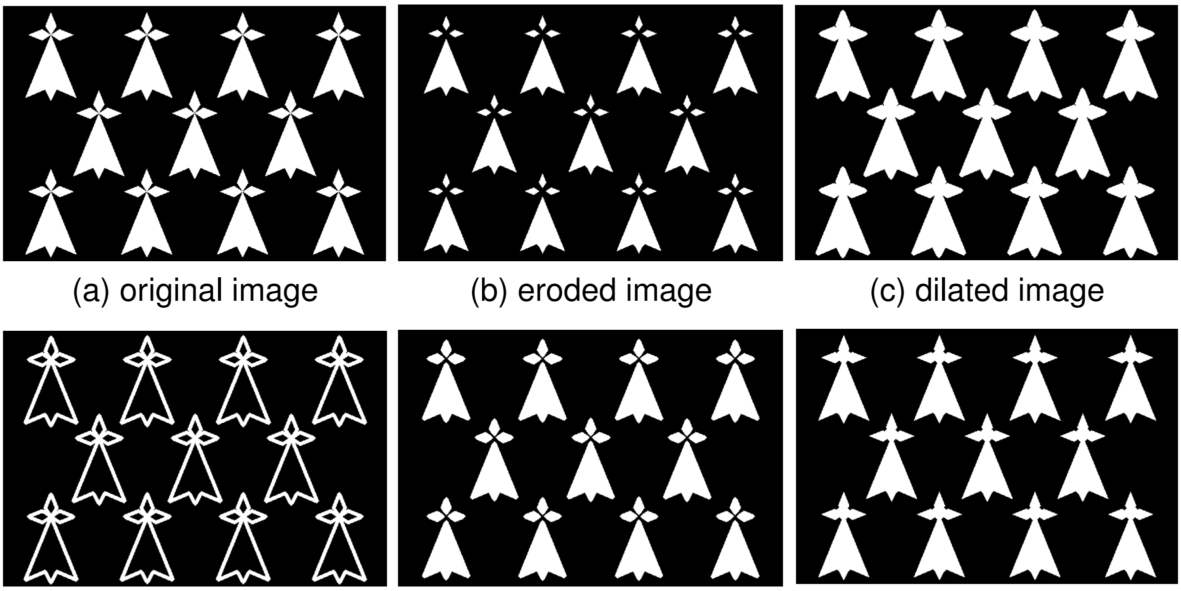
where z and SE denote the processed pixel and the SE. ϵ^2 is a Euclidean space or an integer grid, and I a binary image in ϵ^2. SE_z is the translation of SE by the vector z, I_{se} is the translation of I by the vector se which could be defined as follows:

$$SE_z = \{se + z | se \in SE\}, \forall z \in \epsilon^2 \tag{3.3}$$

$$I_{se} = \{z + se | z \in I\}, \forall se \in SE \tag{3.4}$$

The erosion operation of a binary image produces a new binary image with value 1 in all pixel positions at which that SE fits the image, the value of $origin$ point is set to 1 when all the pixel values in SE_z are 1. The dilation operation produces a new binary image with 1s in all pixel positions at which the SE hits the image. As the structural elements (SE) is scanned over the image, the pixel value of the $origin$ point in SE_z is set to 1 when there is a pixel (in SE_z) whose value is 1.

The Fig. 3.2 (b)(c) gives two examples of erosion and dilation. It can be observed that the erosion operation removes the pixels on the object boundaries and reduces the size of object. On the contrary, the dilation operation enlarges the object size by adding pixels on the edges. Both operations conserve the object shape in the image. One of the simplest application of these operations is that the edge image can be obtained by subtracting the erode image from dilated image cf. Fig 3.2 (d).



(a) original image (b) eroded image (c) dilated image

(d) edge image (e) opened image (f) closed image

Figure 3.2: Examples of morphological operations on binary images.

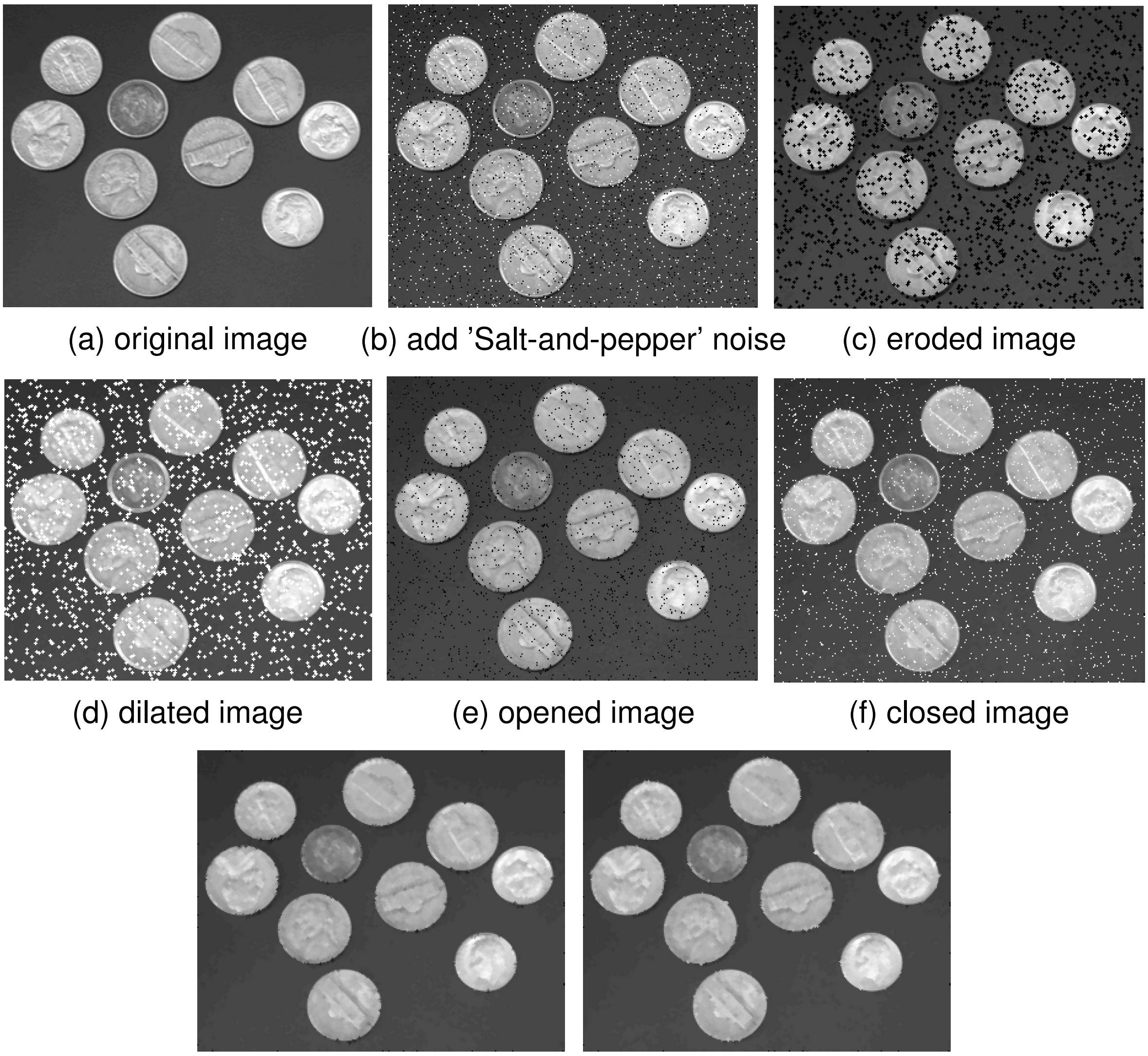
The opening and closing operations are defined as the combinations of erosion and dilation operations with the same SE cf. Eq. 3.5, Eq. 3.6. As shown in Fig. 3.2 (e), (f), the opening operation opens the gap between the objects with thin connections. The closing operation removes the small holes between the objects while keeping the original object's size.

$$Opening : I \circ SE = (I \ominus SE) \oplus SE \tag{3.5}$$

$$Closing : I \bullet SE = (I \oplus SE) \ominus SE \tag{3.6}$$

3.1.2 Morphological operations on gray level images

The morphological operations are firstly developed for binary images, and then extended to gray level images. The binary morphology can be recognized as the gray level morphology applied on binary images. The morphological operations on gray level images are the calculations of maximum or minimum values in the neighbourhood of



(a) original image (b) add 'Salt-and-pepper' noise (c) eroded image

(d) dilated image (e) opened image (f) closed image

(g) opened + closed image (h) closed +opened image

Figure 3.3: Examples of morphological operations on gray level images.

the processed pixel. The two basic morphological operations on gray level images are defined in Eq. 3.7 and Eq. 3.8. The opening and closing operation in gray level image are the same as that in Eq; 3.5 and Eq. 3.6.

$$Erosion : I \ominus SE = \min_{se \in SE} \{I(z + se)\} \tag{3.7}$$

$$Dilation : I \oplus SE = \max_{se \in SE} \{I(z - se)\} \tag{3.8}$$

Fig. 3.3 shows some examples of morphological operations on gray level images. In image (b), we add some 'salt and pepper' noise. From the image (c), (d), it can be observed that the erosion operation can remove white noise pixels while the dilation operation can remove the black noise. So we use a opening and a closing operation on the noisy image respectively. The results show that the opening removes most of the while noise pixels, but left some black noise pixels. While the closing operation fills the black noise pixels, but has no influence on the white pixels in the image. In the image (g) and (h), we use a closing operation followed by an opening operation or an opening operation followed by a closing operation, the results show that they perform well on removing the white noise pixels and filling the black noise pixels at the same time. Based on shapes, mathematical morphology is widely used in image processing, such as removing the noise while keeping the object's shape.

3.2 Proposed metric NIQSV

In this section, we propose a new No-reference quality assessment model to evaluate the quality of 3D synthesized views, called NIQSV (No-reference Image Quality assessment of Synthesized Views). It is based on the following image model: a good quality image is assumed to present sharp and regular object borders, smooth values inside the object and large discontinuities at the object borders. Such "perfect" images are insensitive to opening and closing morphological operations while some artifacts such as blurry regions around the object edges and crumbling in the synthesized views are sensitive to such morphological operations. The crumbling is small-sized artifacts which can be easily detected by the morphological operations with Structural Element (SE) larger than their size, as shown in Figure 4; the blurry regions change much more significantly after the opening and closing morphological operations compared to the good quality images with sharp edges and flat areas. Thus, these properties could be used to detect these artifacts.

The principle of NIQSV is the following: it quantifies the distortions in luminance component Y and chrominance components U, V using a set of morphological operations. Then the 3 obtained distortions are pooled into 1 global distortion by a weighted average. Furthermore, an edge image is utilized to weight the final distortion since the distortions of synthesized views mainly happen around object edges. The block scheme is presented in Fig 3.6.

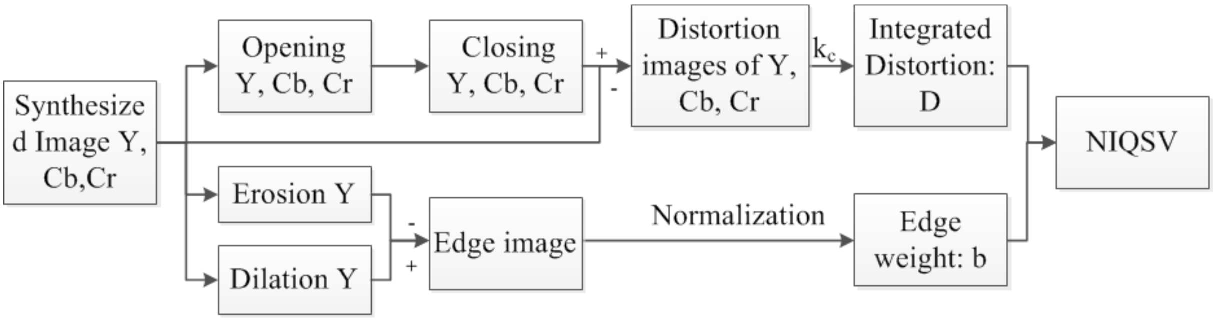


Figure 3.4: Block scheme of NIQSV

The key strategy of NIQSV is a pair of opening and closing operations. The opening operation used on the synthesized image can help to remove some thin blurry regions, and the following closing operation with a relatively larger Structural Element (SE) can fill the holes in the disoccluded areas. The distortion of each component is obtained by measuring the difference between the original component I_X and the processed component I'_X after the opening and closing operation. It can be computed as follows:

$$I'_X = (I_X \circ SE_o) \bullet SE_c, X \in (Y, Cb, Cr) \tag{3.9}$$

$$D_X = |I'_X - I_X|, X \in (Y, Cb, Cr) \tag{3.10}$$

where D_X denotes the difference of each color component, I_X is the corresponding color component of the synthesized image, SE_o is the SE used for opening and SE_c is the SE used for closing. In this paper, the shape of SE_o and SE_c is a circle, the size of SE_o and SE_c is 3 and 8 respectively.

In order to obtain the overall distortion, the distortions of all components are integrated as in Eq 3.11:

$$D = (1 - w_c) \cdot D_Y + \frac{w_c}{2} \cdot (D_{Cb} + D_{Cr}) \tag{3.11}$$

which is a weighted sum of the distortions computed on each color component where the weight is related to the parameter w_c. The value of w_c is set to 0.5 which means that the distortion in luma component weights 50% in the overall distortion.

Since the artifacts mostly happen around the edges, the image edges must be taken into consideration. To reduce computational complexity, the edge image is firstly extracted by a pair of morphological operators as described in Eq. 3.12. Then, they are normalized to $[0, 1]$ using Eq. 3.12:

$$E = (I_Y \oplus SE) - (I_Y \ominus SE) \tag{3.12}$$

$$e = E/Vmax; Vmax = 255, e \in (0, 1) \tag{3.13}$$

where SE is the structural element used for erosion and dilation, the symbols $\oplus$ and $\ominus$ denote the morphological dilation erosion operation respectively. The shape of SE is a circle and its diameter is set to 4. $E/Vmax$ (where Vmax is the maximum value that an edge-detector may provide for 8-bit images: 255) is used as the edge weight. The final edge weight e is used to weight the overall difference D in the whole image. The pixels with higher edge value have more weight on the distortion map. Especially, for the pixels with no edge, the distortion on it will not be considered.

Finally, the overall image quality score $NIQSV$ is computed as follows:

$$MSE' = \frac{\sum_{(i,j)\in I} e(i,j) \cdot D(i,j)^2}{\sum_{(i,j)\in I} e(i,j)} \tag{3.14}$$

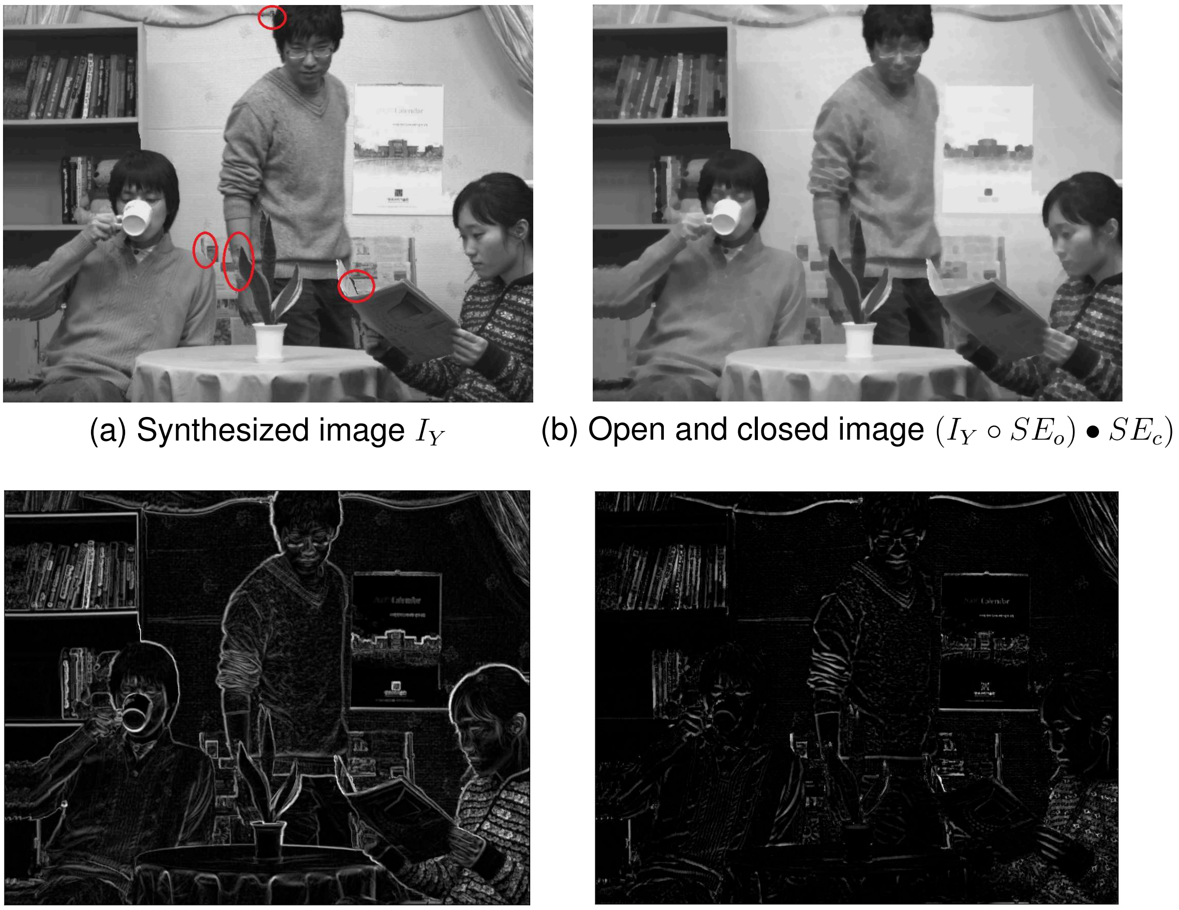
$$NIQSV = 10 \cdot log_{10}(\frac{255 \times 255}{MSE'}) \tag{3.15}$$

Fig. 3.5 shows the processed images of one synthesized view in the “Newspaper” sequence as an example.

3.3 Proposed metric NIQSV+

This section presents the details of the proposed metric (NIQSV+). As an extended version of NIQSV, it is also based on the assumption that the images with good quality are composed of flat regions separated by sharp and regular edges.

A block diagram of the proposed method is presented in Fig. 3.6. The proposed method can be divided into three parts. Part A is designed to detect the blurry regions and crumbling around the object edges, which has been introduced in the previous section NIQSV; part B is related to the unfilled black holes in the dis-occluded areas; and part C is the detection of stretching distortion which always occurs in the left or right side of the synthesized view.



(a) Synthesized image I_Y

(b) Open and closed image $(I_Y \circ SE_o) \bullet SE_c)$

(c) Normalized edge weight b

(d) Overall Difference D

Figure 3.5: Examples of intermediate results in the NIQSV measurement for one synthesized view in the “Newspaper”sequence. The distortions marked in (a) are well detected in (d), while in the non-distortion regions, such as the girl’s hair, the distortion values are very low.

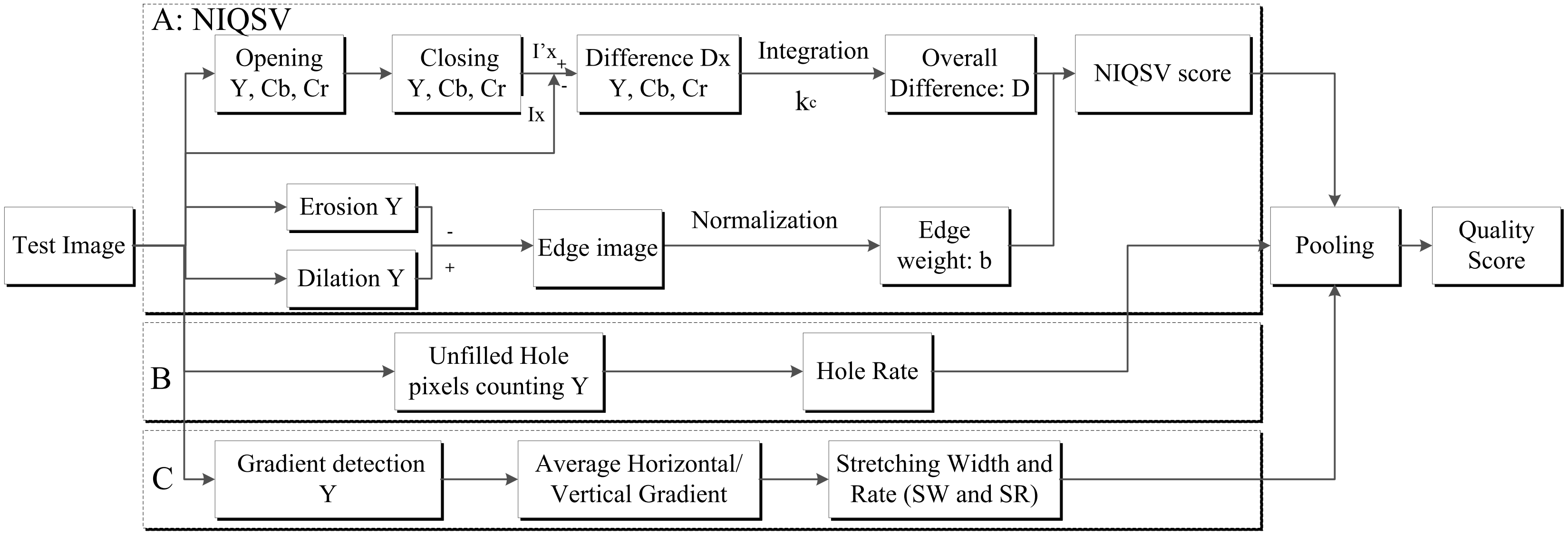


Figure 3.6: Block Diagram of NIQSV+

3.3.1 Detection of black holes

In this part, the distortion of unfilled black hole pixels is taken into consideration. Normally, most natural images do not contain pixels with 0 luminance value. Thus, we use the proportion of black hole pixels in the whole image to measure this type of distortion, as defined in Eq. 3.16:

$$Zrate = NumofBHpixels/(W \times H) \tag{3.16}$$

where $NumofBHpixels$ denotes the number of black hole pixels in the whole images, W and H are the width and the height of the image.

3.3.2 Detection of Stretching

The stretching may happen around the left or right side of the image due to lack of the corresponding texture information, as shown in Fig 1.8 (b).

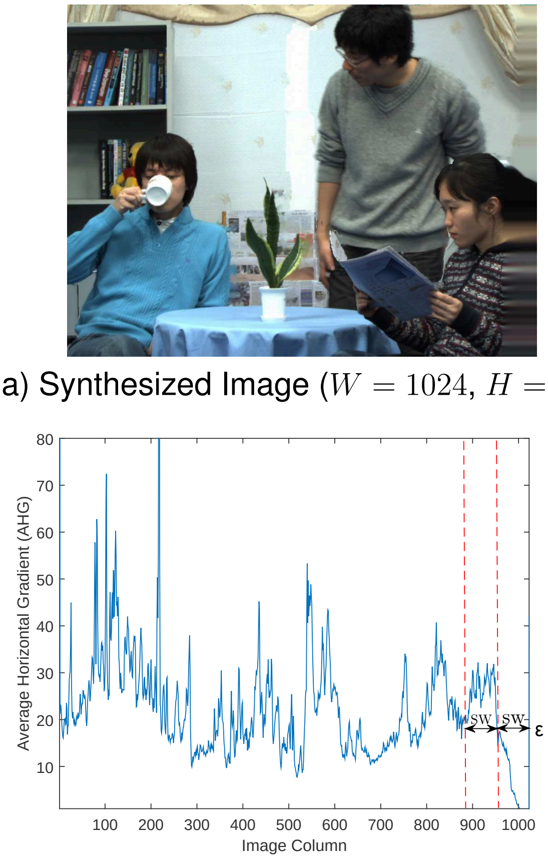
In this part, a stretching measurement is defined to estimate the level of stretching in the synthesized image. The stretching is detected by measuring the crash of horizontal gradient in the stretching area. Firstly, the horizontal and vertical gradients are calculated with the Sobel operator.

$$\begin{cases} \nabla_{ver} = I_y * G_{ver} \\ \\ \nabla_{hor} = I_y * G_{hor} \end{cases} \tag{3.17}$$

where I_y is the Y component of the synthesized image, G_{hor} and G_{ver} denote the Sobel horizontal and vertical gradient operator. The Average Horizontal / Vertical gradient ($\bar{g_H}$ / $\bar{g_V}$) in column are defined in Eq. 3.18.

$$\begin{cases} \bar{g_H} = \sum_{j=1}^{H} \nabla_{hor}/H \\ \\ \bar{g_V} = \sum_{j=1}^{H} \nabla_{ver}/H \end{cases} \tag{3.18}$$

where H denotes the height of the image. Since stretching mainly happens in the horizontal direction, the average horizontal gradient in the column can be used to detect the stretched regions.



(a) Synthesized Image ($W = 1024$, $H = 768$)

(b) $\bar{g_H}$ distribution

Figure 3.7: Synthesized Image and its corresponding average horizontal gradient

As shown in Fig. 3.7, the average values of the horizontal gradient in the stretched area (in the right side) are very low. The Stretching Width (W_s) is obtained by calculating the width of this area.

$$W_s = \sum_{j=1}^{0.1\times W} S + \sum_{j=0.9\times W}^{W} S \tag{3.19}$$

where S is a index to mark the stretching areas. Since the stretching artifacts only occur in the left or the right side of the image, $0.1 \times W$ and $0.9 \times W$ are used to take into account only the side portions of the image.

$$S = \begin{cases} 1, \bar{g_H} < \varepsilon \\ 0, \ else \end{cases} \tag{3.20}$$

where ε is a threshold used to extract the stretching regions, its value is set to 50% of the mean value of $\bar{g_H}$. However, even with the same stretching width, the perceptual annoyance could be different in different textures. In order to handle this issue, a Stretching Rate (R_s) is defined by comparing the average gradient in the stretching regions and those in the adjacent non-stretching regions with the same width, as shown in Fig. 3.7. The more similar these two regions are, the less the stretching will be perceptible. Since the mean horizontal gradient of these two regions is quite different, we only compare the vertical gradients.

$$R_s = \frac{\nabla_{ref} - \nabla_{str}}{\nabla_{ref}} \tag{3.21}$$

where ∇_{str} presents the average $\bar{g_V}$ value in the stretching area, ∇_{ref} is the average $\bar{g_V}$ value in the adjacent non-stretching regions with the same width. When the ∇_{str} and ∇_{ref} values are closer, this type of distortion is less significant, and the SR value is lower. The final stretching distortion is calculated as follows:

$$S\_index = (log_{10}(W_s + 1) + 1) \times (R_s + 1) \tag{3.22}$$

3.3.3 Overall quality measurement

Finally, the integrated overall quality score is computed as Eq. 3.23 since higher stretching and black hole rate indicate bad image quality.

$$NIQSV+ = \frac{NIQSV}{S\_index \times (1 + k_z \times Zrate) + C} \tag{3.23}$$

where k_z denotes the weight of black hole distortion to the final measurement. Since the black hole pixels hold a very low proportion in the whole image, k_z should be a large value. C is a constant used to adjust the difference between the images with “black hole” or “stretching” artifacts and those without. The dependency of these two parameters (k_z, C) are discussed in Section 2.3. The evaluation of the NIQSV and NIQSV+ proposed metrics is presented in the next section.

3.4 Results and discussions

This section evaluates the performance of proposed metrics: NIQSV and NIQSV+. In the following, we will firstly introduce the used database, and then present the performance comparison with other state-of-the-art metrics.

3.4.1 Database

The performances of the metrics are evaluated using IRCCyN/IVC DIBR image database [37, 12]. It contains the frames from 3 different MVD sequences: Book Arrival (1024×768, 16 cameras with 6.5 cm spacing), Lovebird1 (1024×768, 12 cameras with 3.5 cm spacing) and Newspaper (1024×768, 9 cameras with 5 cm spacing). For each sequence, there are four virtual views generated from another viewpoint using the following seven DIBR synthesis algorithms A1-A7:

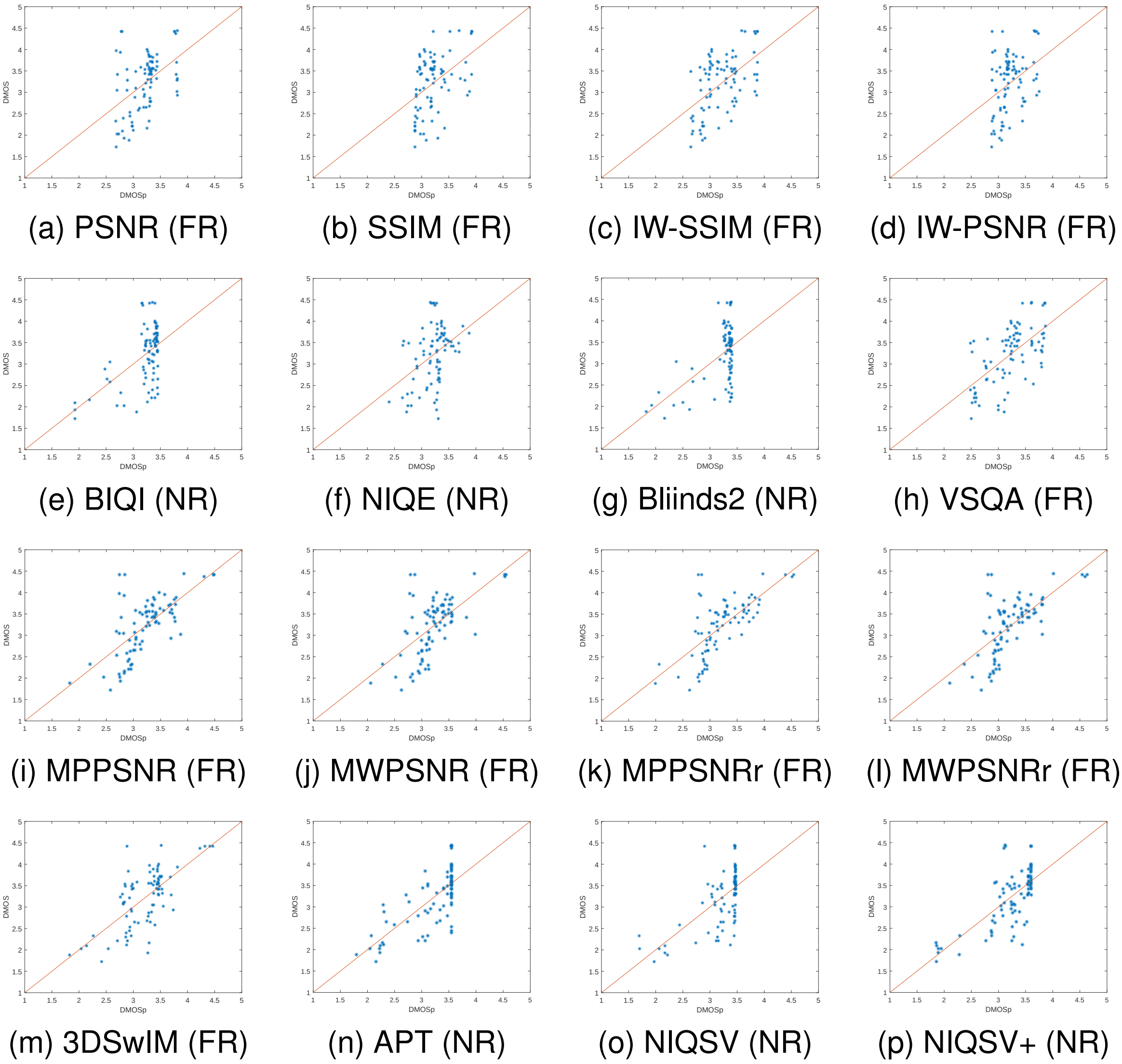
- A1 [23]: the depth map is filtered to remove depth discontinuities; borders are cropped and then the image is interpolated to reach its original size. This may lead to shifting and global radial artifacts.

- A2: the depth map is pre-processed in the same way as in A1, and the borders are in-painted as described in [94] instead of being cropped. This may induce blurring and geometry distortions around the object discontinuities since the depth map is pre-processed by a low-pass filter.

- A3: Tanimoto et al. [62] proposed a 3D view generation system which is adopted as a reference software by the MPEG 3D video group. The blended mode was not used, thus meaning only one image was used to interpolate the virtual view. The in-painting method[94] is also used in A3, which may induce blur into the disoccluded regions.

- A4: Muller et al.[63] proposed a hole filling method aided by depth information. The corresponding depth values at the hole boundary are examined row-wise to find background color samples to be copied into the hole. This may fail to reconstruct the vertical or oblique structures and complex textures. Some foreground color may be propagated into the hole owing to the depth estimation errors.

- A5: Ndjiki-Nya et al. [64] used a patch-based texture synthesis method to fill the missing part in the virtual view. Since the used patches are rectangular, which may lead to block artifacts and only straight edges could be accurately reconstructed.



(a) PSNR (FR)

(b) SSIM (FR)

(c) IW-SSIM (FR)

(d) IW-PSNR (FR)

(e) BIQI (NR)

(f) NIQE (NR)

(g) Bliinds2 (NR)

(h) VSQA (FR)

(i) MPPSNR (FR)

(j) MWPSNR (FR)

(k) MPPSNRr (FR)

(l) MWPSNRr (FR)

(m) 3DSwIM (FR)

(n) APT (NR)

(o) NIQSV (NR)

(p) NIQSV+ (NR)

Figure 3.8: Scatter plots of DMOS versus DMOSp of each IQA method

- A6: Koppel et al. [42] extended A5 by a background sprite which takes the temporal information into consideration to improve the synthesis.
- A7: holes in virtual views are left unfilled.

For each of the synthesized viewpoints a reference view is available as the chosen virtual viewpoints conrrespond to viewpoints also acquired with a real camera. Fig. 3.9 gives some example results of the proposed NR metrics on the reference image and synthesized images. We can easily observe that the obtained score of the proposed NR quality decreases along with the decrease of image quality (MOS).

3.4.2 PLCC, RMSE and SROCC Performance Comparison

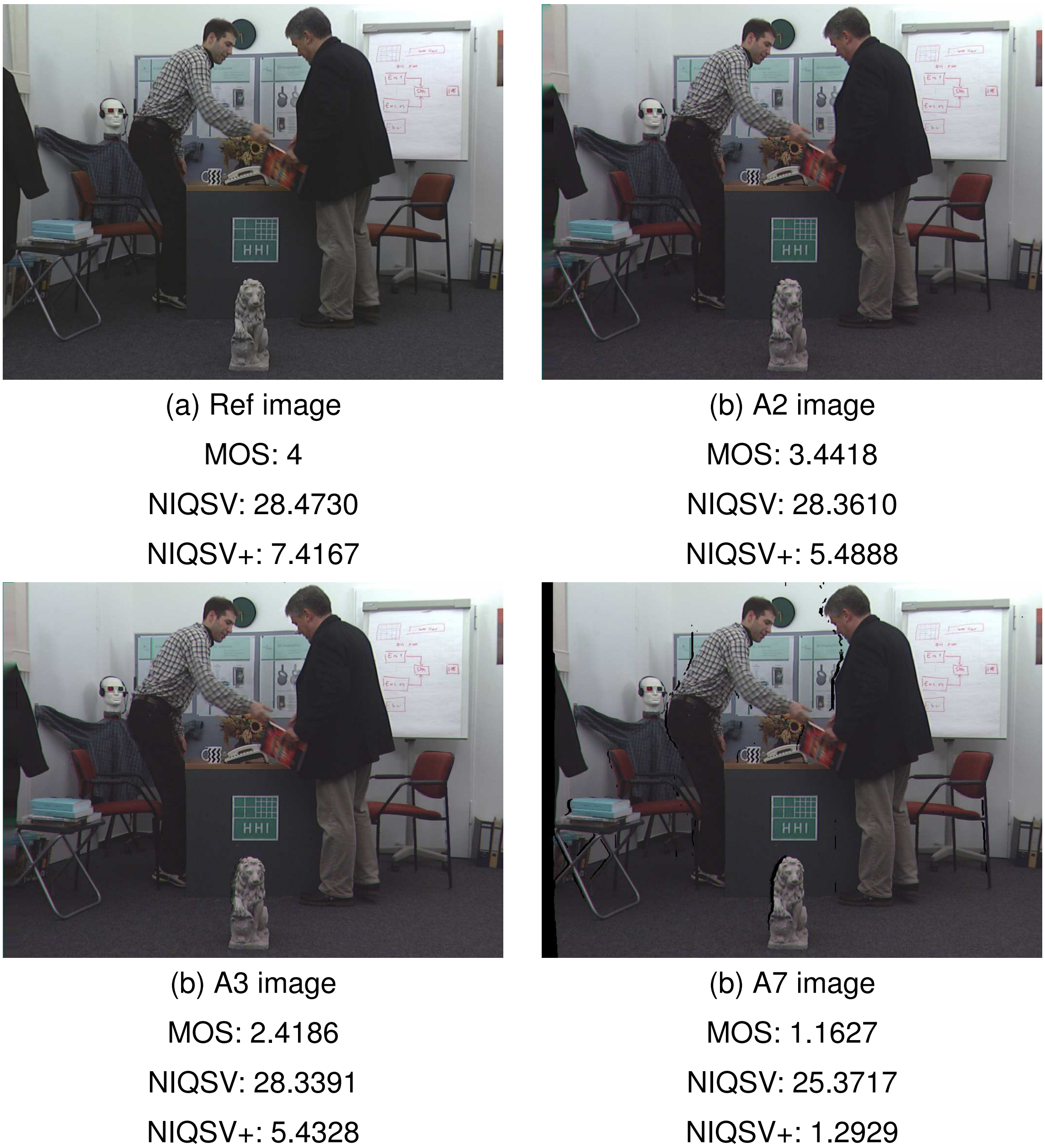
In order to evaluate the performance of the proposed objective IQA method for DIBR-synthesized views, the following methods are also tested for comparison. As a preliminary study, here we only focus on still synthesized images. Thus we did not test the VQA metrics proposed in[117, 52, 84] which include temporal analysis. Due to the lack of depth map in the tested database, the IQA metric proposed in[85], which uses a depth map is not compared in this study either. Below, the proposed metrics NIQSV and NIQSV+ are compared with four full-reference 3D (synthesized view dedicated) metrics, five full-reference 2D metrics and three no-reference 2D metrics. Besides the other metrics introduced in Chapter 1, we compare the proposed metrics with some state-of-the-art 2D image quality metrics, which are presented as follows:

Tested NR 2D metrics include:

- BIQI: Blind Image Quality Index, a Blind/NR objective IQA method proposed by Moorthy et al. in [59].
- BliindSII: BLind Image Integrity Notator using DCT Statistics -II, a Blind/NR objective IQA method proposed by Saad et al. in [71].
- NIQE: Natural Image Quality Evaluator, a Blind/NR objective IQA method proposed by Mittal et al. in [58].

Tested FR 2D metrics include:

- SSIM: Structure SIMilarity, a widely used objective FR IQA metric calculating the structure similarity between the tested image and the reference image proposed by Wang et al. in [112].



(a) Ref image

MOS: 4

NIQSV: 28.4730

NIQSV+: 7.4167

(b) A2 image

MOS: 3.4418

NIQSV: 28.3610

NIQSV+: 5.4888

(b) A3 image

MOS: 2.4186

NIQSV: 28.3391

NIQSV+: 5.4328

(b) A7 image

MOS: 1.1627

NIQSV: 25.3717

NIQSV+: 1.2929

Figure 3.9: Proposed quality metric scores of reference image and images synthesized by A2, A3 and A7

- MS-SSIM: Multi-Scale Structure SIMilarity, a multi-scale approach of SSIM proposed by Wang et al. in [111].
- PSNR: Peak Signal to Noise Ratio, a widely used pixel-based metric.
- IW-PSNR, IW-SSIM: Information content Weighted FR IQA Metric based on PSNR and SSIM separately, proposed by Wang et al. in [110].

Table 3.1: performance dependency of NIQSV+ on k_z and C

k_z	C	PLCC	RMSE	SROCC
100	0	0.7001	0.4754	0.6370
200	0	0.7166	0.4644	0.6591
300	0	0.7141	0.4661	0.6752
100	1	0.7155	0.4652	0.6677
200	1	**0.7274**	**0.4569**	**0.6872**
300	1	0.7214	0.4611	0.6777
100	3	0.7180	0.4634	0.6849
200	3	0.7175	0.4638	0.7047
300	3	0.7032	0.4734	0.7096
100	5	0.7018	0.4743	0.6809
200	5	0.6977	0.4770	0.7000
300	5	0.6840	0.4857	0.7052

For the implementation of these metrics, we used the source code provided in [20], [75], [115], [53], respectively. The execution time of each metric is normalized based on PSNR as shown in Table 3.4.

Table 3.1 gives the performance dependency of NIQSV+ on k_z and C. It shows that the performances of NIQSV+ are optimal when k_z equals 200 and C equals 1. For a fair comparison with the other metrics, we use a cross-validation scenario to obtain the performance of the proposed metric: the adopted database is partitioned into two non-overlapping sets with randomly selected 50% images as a training set and the other 50% as a test set. This random train-test procedure was repeated 100 times and the average performance on the test set across the 100 iterations was reported as the performance of our proposed method. Compared to the best results given in Table 3.1, the results of cross-validation are less favorable (PLCC 0.016 lower), but they are more realistic.

The scatter plots of the $DMOS$ versus the fitted score $DMOS_p$ of all the tested IQA metrics are shown in Fig. 3.8, the PLCC, RMSE, SROCC values are shown in Table 3.2,

Table 3.2: PLCC, RMSE and SROCC between DMOS and objective metrics. (The best three results are marked in bold), NIQSV+_s means NIQSV with stretching detection, NIQSV+_b means NIQSV with black hole detection

Metric		PLCC	RMSE	SROCC
NR 3D metrics	**NIQSV+**	**0.7114**	**0.4679**	**0.6668**
	NIQSV+_s	0.6886	0.4828	0.6497
	NIQSV+_b	0.6423	0.5103	0.4806
	NIQSV	0.6346	0.5146	0.6167
	APT	**0.7307**	**0.4546**	**0.7157**
FR 3D metrics	3DSwIM	0.6864	0.4842	0.6125
	MP-PSNR	0.6729	0.4925	0.6272
	MP-PSNRr	**0.6954**	**0.4784**	**0.6606**
	MW-PSNR	0.6200	0.5224	0.5739
	MW-PSNRr	0.6625	0.4987	0.6232
	VSQA	0.6122	0.5265	0.6032
FR 2D metrics	PSNR	0.4557	0.5927	0.4417
	SSIM	0.4348	0.5996	0.4004
	MS-SSIM	0.5406	0.5602	0.5021
	IW-PSNR	0.3608	0.6210	0.3460
	IW-SSIM	0.5337	0.5631	0.4795
NR 2D metrics	NIQE	0.4022	0.6096	0.3673
	BIQI	0.5273	0.5657	0.3555
	BliindSII	0.5331	0.5633	0.1800

Table 3.3: Ranking of view synthesis algorithms according to DMOS and objective metrics. The green colored algorithms indicate that the changing of ranking position is acceptable; the blue ones are medium; and the red ones are non acceptable.

Metric		Ranking of synthesis algorithms						
DMOS		**A1**	**A5**	**A4**	**A6**	**A2**	**A3**	**A7**
		3.57	**3.49**	**3.40**	**3.32**	**3.31**	**3.15**	**2.28**
NR 3D metrics	**NIQSV+**	A1	A6	A5	A4	A2	A3	A7
	NIQSV	A1	A4	A5	A2	A6	A3	A7
	APT	A1	A2	A4	A3	A5	A6	A7
FR 3D metrics	3DSwIM	A1	A4	A5	A6	A3	A2	A7
	MP-PSNR	A4	A5	A6	A3	A2	A1	A7
	MP-PSNRr	A4	A5	A6	A3	A2	A1	A7
	MW-PSNR	A4	A5	A6	A2	A3	A1	A7
	MW-PSNRr	A4	A5	A6	A2	A3	A1	A7
	VSQA	A6	A5	A4	A3	A2	A7	A1
FR 2D metrics	PSNR	A6	A5	A4	A3	A2	A7	A1
	SSIM	A3	A4	A5	A6	A2	A7	A1
	MS-SSIM	A3	A4	A5	A6	A2	A7	A1
	IW-PSNR	A6	A5	A4	A3	A2	A7	A1
	IW-SSIM	A4	A6	A5	A3	A2	A7	A1
NR 2D metrics	BIQI	A1	A2	A5	A4	A6	A3	A7
	BliindSII	A1	A2	A3	A4	A5	A6	A7
	NIQE	A1	A2	A3	A4	A5	A6	A7

from which we can see that the NIQSV metric performs much better than PSNR and SSIM and achieves very closely to the other three full-reference metrics: 3DSwIM, MW-PSNR and MP-PSNR, the SROCC value is even a little better than 3DSwIM. And compared to NIQSV, the extended version NIQSV+ improves the performance a lot with additional steps (detection of black holes and stretching).

Here, the proposed metric NIQSV+ and APT have the best performances in terms of PLCC, RMSE and SROCC, which indicates that they have the best accuracy and monotonicity estimation compared to FR metrics, even though they are NR metrics. APT performs a little better than our proposed method (PLCC 0.019 higher), but the proposed method executes much faster cf. Table 3.4.

Table 3.4: Execution time of each IQA metric normalized base on PSNR. The metrics A-Z indicate NIQSV+, NIQSV, APT, 3DSwIM, MP-PSNR, MP-PSNRr, MW-PSNR, MW-PSNRr, VSQA, PSNR, SSIM, IW-PSNR, IW-SSIM, NIQE, BIQI and BliindS2 respectively.

Metric	A	B	C	D	E	F	G	H
time	21	18	13k+	90	100	35	12.4	9.6
Metric	I	J	K	L	M	N	O	P
time	140	1	7.4	75	75	45	67.5	6.8

Table 3.5: Variance of residuals between $DMOS$ and $DMOS_p$. (Res. Var. denotes Residual Variance), the metrics A-Z indicate NIQSV+, NIQSV, APT, 3DSwIM, MP-PSNR, MP-PSNRr, MW-PSNR, MW-PSNRr, VSQA, PSNR, SSIM, IW-PSNR, IW-SSIM, NIQE, BIQI and BliindS2 respectively.

Metric	A	B	C	D	E	F	G	H
Residual Variance	**0.222**	0.268	**0.212**	0.237	0.246	**0.232**	0.276	0.2112
Metric	I	J	K	L	M	N	O	P
Residual Variance	0.281	0.356	0.364	0.390	0.321	0.376	0.324	0.321

3.4.3 Ranking performance comparison

In order to further compare the performance of these image quality metrics, we also noted the ranking of the synthesis algorithms according to the DMOS and the objective metric scores in this database. The ranking is based on the average quality score of the views synthesized by the corresponding algorithm (from A1 to A7). As shown in Table 3.3, the first two lines offer the rankings according to DMOS which can be

Table 3.6: Statistical significance table based on residuals between model predictions $DMOS_p$ and $DMOS$, The symbol ”1” indicates that the statistical performance of the IQA metric in the row is significantly superior to the one in the column, the symbol ”-1” means the opposite, while ”0” indicates that there is no significant difference between the metrics in the row and in the column. The metrics A-Z indicate NIQSV+, NIQSV, APT, 3DSwIM, MP-PSNR, MP-PSNRr, MW-PSNR, MW-PSNRr, VSQA, PSNR, SSIM, IW-PSNR, IW-SSIM, NIQE, BIQI and BliindS2 respectively.

Metric	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
A	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
B	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0
C	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
D	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
E	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0
F	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
G	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
H	0	0	0	0	0	0	0	0	0	1	1	1	0	1	0	0
I	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
J	-1	0	-1	-1	-1	-1	0	-1	0	0	0	0	0	0	0	0
K	-1	-1	-1	-1	-1	-1	0	-1	0	0	0	0	0	0	0	0
L	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0
M	-1	0	-1	-1	0	-1	0	0	0	0	0	0	0	0	0	0
N	-1	-1	-1	-1	-1	-1	-1	-1	-1	0	0	0	0	0	0	0
O	-1	0	-1	-1	0	-1	0	0	0	0	0	0	0	0	0	0
P	-1	0	-1	-1	0	-1	0	0	0	0	0	0	0	0	0	0

regarded as the ground truth and the average DMOS score of each algorithm, and the following lines give the rankings based on the tested objective metrics. The green colored algorithms indicate that the changing of ranking position is acceptable; the blue ones are medium; and the red ones are non acceptable. It can be noticed that the proposed metric NIQSV ranks very closely to DMOS scores except A5/A4 and A6/A2, the proposed metric NIQSV+ ranks the most closely to DMOS except for A6, compared to the rankings of the other tested metrics. Since one of the most important roles of the IQA metric is to provide the same rank order as a human does among different image processing algorithms, this result shows the proposed method has a desirable character as an IQA metric.

3.4.4 Statistical significance test

An F-test is used to examine the statistical significance between each tested IQA method. It is based on the residuals between model predictions $DMOS_p$ and the $DMOS$ values, as described in

$$Res(i) = DMOS_p(i) - DMOS(i) \tag{3.24}$$

where $Res(i)$ denotes the residual of each IQA metric. Table 3.5 gives the variances of residuals. The statistical significance results are shown in Table 3.6. The symbol ”1” indicates that the statistical performance of the IQA metric in the row is significantly superior to the one in the column, the symbol ”-1” means the opposite, while ”0” indicates that there is no significant difference between the metrics in the row and in the column. It can be noticed that the proposed IQA method is significantly superior to all the 2D metrics both FR and NR, and has no significant difference compared to the tested 3D FR and NR methods.

3.4.5 Analysis of failure cases

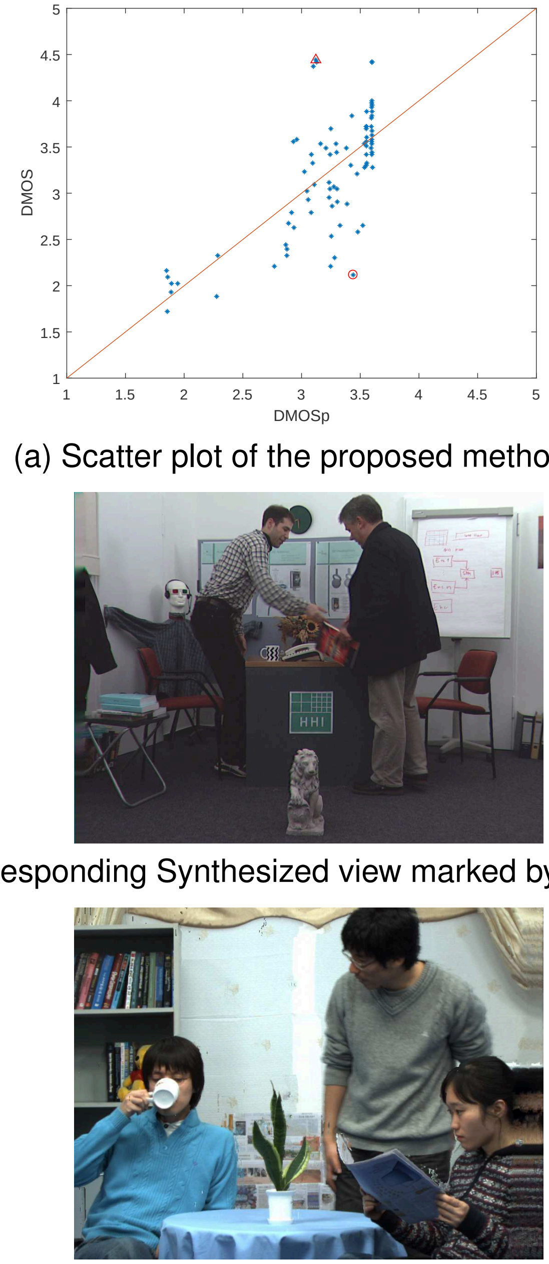
It can be noticed in the scatter plot of the proposed method that some images with very low $DMOS$ obtain a high objective quality score by the proposed method, while some images with very high $DMOS$ obtain relatively low predicted scores. Fig. 3.10 gives two examples of these images, where the positions of these synthesized images in the scatter plot are marked by a red circle and a red triangle.

The main distortions of the image in (b) of Fig. 3.10 are stretching on the left side, object shifting of the posters on the wall and crumbling around the the chair leg. One reason for its extremely high subjective score could be that the subjects focused on the center and foreground objects of the image, thus they did not notice these distortions.

Considering the image in (c) of Fig. 3.10, its main distortions are stretching and some thin distortions around the object edges. While these thin distortions around the edges can be easily detected by opening and closing operations, the stretching with similar texture to that of a complex occluded area (eg. clothes and hair) is hard to detect without a reference image.

3.5 Conclusion

In this chapter, we proposed two totally No-reference IQA metrics NIQSV and NIQSV+ for 3D synthesized views by measuring the blurry regions, crumbling, black holes and stretching. The experimental results show that the proposed metric NIQSV outperforms the traditional 2D metrics and approaches the results of 3D synthesized view aimed full reference metrics very closely. The metric NIQSV+ significantly outperforms the widely used FR 2D IQA metrics (SSIM, PSNR and MS-SSIM), the NR 2D IQA metrics (BIQI, BliindSII, NIQE). Compared to the state-of-the-art DIBR dedicated FR and NR metrics, the proposed NIQSV+ comes in the second place (slightly less favorable than the APT) in terms of PLCC, RMSE and SROCC, while there is no significant difference between the DIBR dedicated metrics. Compared to NIQSV, the proposed metric achieves a gain of 7.68% in correlation with subjective measurements (PLCC). In terms of their approximation of human ranking, the proposed metric achieves the best performance in the experimental test. Moreover, since the morphological operators only contain integer operations, while black hole counting and stretching detection only contain pixel operations, the proposed metric bears very low computational complexity. That is why it is much faster than the APT. All these characteristics of the proposed method make it promising not only for the benchmark application but also for the algorithm optimization application (eg. it can be integrated into the 3D image compression schema for a better perceptual rate-distortion control). In this chapter, we do not consider all the distortions which may occur in the DIBR applications, only blurry regions, crumbling and stretching are taken into account, it still can not achieve a satisfactory results. In the next chapter, we propose two FR quality metrics which achieve much better results.



(a) Scatter plot of the proposed method

(b) Corresponding Synthesized view marked by red triangle

(c) Corresponding Synthesized view marked by red circle

Figure 3.10: Bad performance images

CHAPTER 4

A BENCHMARK OF DIBR-SYNTHESIZED VIEW QUALITY ASSESSMENT METRICS

Several objective quality metrics have been presented in the previous chapters, now we focus on a new DIBR database to better benchmark the existing DIBR-synthesized view quality assessment quality. In this chapter, we firstly introduce the existing DIBR related databases, then we present a new DIBR-synthesized image database with the associated subjective scores. This work focuses on the distortions only induced by different DIBR synthesis methods which determine the quality of experience (QoE) of these DIBR related applications. Seven state-of-the-art DIBR algorithms, including inter-view synthesis and single view based synthesis methods, are considered in this database. The quality of synthesized views was assessed subjectively by 41 observers and objectively using 14 state-of-the-art objective metrics. Subjective test results show that the interview synthesis methods, having more input information, significantly outperform the single view based ones. We also conduct a relatively complete bench marking of the state-of-the-art objective metrics for DIBR-synthesized image quality assessment on this database. Correlation results between the tested objective metrics and the subjective scores on this database reveal that further studies are still needed for a better objective quality metric dedicated to the DIBR-synthesized views.

This chapter is organized as follows. The existing DIBR databases are firstly introduced in the Section 4.1. Section 4.2 states briefly the main contributions of this database. Section 4.3 introduces the seven DIBR algorithms used in this database in detail. The subjective experiments and the objective metrics performance study are described in Section 4.4 and Section 4.5 respectively. In the end, the conclusions are drawn in Section 4.6.

4.1 Introduction to existing DIBR databases

There exist several DIBR related databases, as shown in Table 4.1. Each database has its own focus. The IVC databases focus on the distortions caused by different DIBR synthesis algorithms, the MCL-3D and SIAT database investigate the influence of traditional 2D distortions of original texture and depth map on the DIBR-synthesized views.

4.1.1 IRCCyN/IVC DIBR image and video databases

Bosc et al. proposed the IRCCyN/IVC DIBR image database [37, 12] and IRCCyN/IVC DIBR video database [13]. The source content of these two database is extracted from 3 different MVD sequences: $BookArrival$, $Lovebird1$ and $Newspaper$. For each sequence, four virtual views are synthesized by seven DIBR view synthesis algorithms [23, 94, 62, 63, 64, 65, 42]. That is to say, there are 84 synthesized images or videos in each database. For each synthesized virtual image or video, the image/video captured by a real camera on the same viewpoint is used as reference. One big issue is that several DIBR algorithms tested in this database introduce some “old-fashioned” artifacts (such as “black holes”) which no longer exist when the state-of-the-art DIBR algorithms are used.

4.1.2 MCL-3D database

Song et al. proposed a publicly accessible stereoscopic 3D Database (MCL-3D database) for the quality assessment of DIBR-synthesized stereoscopic images in [87]. The DIBR technology is used to generate the left and the right views by using the 2D-image-plus-depth source. Many types of distortions are considered in this database, such as Gaussian blur, additive white noise, down-sampling blur, JPEG and JPEG-2000 (JP2K) compression and transmission error. These distortions are applied on either the original texture images or the depth images before the view synthesis. Nine MVD sequences are collected, among which $Kendo$, $Lovebird1$, $Balloons$, $PoznanStreet$ and $PoznanHall2$ are natural images; $Shark$, $Microworld$, $GTFly$ and $Undodancer$ are Computer Graphics images.

Table 4.1: Summary of existing DIBR related database

Name	No. seq.	No. DIBR algos	DIBR algos		other distortions	size	Reference	display
			Name	Year				
IVC DIBR-image	3	7	Fehn's	2004	No	84	original	2D
			Telea's	2003				
			VSRS	2009				
			Müller	2008				
			Ndjiki-Nya	2010				
			Köppel	2010				
			Black hole	—				
IVC DIBR-video	3	7	idem		H.264	84	original	2D
MCL-3D	9	4	Fehn's	2004	Additive White Noise	693	synthesized	Stereo.
			Telea's	2003	Blur			
			HHF	2012	Down sampling			
			Black hole	—	JPEG			
					JPEG2k			
					Translation Loss			
SIAT video	10	1	VSRS	2009	3DV-ATM coding	140	original	
IVY	7	4	Criminisi	2004	No	84	original	Stereo.
			Ahn's	2013				
			VSRS	2009				
			Yoon	2014				
Proposed	10	7	Criminisi	2004	No	140	original	2D
			VSRS	2009				
			LDI	2011				
			HHF	2012				
			Ahn's	2013				
			Luo's	2016				
			Zhu's	2016				

Four DIBR view synthesis algorithms ([23, 94, 86] plus DIBR without hole filling) were used. This database contains various types of distortions, but distortion types directly related to DIBR algorithms are very limited. The tested DIBR algorithms also produce some “old-fashioned” artifacts.

4.1.3 SIAT DIBR video database

The SIAT Synthesized Video Quality Database [52] proposed by Liu et al. focused on the distortions introduced by compressed texture and depth images. For each of the ten different MVD sequences, 14 different texture/depth quantization combinations were used to generate the texture/depth view pairs with compression distortions. Then, the virtual videos are synthesized using the VSRS-1D-Fast software implemented in the 3D-HEVC [89] reference software HTM. Here, only compression distortions are evaluated.

4.1.4 IVY stereoscopic database

Jung et al. proposed another IVY stereoscopic 3D image database to assess the quality of DIBR synthesized stereoscopic images [41]. A total of 7 sequences are selected from four Middlebury datasets [81] ($Aloe$, $Dolls$, $Reindeer$, and $Laundry$) and three MVD sequences ($Lovebird1$, $Newspaper$ and $Bookarrival$). 84 stereo image pairs are synthesized by four DIBR algorithms [17], [1], [92], [113] in this database. Note that in this database, virtual views were only generated by view extrapolation.

4.2 Contribution of our proposed database

In this work, the proposed image database focuses on the distortions only caused by DIBR algorithms (like the IRCCyN/IVC DIBR database), but with state-of-the-art DIBR algorithms. The “view" or stimuli in this subjective test indicates an individual synthesized image. In total, we tested seven DIBR algorithms, including both the interview synthesis and the single view synthesis methods. We selected those DIBR algorithms which produce no longer “old-fashioned” artifacts and of which the code sources were provided by their authors. Note that the SIAT database focuses on the effect of texture and depth compression on the synthesized views and it contains only one DIBR algorithm. Compared to the MCL-3D and the IVY databases, the proposed new database (1) includes not only virtual views generated by view extrapolation, but also by view interpolation; (2) tests more and newer DIBR algorithms; (3) shows the views on a 2D display to avoid the 3D display settings and configurations influences (same approach was used in the IRCCyN/IVC DIBR database). The IRCCyN/IVC DIBR database also focuses on the comparison of different DIBR algorithms, but it contains some "old-fashioned" DIBR artifacts (eg. black holes) and it contains less source images than ours. The proposed database can be used along with the IRCCyN/IVC database for this type of usage. To sum up, the main contributions with the proposed DIBR database are: (1) a new publicly accessible DIBR synthesized image quality database with more recent DIBR algorithms; (2) a relatively complete bench marking of the state-of-the-art objective metrics for DIBR synthesized image quality assessment.

4.3 Tested DIBR algorithms

In this section, the tested DIBR algorithms are introduced. As introduced in Chapter 1, due to the lack of original texture information, a synthesized image often contains disocclusion holes which significantly degrades the quality. The processing of these disocclusion holes plays an important role in generating a synthesized view of high quality. Here, both inter-view interpolation and single view synthesis methods are taken into consideration. The interview DIBR algorithm uses the two neighboring views to synthesize the virtual viewpoint, while the single view synthesis methods only use one neighboring view to extrapolate the synthesized view.

4.3.1 Criminisi's Examplar based inpainting

Criminisi et al. proposed a new algorithm for $image\ inpainting$. As shown in Fig. 4.1, it employs an exemplar-based texture synthesis technique [17]. A $confidence$ is used to compute patch priorities, and to optimize the fill order of the target regions according to their priorities. The actual color values are computed using exemplar-based synthesis. After the target patch has been filled with new values, the $confidence$ in this patch is updated. The $confidence$ in the synthesized pixel values is propagated in a manner similar to the propagation of information in inpainting.

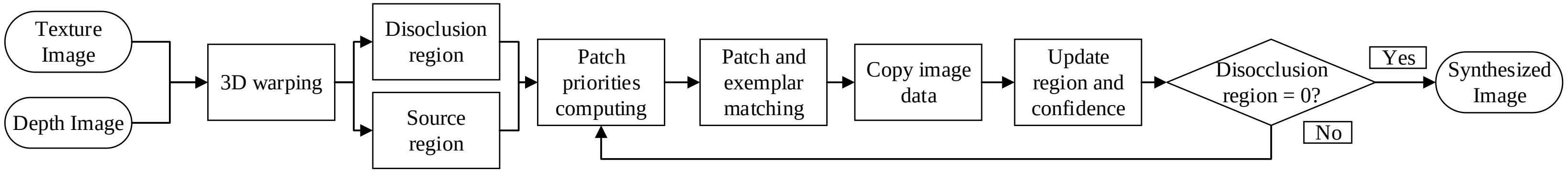


Figure 4.1: Block diagram of Criminisi

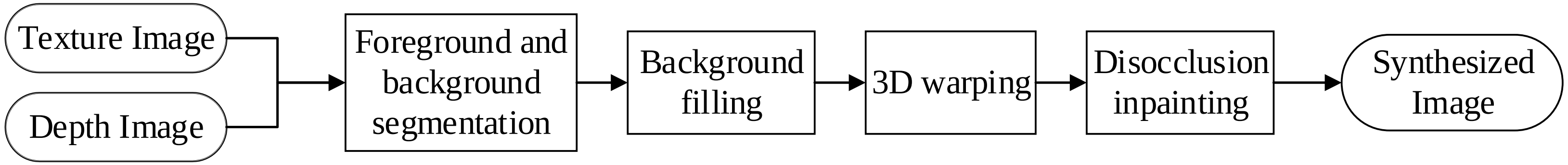


Figure 4.2: Block diagram of LDI

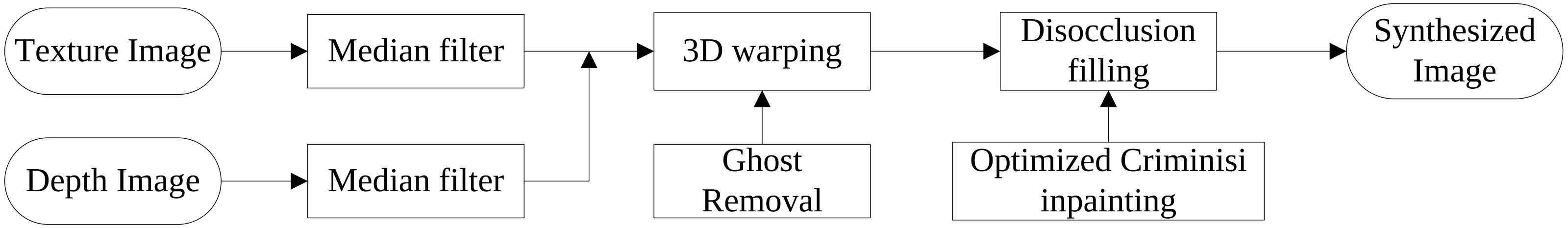


Figure 4.3: Block diagram of Ahn's method

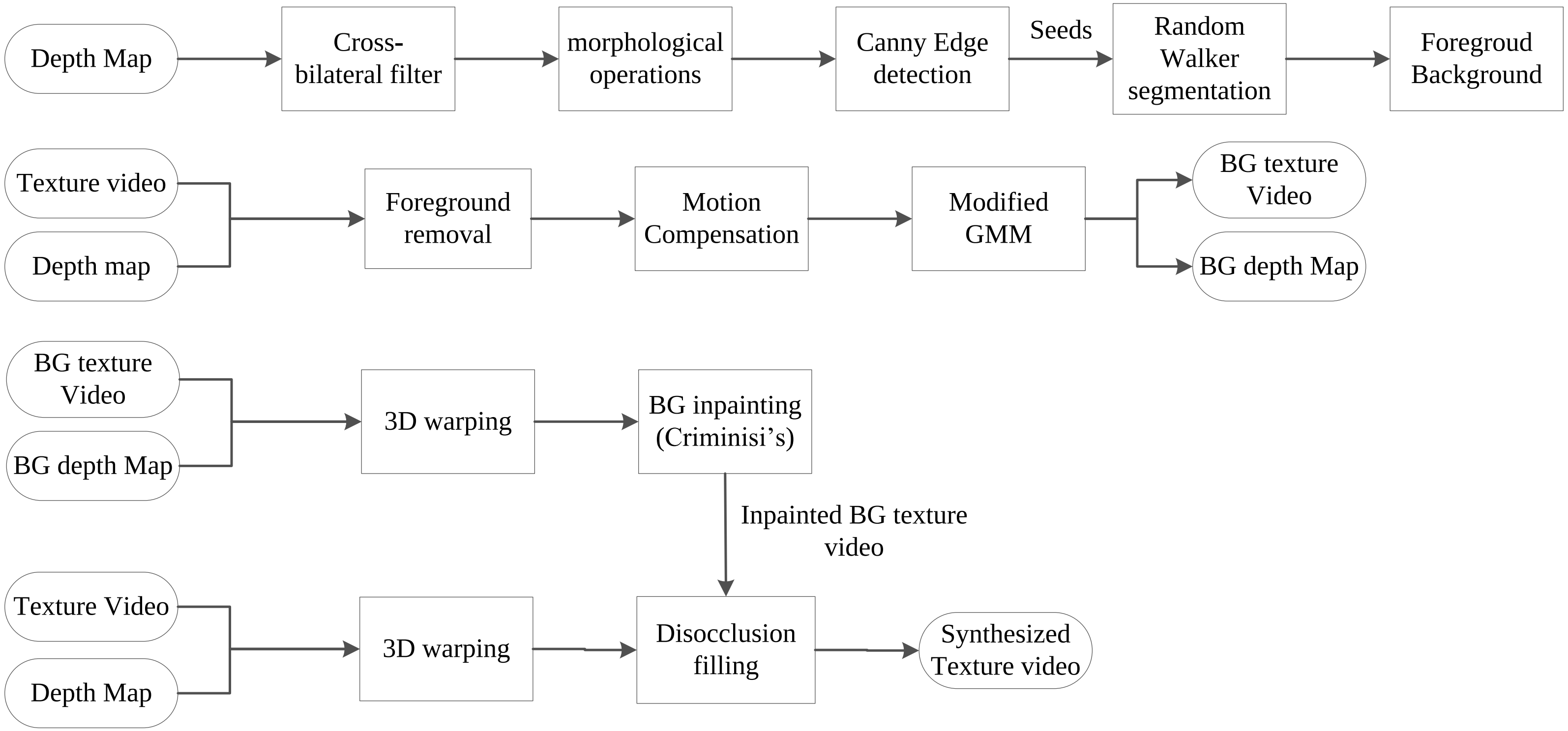


Figure 4.4: Block diagram of Luo's method

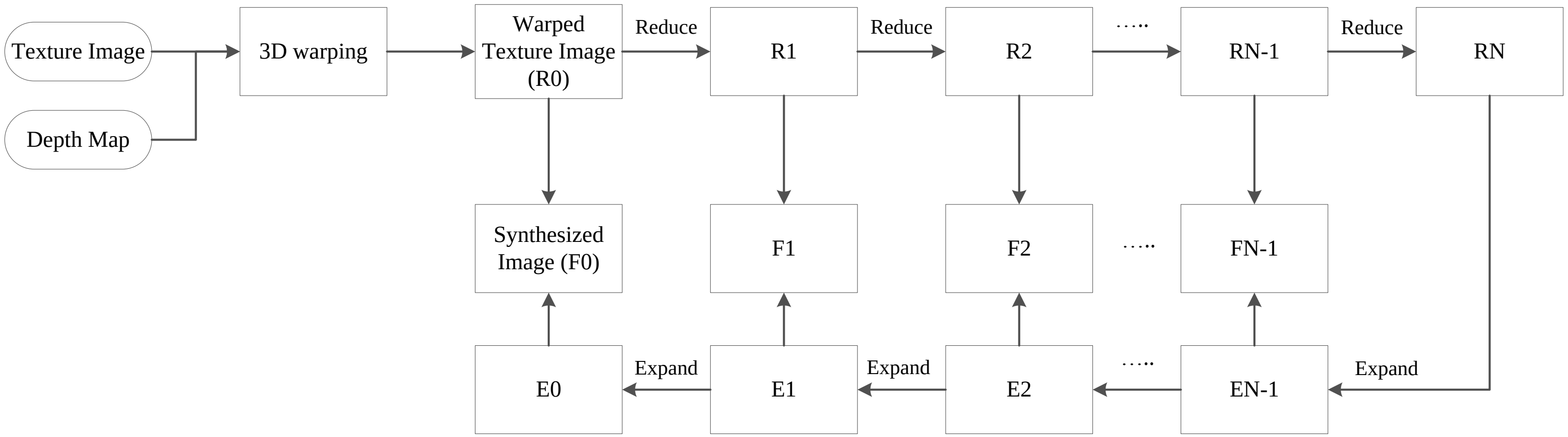


Figure 4.5: Block diagram of HHF

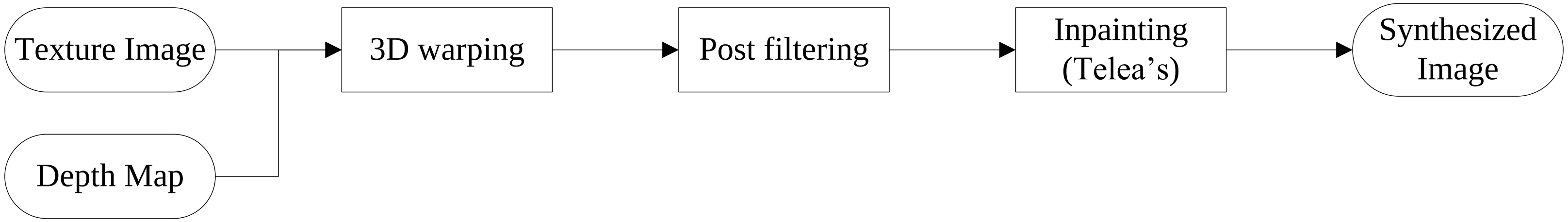


Figure 4.6: Block diagram of VSRS

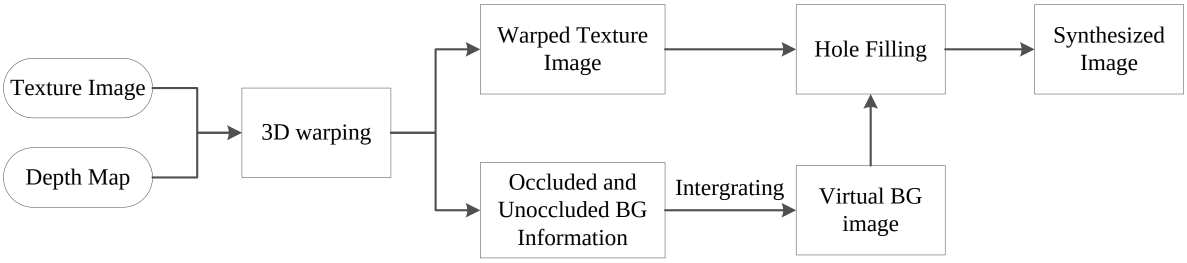


Figure 4.7: Block diagram of Zhu's method

As filling proceeds, confidence values decay, which indicates that the pixel color values are less reliable near the center of the target region.

4.3.2 LDI

Jantet et al. proposed an object-based Layered Depth Image (LDI) representation to improve the quality of virtual synthesized views [38]. As shown in Fig. 4.2, they firstly segment the foreground and background based on a region growing algorithm, which allows organising LDI pixels into two object-based layers. Once the extracted foreground is obtained, an inpainting method is used to reconstruct the complete background image on both depth and texture images. Several inpainting method can be chosen, for example, Navier-Stoke based method [7], Telea method [94] and Gautier method [45]. In this work, the Gautier inpainting method is used in the LDI.

4.3.3 Ahn's method

Ahn et al. proposed a depth based disocclusion filling method using patch-based texture synthesis [1]. Firstly, a median filtering is applied to texture and depth images to remove the small cracks caused by rounding errors in the 3D warping process. In order to handle the ghost effect due to mismatch of the boundaries of the foreground objects in the texture and depth image, a ghost effect removal method is added in the 3D warping process. During the $disocclusion\ inpainting$ procedure, the Criminisi's method is improved by optimizing the filling priority and the patch-matching measure. The new priority term uses the Hessian matrix structure tensor which is robust to noise and reflects the overall structure of an image area. The optimized matched patch is selected through the data term on the background regions which were extracted using warped depth map. The filling of disoccluded holes in a depth map is conducted simultaneously with filling holes in the texture image. The block diagram of Ahn's view synthesis method is shown in Fig. 4.3.

4.3.4 Luo's method

Luo et al proposed a hole filling approach for DIBR systems based on background reconstruction [54]. As shown in Fig. 4.4, in order to extract the foreground, the depth map is firstly preprocessed by a cross-bilateral filter and morphological operations. Then the Canny's edge detection is employed to extract the initial seeds for random walker, and the foreground is finally extracted from the depth map by random walker segmentation. After the removal of foreground, the temporal correlation information in both the 2D video and its corresponding depth map is exploited to construct a background video based on motion compensation and modified Gaussian Mixture model. Finally, the reconstructed background video is warped to the virtual viewpoint to eliminate the disocclusion holes.

4.3.5 HHF-Hierarchical hole-filling

Solh et al. proposed two pyramid-like approaches, namely Hierarchical Hole-Filling (HHF) and Depth Adaptive Hierarchical Hole-Filling, to eliminate the disoccluded holes in DIBR synthesized views [86]. The block diagram of HHF is shown in Fig. 4.5, which can be divided into four steps. Firstly, a sequence of images R_0,..., R_N are low-pass filtered using a pseudo Gaussian plus zero elimination filtering operation (reduce), in which the original 3D warped image is marked as R_0. R_1 is the $reduced$ version of R_0, and so on. The Gaussian pyramid is generated by this $reduce$ operation when the holes do not influence the calculations. Secondly, they start from the highest level of this pyramid R_N, an $Expand$ operation is utilized to get an interpolated image E_{N-1}, whose size is equal to R_{N-1}. Then, this interpolated image E_{N-1} is used to fill the disoccluded holes in R_{N-1} to obtain the filled image F_{N-1}. Finally, the filled image in each scale, F_{N-1}, ... , F_0 can be obtained by repeating the operations upon, and F_0 is the final inpainted result. The DAHHF method adds a depth adaptive preprocessing before the $reduce$ and $expand$ operations. Since the disoccluded regions are more likely to be the background regions, a depth map is employed to assign higher weights to the pixels belonging to the background. The following steps are similar to HHF except that the starting image is the preprocessed image and the depth weight must be considered during the $Fill$ operations.

4.3.6 VSRS-View Synthesis Reference Software

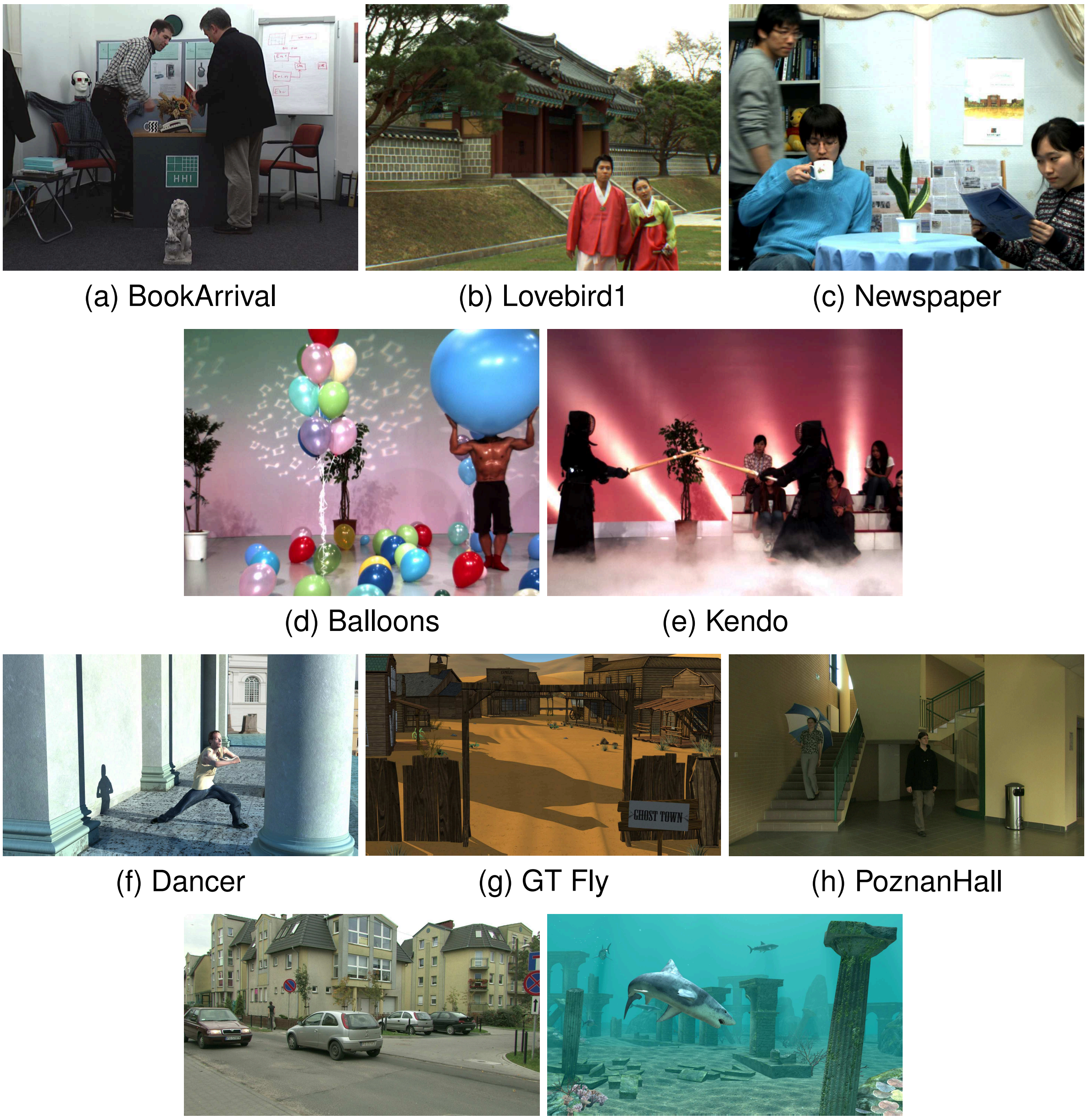
Tanimoto et al. proposed a DIBR method [62] which has been adopted by the MPEG 3D video Group, known as View Synthesis Reference Software (VSRS) [92]. The depth discontinuity artifacts are solved by performing a post-filter on the projected depth map. Then, the inpainting method proposed in [94] is used to fill the holes in the disoccluded regions. This approach is primarily used in the inter-view synthesis applications which have just small holes to be filled, but it can also be used in single view based rendering cases. In this paper, both the interview mode (VSRS2) and the single view based mode (VSRS1) are used.

4.3.7 Zhu's method

Zhu et al. proposed a novel depth-enhanced hole filling approach for DIBR view interpolation [120]. Instead of inpainting the warped images directly, they focus on the use of the occluded information to identify the relevant background pixels around the holes. Firstly, the occluded background information is registered in both texture and depth during the 3D warping process, and the background pixels around the holes are found. Then, the unoccluded background information around the holes is extracted based on the depth map. After that, a virtual image is generated by integrating the occluded background and unoccluded background information. The disoccluded holes are filled based on this generated image with the help of a depth-enhanced Criminisi's inpainting method and a simplified block-averaged filling method. Finally, the pre-stored foreground information is recovered in the virtual synthesized image.

Among the DIBR algorithms mentioned above, Zhu's method is an interview synthesis method, VSRS is used both as interview synthesis and single view based synthesis (marked as VSRS2 and VSRS1 recpectively in this paper), the others are only single view based synthesis methods.

4.4 Subjective Experiment



(a) BookArrival (b) Lovebird1 (c) Newspaper

(d) Balloons (e) Kendo

(f) Dancer (g) GT Fly (h) PoznanHall

(i) Pozan Street (j) Shark

Figure 4.8: The used MVD sequences

Ten MVD test sequences provided by MPEG for the 3D video coding are used in this experiment. The *Balloons*, *BookArrival*, *Kendo*, *Lovebird1*, *Newspaper*, *Poznan Street* and *PoznanHall* sequences are natural images while the *Undo Dancer*, *Shark* and

Table 4.2: Introduction of the tested MVD sequences

Sequence	Resolution	Frame No.	View ref. Position	View sys. Position	SI
BookArrival	1024 × 768	58	8, 10	9	60.2348
Lovebird1	1024 × 768	80	4, 8	6	64.9756
Newspaper	1024 × 768	56	2, 6	4	61.1012
Balloons	1024 × 768	6	1, 5	3	47.6410
Kendo	1024 × 768	10	1, 5	3	48.6635
Undo Dancer	1920 × 1088	66	1, 9	5	64.1033
GT Fly	1920 × 1088	150	1, 9	5	55.5549
Poznan street	1920 × 1088	26	3, 5	4	61.3494
Poznan Hall2	1920 × 1088	150	5, 7	6	23.5174
Shark	1920 × 1088	220	1, 9	5	48.6635

$Gt\ Fly$ are computer animation images, as shown in Fig. 4.8. The characteristics of the sequences are summarized in Table 4.2.

For each single view based DIBR algorithm, a single virtual viewpoint is extrapolated from the neighboring two views separately. For the interview DIBR algorithms, the virtual viewpoint is synthesized based on both the two neighboring views, as shown in Table 4.3. We consider thus for each reference image, 2 virtual views synthesized by 2 interview synthesis algorithms and 12 virtual views synthesized by 6 single view based DIBR algorithm, which leads to 14 degraded images.

Table 4.3: Type of DIBR method

DIBR method	inter-view or single view (extrapolation)
VSRS2	inter-view
Zhu's	inter-view
Criminisi's	single view (extrapolation)
Luo's	single view (extrapolation)
HHF	single view (extrapolation)
LDI	single view (extrapolation)
VSRS1	single view (extrapolation)
Ahn's	single view (extrapolation)

4.4.1 Subjective Test Methodology

There are several subjective testing methods to obtain the perceived quality scores, such as the subjective assessment methodology for video quality (SAMVIQ) [8], the absolute categorical rating (ACR), etc. In this test, we choose to follow the SAMVIQ protocol because of its stability, reliability and relatively higher discriminability. The SAMVIQ results have a greater accuracy than the ACR scores for the same number of observers (on average 30% fewer observers were required for SAMVIQ than ACR for the same level of accuracy) [6].

In the SAMVIQ protocol, there is much more freedom for the observers who can view each image several times and correct the notation at any time they want. The observers can compare the degraded versions with each other, as well as with the explicit reference. In each trial, there is also a hidden reference which helps to evaluate the intrinsic quality of the reference when the perceived quality of the reference is not perfect. A continuous quality rating scale ranging from 0 to 100 is used during the test. It can be categorized according to the five quality levels: Bad, Poor, Fair, Good and Excellent. (See Table 4.4) The experiment was conducted on a NEC MultiSync PA322UHD monitor with resolution 3840×2160. The environment of the subjective experiment was controlled as recommended in the ITU-R Rec. BT.1788 [36].

Table 4.4: Comparison scale for SAMVIQ

10	Bad
30	Poor
50	Fair
70	Good
90	Excellent

Altogether, 42 naive observers (28 males and 14 females with an age varying from 19 to 52 years old) participated in the subjective assessment experiment. All the observers have no prior knowledge of the view synthesis methodology domain. Prior to the test, the observers were screened for normal visual acuity on the Snellen chart, and for normal colour vision using the Ishihara chart. A training session was conducted before the test session. The observers could have a rest at any time they want during the test. The total duration of the experiment varied from 30 to 45 minutes for each observer.

4.4.2 Processing of Subjective Scores

The subjective scores were firstly processed using the observer screening method recommended in the ITU-R Rec. BT.1788 [36]. In this experiment, only one observer is eliminated after the observer screening. That leads to 41 observers finally for this database.

The primary quality scores of the tested image are obtained as the difference between the score of the hidden reference image and the score of the tested image as shown in Eq. (4.1):

$$S_{i,j} = Score_{hr,j} - Score_{i,j} \tag{4.1}$$

where $S_{i,j}$ denotes the primary quality score of the ith tested synthesized image, $Score_{hr}$ and $Score_i$ denote the score of the hidden reference and the ith tested synthesized image respectively, and the subscript j denotes the jth observer.

Then, the primary quality scores are normalized to z-score per person cf. Eq. 4.2.

$$Zscore_{i,j} = \frac{S_{i,j} - \mu_j}{\sigma_j} \tag{4.2}$$

where μ_j and σ_j denotes the mean value and variance value of the jth observer respectively. To make the data more intuitive, the normalized zscores are scaled to (0,1).

The final quality score differential mean opinion score (DMOS) is calculated by averaging the normalized z-scores of all the observers, as shown in Eq. 4.3:

$$DMOS_i = \sum_{j=1}^{N} Zscore_{i,j}/N \tag{4.3}$$

where $DMOS_i$ denotes the final subjective quality score of the ith tested synthesized image, $S_{i,j}$ is the obtain primary quality score in Eq. (4.1), and N is the number of observers.

The obtained $DMOS$ score distributions and their confidence intervals are shown in Fig. 4.9. Generally, the interview synthesis methods outperform the single view based synthesis methods in most sequences. However in some sequences, such as $BoolArrival$, the VSRS1 get better results than VSRS2 and Zhu's methods, but not very significantly according to the corresponding confidence intervals. One reason could be that, owing to the inaccuracy of depth map, the same object in the two base views are rendered to different positions which results in a “ghost” effect in the synthesized view. However, this situation does not happen in single view based synthesis method VSRS1. As shown in Fig. 4.10, there exists a “ghost” effect of the “chat flow” on the board marked by red blocks in (c) and (d); but according to the synthesized content marked by red circles, the interview synthesis methods (c), (d) works better than the single view based ones (a), (b) in generating the object texture.

A statistical analysis (student T-test here) was also made over the obtained $DMOS$ scores, to show the statistical equivalence information of the tested algorithms. The scores of single view based methods are obtained by averaging the scores of the two images synthesized from the viewpoints at the two sides. As shown in Table 4.5, the view interpolation methods (VSRS2 and Zhu's), which use the two neighboring views as reference views, perform much better than the single view based methods. Among the single view based approaches, VSRS1 and Ahn's methods are significantly superior to the others.

Table 4.5: Student T-test with obtained $DMOS$ scores, where the symbol 1 indicates that the DIBR synthesis method in the row is significantly superior to the one in the column, the symbol -1 means the opposite, while 0 indicates that there is no significant difference between the DIBR synthesis methods in the row and in the column.

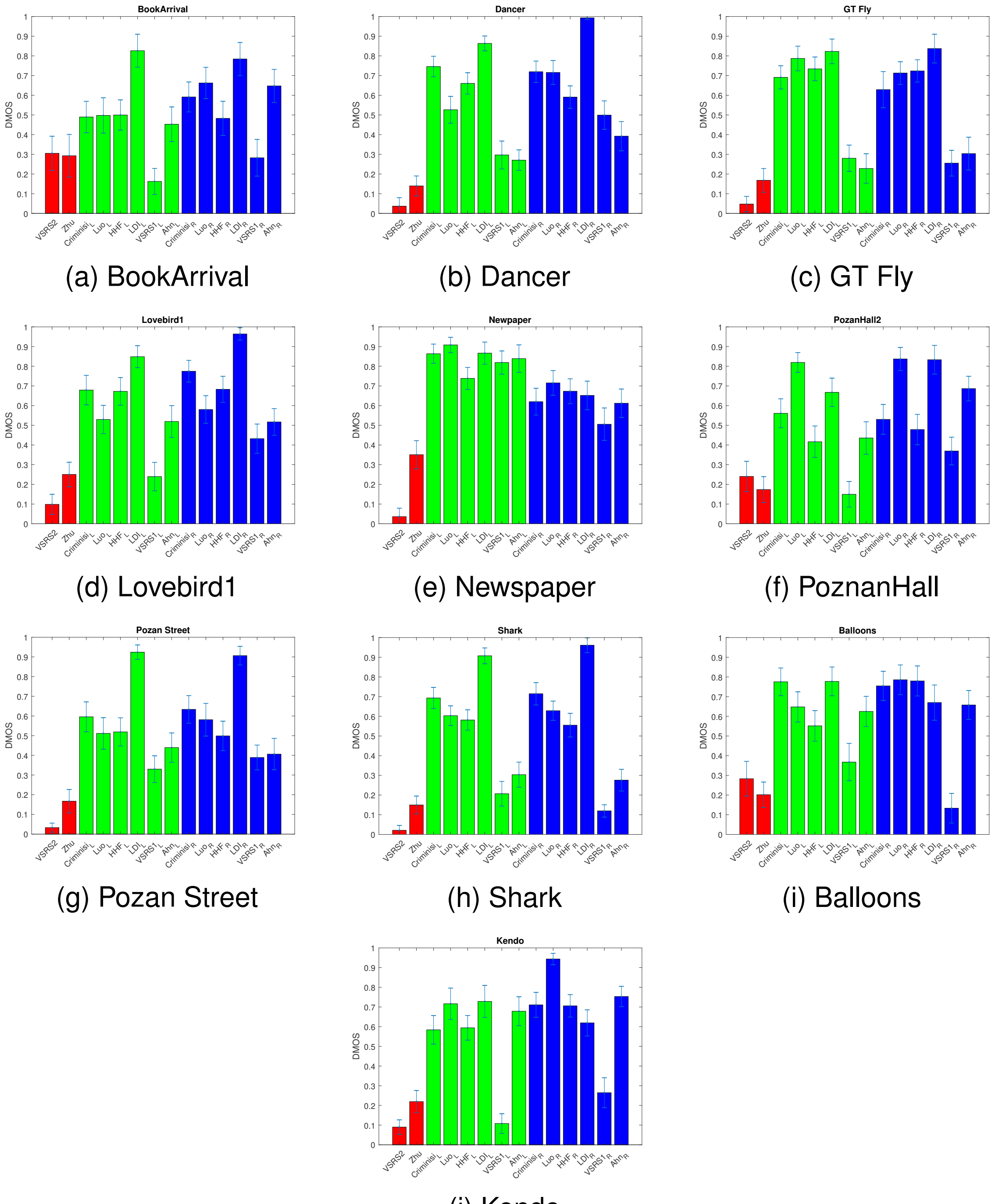
	VSRS2	Zhu	Cri.	Luo	HHF	LDI	VSRS1	Ahn
VSRS2	0	1	1	1	1	1	1	1
Zhu	—	0	1	1	1	1	1	1
Cri.	—	—	0	0	-1	1	-1	-1
Luo	—	—	—	0	-1	1	-1	-1
HHF	—	—	—	—	0	1	-1	-1
LDI	—	—	—	—	—	0	-1	-1
VSRS1	—	—	—	—	—	—	0	1
Ahn	—	—	—	—	—	—	—	0

4.5 Objective Measurement

In this section, we compare the performances of several existing objective image quality assessment metrics on the proposed database.

4.5.1 Objective Metrics

In addition to the DIBR-synthesized image dedicated quality metrics which has already been introduced in Chapter 1, to be robust, we also test the following state-of-the-art 2D image quality metrics:



(a) BookArrival

(b) Dancer

(c) GT Fly

(d) Lovebird1

(e) Newspaper

(f) PoznanHall

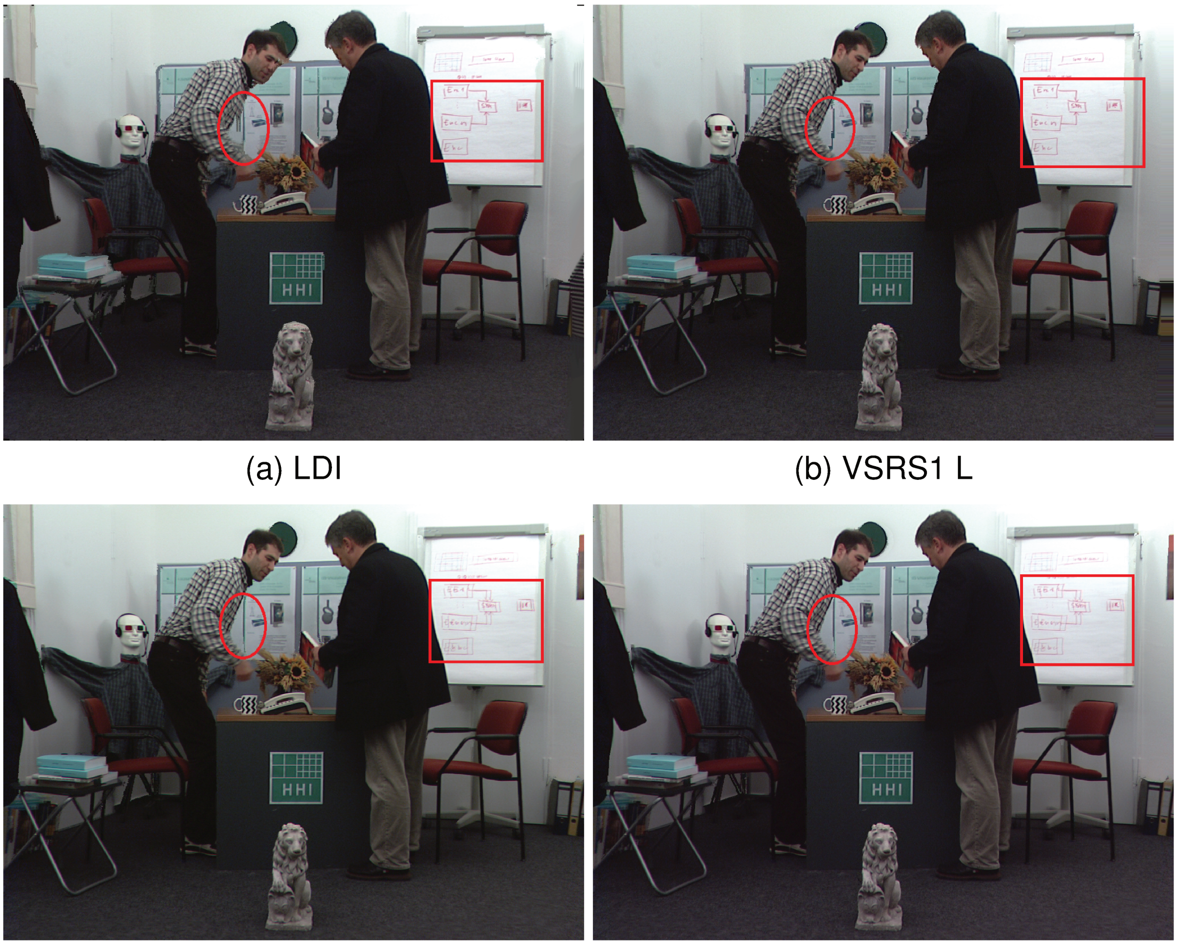
(g) Pozan Street

(h) Shark

(i) Balloons

(j) Kendo

Figure 4.9: DMOS distribution and confidence intervals of the synthesized views of different MVD sequences and different view synthesis methods. The x-labels are $VSRS2$, Zhu, $Criminisi_L$, Luo_L, HHF_L, LDI_L, $VSRS1_L$, Ahn_L, $Criminisi_R$, Luo_R, HHF_R, LDI_R, $VSRS1_R$, Ahn_R ordinally. The subscript L means this virtual view is synthesized from the neighboring left view, while the subscript R means from the right. $VSRS2$ denotes the view interpolation inter-view mode of VSRS. The error bars indicates the corresponding confidence intervals of the tested images. The bars referring to interview synthesize views are marked by red, the left-extrapolated views marked by green, and the right-extrapolated views marked by blue.



(a) LDI

(b) VSRS1 L

(c) Zhu’s method

(d) VSRS2

Figure 4.10: Examples of synthesized images

The Full-Reference (FR) 2D metrics include:

- SSIM: Structure SIMilarity, a widely used objective FR IQA metric calculating the structure similarity between the tested and the reference images proposed by Wang et al. in [112].
- MS-SSIM: Multi-Scale Structure SIMilarity, a multi-scale approach of SSIM proposed by Wang et al. in [111].
- PSNR: Peak Signal to Noise Ratio, a widely used pixel-based metric.
- IW-PSNR, IW-SSIM: Information content Weighted FR IQA Metric based on PSNR and SSIM separately, proposed by Wang et al. in [110].
- UQI: Universal Quality Index proposed by Wang et al. in [107], models the image distortions by integrating loss correlation, luminance distortion and contrast distortion.
- PSNR-HVS: based on PSNR and UQI, takes the Human Vision System (HVS) into account [19] [66].

The No-Reference (NR) 2D metrics include:

- BIQI: Blind Image Quality Index, a NR IQA metric proposed by Moorthy et al. in [59].
- BliindSII: BLind Image Interfrity Notaor using DCT Statistics-II proposed by Saad et al. in [71].
- NIQE: Natural Image Quality Evaluator, a NR IQA metric proposed by Mittal et al. in [58].

4.5.2 Correlation between the objective and subjective measurements

As introduced in the previous chapters, the performance of objective quality assessment metrics can be evaluated by their correlations with the subjective test results. These correlations methods compare the performance of each metric by calculating their correlations with the subjective results, however they just take the mean value of subjective scores into consideration, the uncertainty of the subjective scores has been ignored. In addition, the quality scores need to be regressed by a regression function cf. Eq. 1, that is not the way they are exactly used in real scenarios. Thus, we further conduct a statistical test proposed by Krasula et al. in [43] which does not suffer from the drawbacks of the above methods. The performances of objective metrics are evaluated by their classification abilities.

PLCC, RMSE, SROCC performance comparison

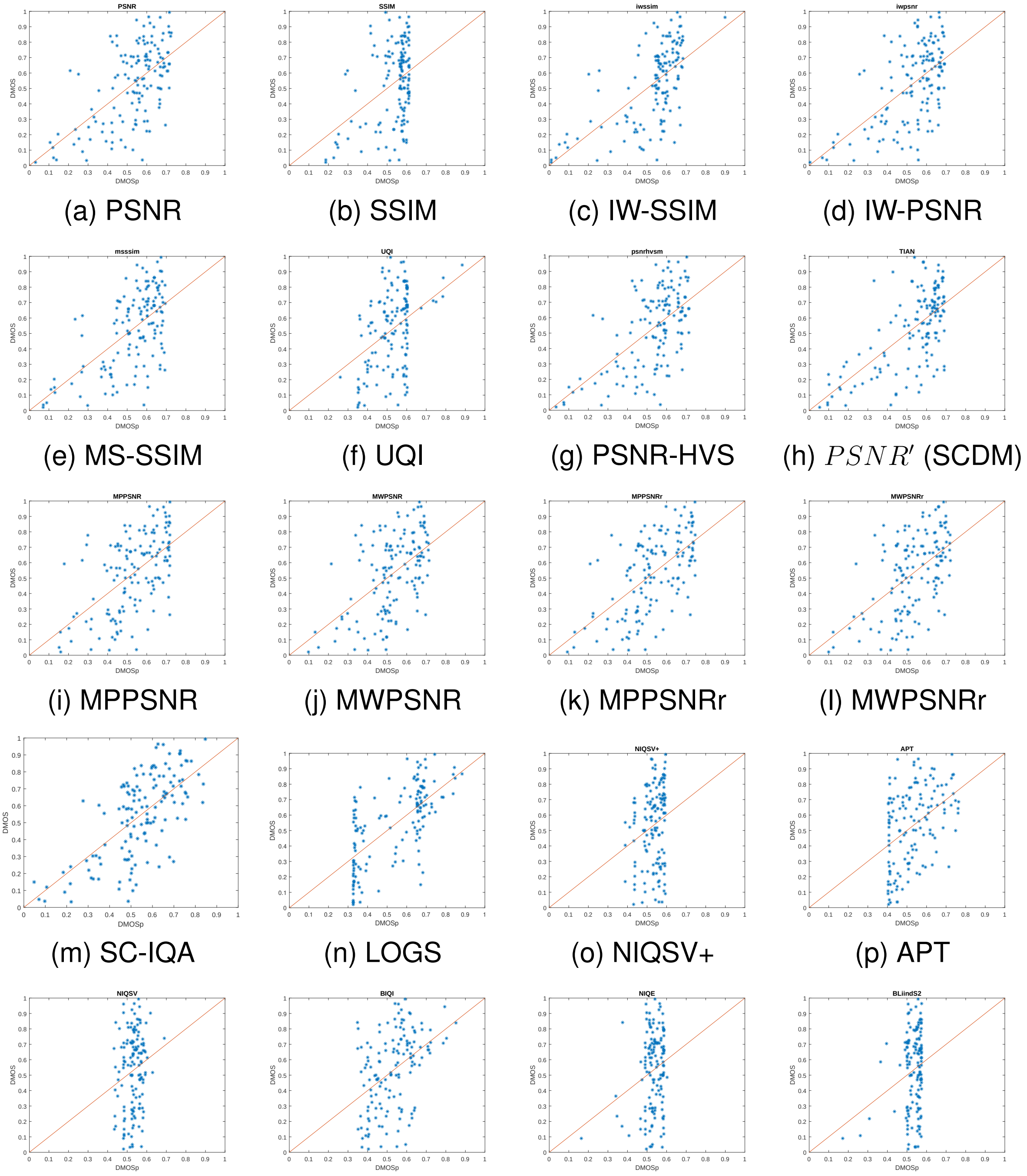
The obtained PLCC, RMSE, SROCC values are given in Table 4.6. It can be noticed at once that the performances of these metrics on the presented database are quite bad (no PLCC value more than 70%). Among which, the proposed metrics $PSNR'$ (SCDM), SC-IQA and the side view based FR metric LOGS perform the best in terms of the PLCC on this database. Especially for NIQSV+, NIQSV, NIQE and BliindS2 NR metrics, they show weak correlations with the subjective results.

In Table 4.6, we use the parameters provided by the authors, which make the algorithms achieve their best performance on the IVC DIBR database. In Table 4.7, we investigate the performance dependency of MWPSNR and MPPSNR on decomposition level and structural element size. It shows that these parameters can be fitted to achieve better performance on the proposed database, but they still cannot get satisfactory results. In addition, we think that a high degree of generality is a desirable feature for a good quality metric. That means a metrics performance cannot be judged only on its best performance on a selected database. This is also why the cross-validation is usually needed for the validation of a metric.

The scatter plot of each IQA metric is shown in Fig. 4.11. It seems that all methods are incapable of predicting worse qualities (bigger DMOS value indicates worse quality), which is however consistent with the results shown in Table 4.6 where no metric has a PLCC value higher than 0.7. While some of them do sometimes succeed in their prediction of high qualities in Fig. 4.11 (consistent with their PLCC values bigger than 0.5). Be similar to the results in Table 4.6, the NR metrics NIQSV+, NIQSV, NIQE and BliindS2 show little correction with the subjective results, there is large empty regions in the corresponding scatter plots (consistent with their PLCC values smaller than 0.3).

Table 4.6: PLCC, RMSE and SROCC between DMOS and objective metrics, where "SV FR metric" indicates the side view based FR metric

Metric		PLCC	RMSE	SROCC
FR 2D metrics	PSNR	0.6012	0.1985	0.5356
	SSIM	0.4016	0.2275	0.2395
	MS-SSIM	0.6162	0.1957	0.5355
	IW-PSNR	0.5827	0.2019	0.4973
	IW-SSIM	0.6280	0.1933	0.5950
	UQI	0.4346	0.2237	0.4113
	PSNR-HVS	0.5982	0.1991	0.5195
FR 3D metrics	MP-PSNR	0.5753	0.2032	0.5507
	MP-PSNRr	0.6061	0.1976	0.5873
	MW-PSNR	0.5301	0.2106	0.4845
	MW-PSNRr	0.5403	0.2090	0.4946
	VSQA	0.5576	0.2062	0.4719
	PSNR' (pro)	0.6685	0.1844	0.5903
	SC-IQA	0.6856	0.1805	0.6423
SV FR metric	LOGS	0.6687	0.1845	0.6683
NR 3D metrics	NIQSV	0.1759	0.2446	0.1473
	NIQSV+	0.2095	0.2429	0.2190
	APT	0.4225	0.2252	0.4187
NR 2D metrics	NIQE	0.2244	0.2421	0.1360
	BLiindS2	0.2225	0.2422	0.1329
	BIQI	0.4348	0.2237	0.4328



(a) PSNR (b) SSIM (c) IW-SSIM (d) IW-PSNR

(e) MS-SSIM (f) UQI (g) PSNR-HVS (h) $PSNR'$ (SCDM)

(i) MPPSNR (j) MWPSNR (k) MPPSNRr (l) MWPSNRr

(m) SC-IQA (n) LOGS (o) NIQSV+ (p) APT

(q) NIQSV (r) BIQI (s) NIQE (t) BliindS2

Figure 4.11: Scatter plots of DMOS versus DMOSp of each IQA method

Table 4.7: Performance dependency of MP-PSNR and MWPSNR, where “SE” indicates Structural Element size

		MW-PSNR			MW-PSNRr		
level	SE	PLCC	RMSE	SROCC	PLCC	RMSE	SROCC
8		0.5500	0.2070	0.5199	0.5602	0.2054	0.5235
7		0.5389	0.2088	0.4875	0.5383	0.2089	0.4953
6		0.6132	0.1958	0.5598	0.6095	0.1965	0.5634
5		0.5981	0.1987	0.5353	0.6014	0.1981	0.5240
level	SE	MP-PSNR			MP-PSNRr		
6	7	0.6037	0.1976	0.5578	0.5659	0.2044	0.5402
6	5	0.6284	0.1929	0.5737	0.5889	0.2004	0.5794
6	3	0.6312	0.1923	0.5914	0.6023	0.1979	0.5745
6	2	0.6134	0.1958	0.5601	0.5945	0.1993	0.5443
5	7	0.6190	0.1947	0.5809	0.5841	0,2012	0.5570
5	5	0.6294	0.1927	0.5951	0.6160	0.1953	0.5870
5	3	0.6246	0.1936	0.5860	0.6314	0.1922	0.5855
5	2	0.6163	0.1952	0.5497	0.6009	0.1982	0.5313
4	7	0.6170	0.1951	0.5909	0,6023	0.1979	0.5569
4	5	0.6230	0.1939	0.5796	0,6247	0.1936	0.5678
4	3	0.6170	0.1951	0.5619	0.6106	0.1963	0.5470
4	2	0.5911	0.2000	0.5319	0.5571	0.2059	0.4928

Analysis of Krasula's model

The above methods compare the performance of each metric by calculating their correlations with the subjective results, however they just take the mean value of sub- jective scores into consideration, the uncertainty of the subjective scores has been ignored. In addition, the quality scores need to be regressed by a regression function cf. Eq. 1, that is not the way they are exactly used in real scenarios. Thus, we further conduct a statistical test proposed by Krasula et al. in [Kra+16] which does not suffer from the drawbacks of the above methods. The performances of objective metrics are evaluated by their classification abilities.

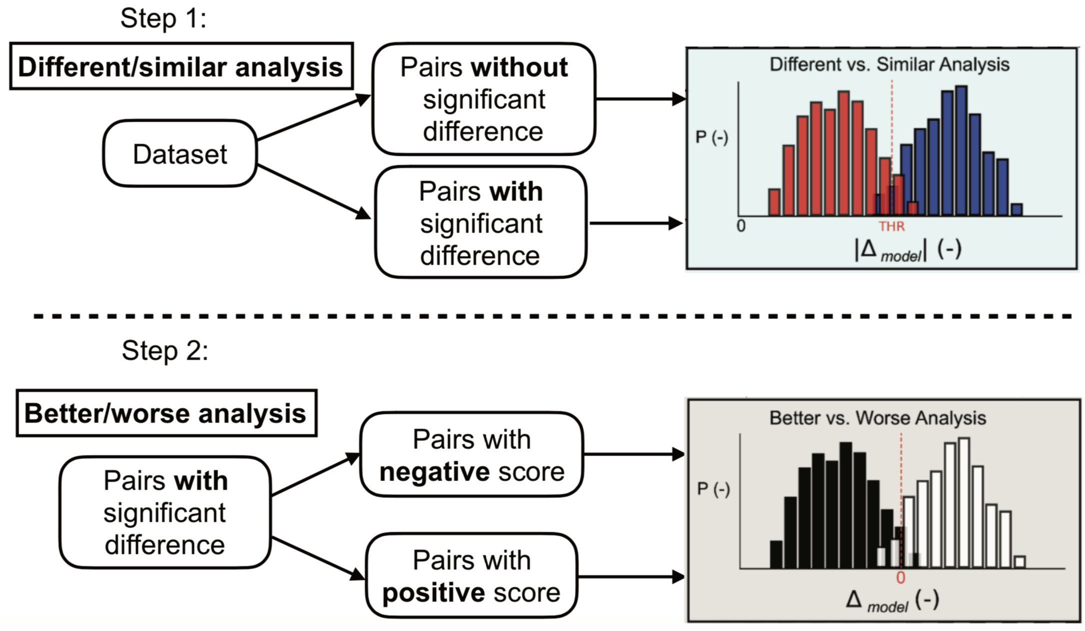

Step 1:
Different/similar analysis
Dataset
Pairs **without** significant difference
Pairs **with** significant difference
Different vs. Similar Analysis
P (-)
0
THR
|Δ $_{model}$| (-)
Step 2:
Better/worse analysis
Pairs **with** significant difference
Pairs with **negative** score
Pairs with **positive** score
Better vs. Worse Analysis
P (-)
0
Δ $_{model}$ (-)

Figure 4.12: Krasula's model for performance evaluation of objective quality metrics.

As shown in Fig. 4.12, firstly, the tested image pairs in the database are divided into two groups: different and similar according to their subjective scores. The cumulative distribution function (cdf) of the normal distribution is used to calculate the probability of image pairs. Then, we consider the pairs with higher than the selected significance level 0.95 to be significantly different. The others will be recognized as similar.

There are two performance analysis, the first performance analysis is conducted by how well the objective metric can distinguish the video pair are significant different or not. If the two videos in the pair are significantly different according to the subjective results. In the second analysis, we determine whether the objective metric can correctly identify the image of higher quality in the pair.

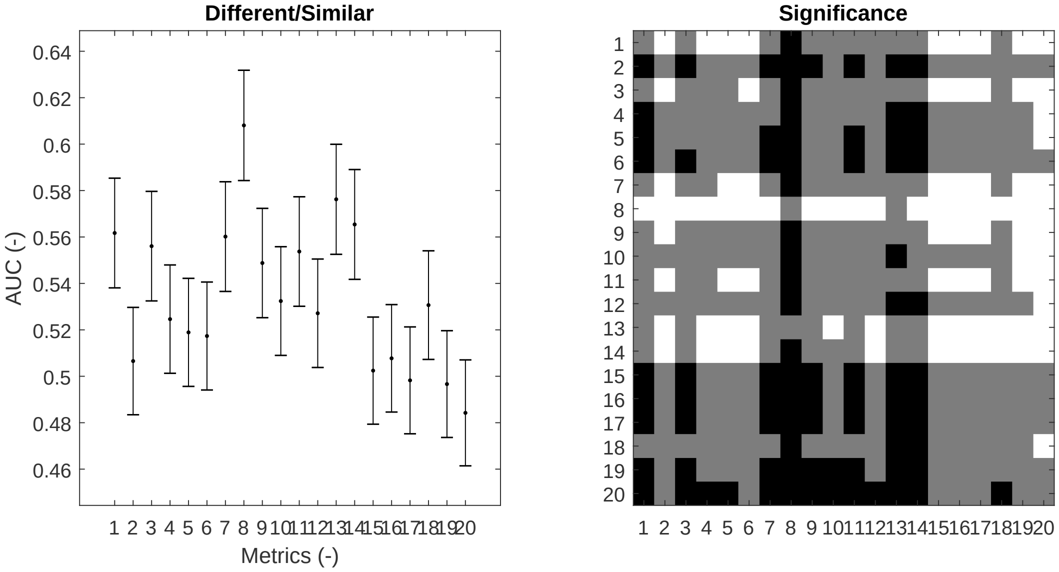
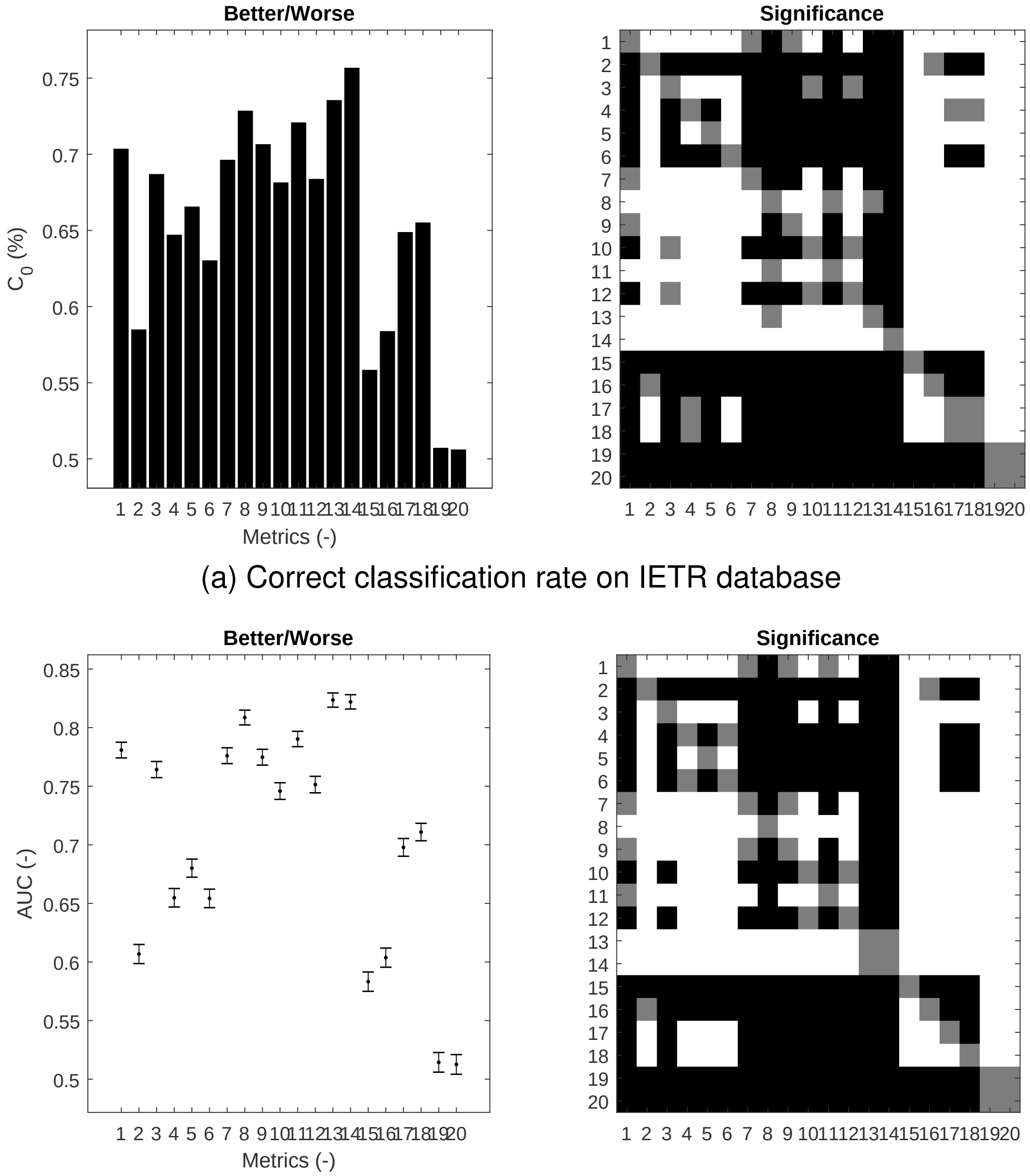


Figure 4.13: Results of different/similar on IETR database (Metric 1-20 represent PSNR, SSIM, IW-PSNR, IW-SSIM, MS-SSIM, UQI, PSNRHVSM, $PSNR'$ (SCDM), MP-PSNR, MW-PSNR, MP-PSNRr, MW-PSNRr, SC-IQA, LOGS, NIQSV, NIQSV+, APT, BIQI, NIQE, BliindS2)

Analysis on the proposed IETR database The obtained results of the tested metrics on the proposed IETR database, the Correct Classification percentage (C_0) and the Area Under the Curves (AUC) are given in Fig. 4.13 and Fig. 4.14 respectively.

In the first different/similar analysis, the AUC values of most metrics are within [0.5, 0.6], of which the metric $PSNR'$ (SCDM) (metric 8 in the figure, the FR quality model introduced in Chapter 3) performs the best. There even exit some metrics whose AUC values are under 0.5.

In the second better/worse analysis, the metric $PSNR'$ (SCDM), MW-PSNR and SC-IQA performs significantly better than the other metrics. Similar to the the first different/similar analysis, the last two metrics (NIQE and BliindS2) have little correlation with the subjective results, and there is no metric whose AUC value is higher than 0.85, which is consistent with results in Table 4.6 and Fig. 4.11 that none of the metrics can achieve a satisfactory correlation with the ground truth. Especially, the better/worse analysis results are quite consistent with the SROCC values, which give the mono-



(a) Correct classification rate on IETR database

(b) Area under the ROC curve (AUC on IETR database)

Figure 4.14: Results of better/worse on IETR database (Metric 1-20 represent PSNR, SSIM, IW-PSNR, IW-SSIM, MS-SSIM, UQI, PSNRHVSM, $PSNR'$ (SCDM), MP-PSNR, MW-PSNR, MP-PSNRr, MW-PSNRr, SC-IQA, LOGS, NIQSV, NIQSV+, APT, BIQI, NIQE, BliindS2)

tonicity estimation, in Table 4.6. The metrics which lower SROCC values performs worse in the better/worse analysis according to Fig. 4.14 (a) and (b).

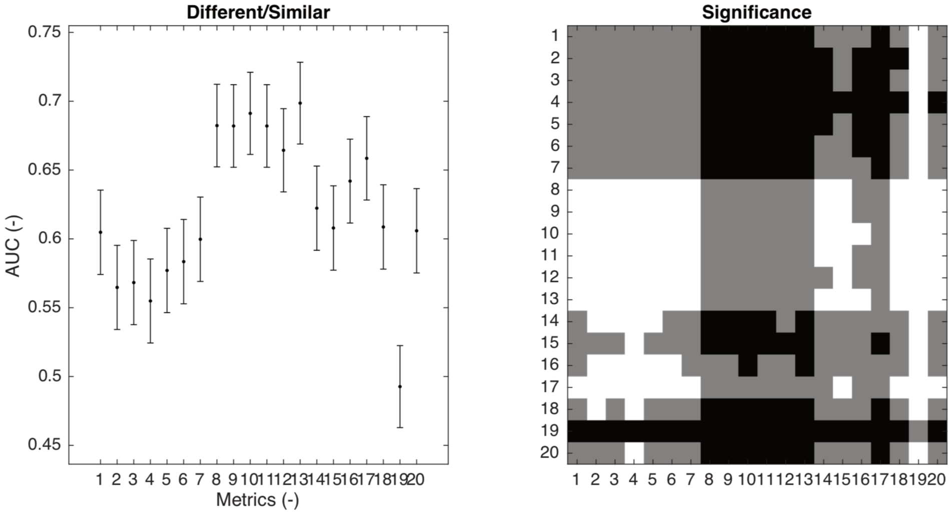
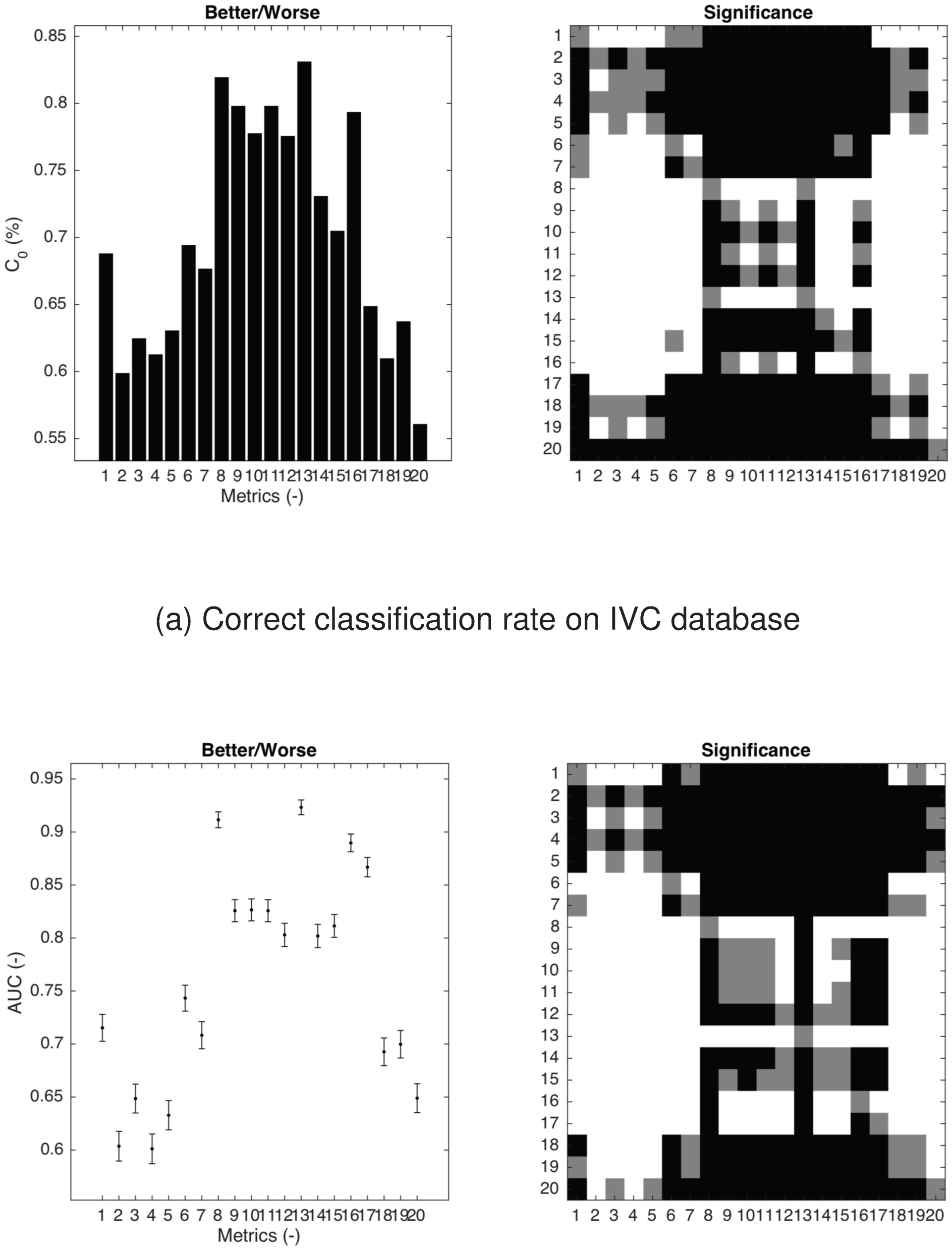


Figure 4.15: Results of different/similar on IVC database (Metric 1-20 represent PSNR, SSIM, IW-PSNR, IW-SSIM, MS-SSIM, UQI, PSNRHVSM, $PSNR'$ (SCDM), MP-PSNR, MW-PSNR, MP-PSNRr, MW-PSNRr, SC-IQA, LOGS, NIQSV, NIQSV+, APT, BIQI, NIQE, BliindS2)

Analysis on the IVC DIBR-image database The obtained results on IVC DIBR-image database are shown in Fig. 4.15 and Fig. 4.16.

In the first difference/similar analysis, cf. Fig. 4.15, the DIBR-synthesized view dedicated FR metrics (metrics 8-13) performs significantly better than the traditional 2D FR quality metrics (metrics 1-7 in the figure) and the SVFR, NR metrics. The AUC values of 2D FR metrics are between 0.5 and 0.6, even the AUC values of DIBR FR metrics can only achieve 0.7. Among which our proposed metric SC-IQA (metric 13) performs the best.

In the second better/worse analysis, cf. Fig. 4.16, the similar conclusions can be drawn that the DIBR FR metrics outperform the other metrics generally except our proposed NR metric NIQSV+ (metric 16) since most of the metrics can achieve a higher



(a) Correct classification rate on IVC database

(b) Area under the ROC curve (AUC) on IVC database

Figure 4.16: Results of better/worse on IVC database (Metric 1-20 represent PSNR, SSIM, IW-PSNR, IW-SSIM, MS-SSIM, UQI, PSNRHVSM, $PSNR'$ (SCDM), MP-PSNR, MW-PSNR, MP-PSNRr, MW-PSNRr, SC-IQA, LOGS, NIQSV, NIQSV+, APT, BIQI, NIQE, BliindS2)

correct classification rate and AUC value. Among these the tested metrics, the proposed $PSNR'$ (SCDM), SC-IQA and NIQSV+ perform the best along with the APT metric. They achieved an AUC value higher than 0.85, especially the AUC values of $PSNR'$ (SCDM), SC-IQA are even higher than 0.9. Similar to the results on IETR dataset, the 2D FR and NR metric perform not well on the DIBR databases. The results are consistent with the correlation results in Tab. 2.1 and 3.2.

Based on the above analysis, it could be referred that:

1. The 2D metrics (including FR and NR) show weak correlation with subjective results on DIBR-synthesized view databases, the main reason could be that the 2D metrics are trained and focus on the traditional artifacts, such as blurry, additive white noise, jpeg etc. they cannot well assess the quality of DIBR synthesized views.

2. The test DIBR-synthesized view dedicated quality metrics (including FR and NR) perform much better on IVC database than on the proposed IETR database. Among which, the performance of NR metrics decrease the most, which is consistent with the results in Tab. 4.6. One reason of this cross datasets performance decrease could be that the DIBR NR metrics NIQSV, NIQSV+ and APT tried to optimize their performances on the IRCCyN/IVC DIBR database where “old-fashioned” artifacts exist. On the new proposed IETR database, they cannot get a good performance when the “old-fashioned” are excluded.

4.6 Conclusion

DIBR is widely used in FVV, VR, AR, and other popular topics considered as the next generation of 3D broadcasting applications, in order to provide a better QoE to users. In this chapter, a new DIBR-synthesized image database which focuses on the distortions induced by different state-of-the-art DIBR view synthesis algorithms is presented. Ten MVD sequences and seven state-of-the-art DIBR view synthesis algorithms are selected to generate the virtual view images. The subjective experiment is conducted following the SAMVIQ protocol in a controlled environment as recommended by ITU-R Rec. BT.1788 [36]. Results show that the inter-view synthesis methods, which have more input information, significantly outperform the single view based synthesis algorithms. Furthermore, several objective measurements were used to assess the quality of synthesized images on this database. Their performance results indicate that further work has to be done to exploit deeply the characteristics of these specific distortions, for new objective metrics with a better correlation with subjective scores. However, in the current database, only the MPEG MVD source images are included; in the future work, more source images, such as the images from Middlebury database [81], will be considered to make the experiment results more benchmark.

All the data of this presented database, including images, the ground truth depth maps and their associated DMOS, is publicly accessible (`https://vaader-data.insa-rennes.fr/data/stian/ieeetom/IETR_DIBR_Database.zip`), for the improvement of the QoE of DIBR related applications.

CHAPTER 5

CONCLUSION AND PERSPECTIVES

5.1 Brief summary

As the needs of the visual experience increase, applications that can provide more immersive perception of 3D visual scene, have gained an exceptional increase in public interest and curiosity in the past decade. As a fundamental technology to synthesis virtual view, DIBR is widely used in 3D applications. It can help to reduce the transmission cost. However, this DIBR process produces new types of distortion, which are far different from those distortions induced by 2D video compression. Therefore, the quality assessment of DIBR-synthesized views is of great importance for a high quality immersive experience.

This thesis is dedicated to assess the quality of 3D synthesized views objectively and subjectively. After a careful analysis of the existing state-of-the-art quality metrics for DIBR-synthesized views, we propose two no-reference and two full-reference image quality assessment metrics for DIBR-synthesized views respectively. Then, we present a novel DIBR image database.

5.2 Contributions and list of publications

The main contributions and novelties of this thesis are:

1. A totally NR quality assessment metric for DIBR-synthesized views (NIQSV) (detailed in Chapter 2). It uses a set of morphological operations to detect the typical distortions in DIBR-synthesized views: “thin distortion”, “blurry region” and “crumbling”. The experimental results show that the proposed NIQSV metric outperforms traditional 2D metrics and ranks among the best of dedicated 3D synthesized and full reference metrics. Moreover, as the morphological operators only

contain integer operations, our metric holds a very low computational complexity. Published in IEEE ICASSP conference [100].

2. An extended NR quality metric (NIQSV+) of the NIQSV (detailed in Chapter 2). NIQSV+ improves the original NIQSV by considering more distortion types: “black hole” and “stretching”. The experimental resuls show that it achieves a gain of 7.68% in correlation with subjective measurement (PLCC) compared to the original NIQSV, and even outperforms the widely used FR and NR 2D IQA metrics. Published in IEEE Transactions on Image Processing [101].

3. A FR quality model for DIBR-synthesized views (SC-DM) (detailed in chapter 3). This model can be used on any pixel based quality assessment metric. The performance improvements of PSNR and SSIM are 36.85% and 13.33% respectively in terms of PLCC. Besides, the improved PSNR outperforms all the tested DIBR quality metrics including NR and FR significantly. Published in IS&T EI conference [99].

4. A FR Shift Compensation based Image Quality Assessment (SC-IQA) metric (detailed in chapter 3). Focusing on the geometric distortion in the DIBR-synthesized views, we use a two step shift compensation method to compensate the global shift and penalize the local geometric distortion at the same time. The experiment results show that the proposed SC-IQA outperforms all the tested DIBR quality metrics including NR and FR significantly. Published in IEEE VCIP conference [103].

5. A new DIBR-synthesized image database and a benchmark of existing DIBR-synthesized view dedicated quality metrics (detailed in Chapter 4). Compared with the existing DIBR databases, firstly, we tested more and newer DIBR algorithms including interpolation and extrapolation; secondly, all the tested images are showed on a 2D display to avoid the 3D display settings and configurations influences, thirdly, some “old fashioned” DIBR artifacts (eg. “black hole” in IVC database) have been excluded in the proposed database. In addition, we conducted a relatively complete bench-marking of the state-of-the-art objective metrics for DIBR-synthesized view quality assessment. Published in IEEE transactions on Multimedia [98].

6. A performance comparison of existing quality metrics on free-viewpoint videos

with different depth coding algorithms. The results show that the DIBR dedicated metric performs better than the conventional 2D quality metrics generally, but none of these metrics can achieve a satisfactory correlation with the ground truth. There is certainly room for the improvement of quality assessment method of free-viewpoint videos with compressed depth maps. Published in SPIE SPIE Optical Engineering + Applications conference [102].

So far, our research work resulted in 6 Publications (2 journal papers and 4 conference papers), which are given below:

Journal papers:

[1] Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. “NIQSV+: A No Reference Synthesized View Quality Assessment Metric”. IEEE Transactions on Image Processing; April 2018; 27(4), Page(s): 1652 – 1664.

[2] Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. “A benchmark of DIBR Synthesized View Quality Assessment Metrics on a new database for Immersive Media Applications”. IEEE Transactions on Multimedia, 2018.

Conference papers:

[3] Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. “NIQSV: A No Reference Image Quality Assessment Metric for 3D Synthesized Views”. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), March 2017, New Orleans, USA.

[4] Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. “A full-reference Image Quality Assessment metric for 3D Synthesized Views”. Image Quality and System Performance Conference, at IS&T Electronic Imaging 2018, 28 January – 1 February 2018, Burlingame, California, USA.

[5] Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. “Performance comparison of objective metrics on free-viewpoint videos with different depth coding algorithms”. SPIE Optical Engineering + Applications, August 2018, San Diego, California, USA.

[6] Shishun Tian, Lu Zhang, Luce Morin, Olivier Déforges. “SC-IQA: Shift compensation based image quality assessment for DIBR-synthesized views”. IEEE international conference on Visual Communications and Image Processing (VCIP), December 2018, Taichung, Taiwan.

5.3 Perspectives

Generally, we focus on the image quality assessment of DIBR-synthesized views objectively and subjectively in this thesis. However, there is still plenty space for the improvement in this research area.

1. Only limited type of distortions are considered in this work, such as blurry regions, crumbling and stretching. It would be interesting to investigate other types of distortions to improve the performance of existing metrics. Besides, for different applications, different factors should also be taken into account. For instance, the binocular disparity distortion for stereoscopic applications, etc.

2. For the video applications, due to the geometric distortions in DIBR-synthesized views, obvious “flickering” distortion usually happens in the DIBR-synthesized videos. In future work, we’ll try to extend our IQA metric to VQA metric by handling the “flickering” distortion.

3. The deep learning approaches, in particular the convolutional neural networks, have shown their great advantages in different computer vision research topics. It becomes possible to learn the representative features directly from image or video data. In recent years, several efforts have been made to assess the quality of traditional 2D images [44, 55], but fewer work has been done on the quality assessment of DIBR-synthesized views. The main reason could be the limitation of database. Unlike the homogeneous distortions in the traditional 2D images, the distortions in the DIBR-synthesized views mostly occur in the dis-occlusion regions. In other words, the majority part of the DIBR-synthesized view hold a perfect quality. We could not split the image into several patches and directly use the quality of the whole image as the quality of all the patches. Recently, Generative Adversarial Nets (GAN) [26] has been used to generate various type of images [18, 35, 67, 32]. During its training step, as the generative network is

trained better and better, the discriminative network is also trained to better discriminate the generated image with various distortions. At the same time, many efforts have been made to use the deep neural network to learn the multi-view representation and generate novel virtual views [118, 104, 116]. So, one possible way could be that using the generative network to generate virtual views with various distortions which may help solve the problem of size limit of existing database, and then use the discriminative network to evaluate the virtual view's quality.

GLOSSARY

Notation	Description
AR	Augmented Reality. 14
CG	Computer Graphics. 25
DIBR	Depth-Image-Based-Rendering. 14, 16, 17, 18, 19, 121, 141
FR	Full-reference. 5, 17, 18, 31, 33, 44, 53
FVV	Free-viewpoint Video. 13, 18, 24
HEVC	High Efficiency Video Coding. 14
HVS	Human Vision System. 18, 34
IQA	Image quality assessment. 31
LF	Light Field. 14
MVD	Multiview Video plus Depth. 14
MVV	Multi-view Video. 13, 24
NR	No-reference. 6, 17, 18, 31, 47, 48
PSNR	Peak signal-to-noise ratio. 18
QoE	Quality of Experience. 13, 14
RANSAC	Random sample consensus. 18, 53
RR	Reduced-reference. 18, 31, 44

Notation	Description
SC-DM	Shift Compensation and Dis-occlusion based Model. 53
SE	Structural element. 67, 68
SSIM	Structural similarity index. 18
SURF	Speeded Up Robust Features. 18, 53
VR	Virtual Reality. 13, 14

BIBLIOGRAPHY

[1] Ilkoo Ahn and Changick Kim, “A novel depth-based virtual view synthesis method for free viewpoint video”, *in*: *IEEE Transactions on Broadcasting* 59.*4* (2013), pp. 614–626.

[2] Ronald T Azuma, “A survey of augmented reality”, *in*: *Presence: Teleoperators and virtual environments* 6.*4* (1997), pp. 355–385.

[3] Amitav Banerjee et al., “Hypothesis testing, type I and type II errors”, *in*: *Industrial psychiatry journal* 18.*2* (2009), p. 127.

[4] Federica Battisti et al., “Objective image quality assessment of 3D synthesized views”, *in*: *Signal Processing: Image Communication* 30 (2015), pp. 78–88.

[5] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, “Surf: Speeded up robust features”, *in*: *European conference on computer vision*, Springer, 2006, pp. 404–417.

[6] Alexandre Benoit et al., “Quality assessment of stereoscopic images”, *in*: *EURASIP journal on image and video processing* 2008.*1* (2009), p. 659024.

[7] Marcelo Bertalmio, Andrea L Bertozzi, and Guillermo Sapiro, “Navier-stokes, fluid dynamics, and image and video inpainting”, *in*: *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, IEEE, 2001, pp. I–I.

[8] Jean-Louis Blin, “New quality evaluation method suited to multimedia context: Samviq”, *in*: *Proceedings of the Second International Workshop on Video Processing and Quality Metrics, VPQM*, vol. 6, 2006.

[9] E. Bosc et al., “An edge-based structural distortion indicator for the quality assessment of 3D synthesized views”, *in*: *2012 Picture Coding Symposium*, 2012, pp. 249–252.

[10] Emilie Bosc, “Compression of Multi-View-plus-Depth (MVD) data: from perceived quality analysis to MVD coding tools designing”, PhD thesis, INSA de Rennes, 2012.

[11] Emilie Bosc et al., “A quality assessment protocol for Free-viewpoint video sequences synthesized from decompressed depth data”, *in*: *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, IEEE, 2013, pp. 100–105.

[12] Emilie Bosc et al., “Towards a new quality metric for 3-D synthesized view assessment”, *in*: *IEEE Journal of Selected Topics in Signal Processing* 5.*7* (2011), pp. 1332–1343.

[13] Emilie Bosc et al., “Visual quality assessment of synthesized views in the context of 3D-TV”, *in*: *3D-TV system with depth-image-based rendering*, Springer, 2013, pp. 439–473.

[14] Ying Chen et al., “The emerging MVC standard for 3D video services”, *in*: *EURASIP Journal on Applied Signal Processing* 2009 (2009), p. 8.

[15] Chi Ho Cheung, King Ngi Ngan, and Lu Sheng, “Spatio-Temporal Disocclusion Filling Using Novel Sprite Cells”, *in*: *IEEE Transactions on Multimedia* 20.*6* (2018), pp. 1376–1391.

[16] Pierre-Henri Conze, Philippe Robert, and Luce Morin, “Objective view synthesis quality assessment”, *in*: *IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, 2012, pp. 82881M–82881M.

[17] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama, “Region filling and object removal by exemplar-based image inpainting”, *in*: *IEEE Transactions on image processing* 13.*9* (2004), pp. 1200–1212.

[18] Chao Dong et al., “Learning a deep convolutional network for image super-resolution”, *in*: *European conference on computer vision*, Springer, 2014, pp. 184–199.

[19] Karen Egiazarian et al., “New full-reference quality metrics based on HVS”, *in*: *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, vol. 4, 2006.

[20] F. Battisti, *3DSwIM Source Code*, `http://www.comlab.uniroma3.it/3DSwIM.html`, last accessed Aug. 30th 2017, [Online].

[21] Muhammad Shahid Farid, Maurizio Lucenteforte, and Marco Grangetto, “Perceptual quality assessment of 3D synthesized images”, *in*: *Multimedia and Expo (ICME), 2017 IEEE International Conference on*, IEEE, 2017, pp. 505–510.

[22] Christoph Fehn, “A 3D-TV approach using Depth-image-based rendering (DIBR)”, *in*: *Proc. of VIIP*, vol. 3, 3, 2003.

[23] Christoph Fehn, “Depth-Image-Based Rendering (DIBR), compression, and transmission for a new approach on 3D-TV”, *in*: *Electronic Imaging 2004*, International Society for Optics and Photonics, 2004, pp. 93–104.

[24] Xiaojun Feng and Jan P Allebach, “Measurement of ringing artifacts in JPEG images”, *in*: *Proceedings of SPIE*, vol. 6076, 2006, pp. 74–83.

[25] Martin A Fischler and Robert C Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography”, *in*: *Readings in computer vision*, Elsevier, 1987, pp. 726–740.

[26] Ian Goodfellow et al., “Generative adversarial nets”, *in*: *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[27] Video Quality Experts Group, “FINAL REPORT FROM THE VIDEO QUALITY EXPERTS GROUP ON THE VALIDATION OF OBJECTIVE MODELS OF MULTIMEDIA QUALITY ASSESSMENT”, *in*: *VQEG* (March 2008).

[28] Ke Gu et al., “Deep learning network for blind image quality assessment”, *in*: *Image Processing (ICIP), 2014 IEEE International Conference on*, IEEE, 2014, pp. 511–515.

[29] Ke Gu et al., “Model-based referenceless quality metric of 3D synthesized images using local image description”, *in*: *IEEE Transactions on Image Processing* (2017).

[30] Philippe Hanhart and Touradj Ebrahimi, “Quality assessment of a stereo pair formed from decoded and synthesized views using objective metrics”, *in*: *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2012*, IEEE, 2012, pp. 1–4.

[31] Ian P Howard, Brian J Rogers, et al., *Binocular vision and stereopsis*, Oxford University Press, USA, 1995.

[32] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa, “Globally and locally consistent image completion”, *in*: *ACM Transactions on Graphics (TOG)* 36.*4* (2017), p. 107.

[33] Wijnand Ijsselsteijn, Pieter J.H. Seuntiëns, and Lydia M.J. Meesters, *Human Factors of 3D Displays*, Jan. 2006, pp. 217 –233, ISBN: 9780470022733.

[34] *Industrial demonstration*, `https://igrv2017.sciencesconf.org/resource/page/id/12`.

[35] Phillip Isola et al., “Image-to-image translation with conditional adversarial networks”, *in*: *arXiv preprint* (2017).

[36] ITURBT ITU, “Methodology for the subjective assessment of video quality in multimedia applications”, *in*: *Rapport technique, International Telecommunication Union* (2007).

[37] IVC-IRCCyN lab, *IRCCyN/IVC DIBR image database*, `http://ivc.univ-nantes.fr/en/databases/DIBR_Images/`, last accessed Aug. 30th 2017, [Online].

[38] Vincent Jantet, Christine Guillemot, and Luce Morin, “Object-based Layered Depth Images for improved virtual view synthesis in rate-constrained context”, *in*: *Image Processing (ICIP), 2011 18th IEEE International Conference on*, IEEE, 2011, pp. 125–128.

[39] Huaizu Jiang et al., “Salient object detection: A discriminative regional feature integration approach”, *in*: *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, IEEE, 2013, pp. 2083–2090.

[40] X. Jiang, M. Le Pendu, and C. Guillemot, “Light field compression using depth image based view synthesis”, *in*: *2017 IEEE International Conference on Multimedia Expo Workshops (ICMEW)*, IEEE, 2017, pp. 19–24.

[41] Yong Ju Jung, Hak Gu Kim, and Yong Man Ro, “Critical binocular asymmetry measure for the perceptual quality assessment of synthesized stereo 3D images in view synthesis”, *in*: *IEEE Transactions on Circuits and Systems for Video Technology* 26.7 (2016), pp. 1201–1214.

[42] Martin Köppel et al., “Temporally consistent handling of disocclusions with texture synthesis for depth-image-based rendering”, *in*: *2010 IEEE International Conference on Image Processing*, IEEE, 2010, pp. 1809–1812.

[43] Lukáš Krasula et al., “On the accuracy of objective image and video quality models: New methodology for performance evaluation”, *in*: *Quality of Multimedia Experience (QoMEX), 2016 Eighth International Conference on*, IEEE, 2016, pp. 1–6.

[44] Patrick Le Callet, Christian Viard-Gaudin, and Dominique Barba, “A convolutional neural network approach for objective video quality assessment”, *in*: *IEEE Transactions on Neural Networks* 17.*5* (2006), pp. 1316–1327.

[45] Olivier Le Meur, Josselin Gautier, and Christine Guillemot, “Examplar-based inpainting based on local geometry”, *in*: *Image Processing (ICIP), 2011 18th IEEE International Conference on*, IEEE, 2011, pp. 3401–3404.

[46] Leida Li et al., “Quality Assessment of DIBR-Synthesized Images by Measuring Local Geometric Distortions and Global Sharpness”, *in*: *IEEE Transactions on Multimedia* 20.*4* (2018), pp. 914–926.

[47] Shuai Li, Ce Zhu, and Ming-Ting Sun, “Hole Filling with Multiple Reference Views in DIBR View Synthesis”, *in*: *IEEE Transactions on Multimedia* (2018).

[48] Wen-Nung Lie, Chia-Yung Hsieh, and Guo-Shiang Lin, “Key-Frame-Based Background Sprite Generation for Hole Filling in Depth Image-Based Rendering”, *in*: *IEEE Transactions on Multimedia* 20.*5* (2018), pp. 1075–1087.

[49] Suiyi Ling and Patrick Le 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*: *2018 25th IEEE International Conference on Image Processing (ICIP)*, IEEE, 2018, pp. 286–290.

[50] Suiyi Ling and Patrick Le Callet, “Image Quality Assessment for DIBR Synthesized Views using Elastic Metric”, *in*: *Proceedings of the 2017 ACM on Multimedia Conference*, ACM, 2017, pp. 1157–1163.

[51] Suiyi Ling, Patrick Le Callet, and Gene Cheung, “Quality assessment for synthesized view based on variable-length context tree”, *in*: *Multimedia Signal Processing (MMSP), 2017 IEEE 19th International Workshop on*, IEEE, 2017, pp. 1–6.

[52] Xiangkai Liu et al., “Subjective and objective video quality assessment of 3D synthesized views with texture/depth compression distortion”, *in*: *IEEE Transactions on Image Processing* 24.*12* (2015), pp. 4847–4861.

[53] LIVE lab, *LIVE Software Releases*, `http://live.ece.utexas.edu/research/Quality/`, last accessed Aug. 30th 2017, [Online].

[54] Guibo Luo et al., “A hole filling approach based on background reconstruction for view synthesis in 3D video”, *in*: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1781–1789.

[55] Kede Ma et al., “End-to-end blind image quality assessment using deep neural networks”, *in*: *IEEE Transactions on Image Processing* 27.*3* (2018), pp. 1202–1213.

[56] Yu Mao, Gene Cheung, and Yusheng Ji, “On Constructing z-Dimensional DIBR-Synthesized Images”, *in*: *IEEE Transactions on Multimedia* 18.*8* (2016), pp. 1453–1468.

[57] P. Merkle et al., “Multi-View Video Plus Depth Representation and Coding”, *in*: *2007 IEEE International Conference on Image Processing*, vol. 1, 2007, pp. I –201–I –204.

[58] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik, “Making a completely blind image quality analyzer”, *in*: *IEEE Signal Processing Letters* 20.*3* (2013), pp. 209–212.

[59] Anush Krishna Moorthy and Alan Conrad Bovik, “A two-step framework for constructing blind image quality indices”, *in*: *IEEE Signal processing letters* 17.*5* (2010), pp. 513–516.

[60] Anush Krishna Moorthy and Alan Conrad Bovik, “Blind image quality assessment: From natural scene statistics to perceptual quality”, *in*: *IEEE transactions on Image Processing* 20.*12* (2011), pp. 3350–3364.

[61] Anush Krishna Moorthy and Alan Conrad Bovik, “Visual importance pooling for image quality assessment”, *in*: *IEEE journal of selected topics in signal processing* 3.*2* (2009), pp. 193–201.

[62] Yuji Mori et al., “View generation with 3D warping using depth information for FTV”, *in*: *Signal Processing: Image Communication* 24.*1* (2009), pp. 65–72.

[63] Karsten Mueller et al., “View synthesis for advanced 3D video systems”, *in*: *EURASIP Journal on Image and Video Processing* 2008.*1* (2009), pp. 1–11.

[64] Patrick Ndjiki-Nya et al., “Depth image based rendering with advanced texture synthesis”, *in*: *2010 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2010, pp. 424–429.

[65] Patrick Ndjiki-Nya et al., "Depth image-based rendering with advanced texture synthesis for 3-D video", *in*: *IEEE Transactions on Multimedia* 13.*3* (2011), pp. 453–465.

[66] Nikolay Ponomarenko et al., "On between-coefficient contrast masking of DCT basis functions", *in*: *Proceedings of the third international workshop on video processing and quality metrics*, vol. 4, 2007.

[67] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", *in*: *arXiv preprint arXiv:1511.06434* (2015).

[68] ITURBT Recommendation, "500-11, Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT. 500-11", *in*: *ITU Telecom. Standardization Sector of ITU* 7 (2002).

[69] Dongni Ren et al., "Anchor view allocation for collaborative free viewpoint video streaming", *in*: *IEEE Transactions on Multimedia* 17.*3* (2015), pp. 307–322.

[70] Howard Rheingold, *Virtual reality: exploring the brave new technologies*, Simon & Schuster Adult Publishing Group, 1991.

[71] Michele A Saad, Alan C Bovik, and Christophe Charrier, "Blind image quality assessment: A natural scene statistics approach in the DCT domain", *in*: *IEEE Transactions on Image Processing* 21.*8* (2012), pp. 3339–3352.

[72] Michele A Saad, Alan C Bovik, and Christophe Charrier, "DCT statistics model-based blind image quality assessment", *in*: *2011 18th IEEE International Conference on Image Processing (ICIP),* IEEE, 2011, pp. 3093–3096.

[73] Neus Sabater et al., "Dataset and Pipeline for Multi-view Light-Field Video", *in*: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, IEEE, 2017, pp. 1743–1753.

[74] D. Sandić-Stanković et al., "Free viewpoint video quality assessment based on morphological multiscale metrics", *in*: *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1–6, DOI: `10.1109/QoMEX.2016.7498949`.

[75] Dragana Sandic-Stankovic, *MW-PSNR and MP-PSNR Source Code*, `https://sites.google.com/site/draganasandicstankovic/`, last accessed Aug. 30th 2017, [Online].

[76] Dragana Sandic-Stankovic, Dragan Kukolj, and Patrick Le Callet, “DIBR-synthesized image quality assessment based on morphological multi-scale approach”, *in*: *EURASIP Journal on Image and Video Processing* 2017.*1* (2016), p. 4.

[77] Dragana Sandić-Stanković, Dragan Kukolj, and Patrick Le Callet, “DIBR synthesized image quality assessment based on morphological pyramids”, *in*: *3DTV-Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON), 2015*, IEEE, 2015, pp. 1–4.

[78] Dragana Sandić-Stanković, Dragan Kukolj, and Patrick Le Callet, “DIBR synthesized image quality assessment based on morphological wavelets”, *in*: *2015 Seventh International Workshop onQuality of Multimedia Experience (QoMEX)*, IEEE, 2015, pp. 1–6.

[79] Dragana Sandić-Stanković, Dragan Kukolj, and Patrick Le Callet, “Multi–Scale Synthesized View Assessment Based on Morphological Pyramids”, *in*: *Journal of Electrical Engineering* 67.*1* (2016), pp. 3–11.

[80] ZM Parvez Sazzad, Yoshikazu Kawayoke, and Yuukou Horita, “No reference image quality assessment for JPEG2000 based on spatial features”, *in*: *Signal Processing: Image Communication* 23.*4* (2008), pp. 257–268.

[81] Damiel Scharstein, R Szeliski, and C Pal, *Middlebury stereo datasets*, 2012.

[82] Michael Schmeing and Xiaoyi Jiang, “Faithful disocclusion filling in depth image based rendering using superpixel-based inpainting”, *in*: *IEEE Transactions on Multimedia* 17.*12* (2015), pp. 2160–2173.

[83] Bong-Soo Sohn, Chandrajit Bajaj, and Vinay Siddavanahalli, “Feature based volumetric video compression for interactive playback”, *in*: *Proceedings of the 2002 IEEE symposium on Volume visualization and graphics*, IEEE Press, 2002, pp. 89–96.

[84] M. Solh, G. AlRegib, and J. M. Bauza, “3VQM: A vision-based quality measure for DIBR-based 3D videos”, *in*: *2011 IEEE International Conference on Multimedia and Expo*, 2011, pp. 1–6, DOI: `10.1109/ICME.2011.6011992`.

[85] Mashhour Solh and Ghassan AlRegib, “A no-reference quality measure for DIBR-based 3D videos”, *in*: *2011 IEEE International Conference on Multimedia and Expo (ICME)*, IEEE, 2011, pp. 1–6.

[86] Mashhour Solh and Ghassan AlRegib, “Hierarchical hole-filling for depth-based view synthesis in FTV and 3D video”, *in*: *IEEE Journal of Selected Topics in Signal Processing* 6.*5* (2012), pp. 495–504.

[87] Rui Song, Hyunsuk Ko, and CC Kuo, “MCL-3D: A database for stereoscopic image quality assessment using 2D-image-plus-depth source”, *in*: *arXiv preprint arXiv:1405.1403* (2014).

[88] Milan Sonka, Vaclav Hlavac, and Roger Boyle, *Image processing, analysis, and machine vision*, Cengage Learning, 2014.

[89] Gary J Sullivan et al., “Standardized extensions of high efficiency video coding (HEVC)”, *in*: *IEEE Journal of selected topics in Signal Processing* 7.*6* (2013), pp. 1001–1016.

[90] Huixuan Tang, Neel Joshi, and Ashish Kapoor, “Learning a blind measure of perceptual image quality”, *in*: *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* IEEE, 2011, pp. 305–312.

[91] Masayuki Tanimoto et al., “Free-viewpoint TV”, *in*: *IEEE Signal Processing Magazine* 28.*1* (2011), pp. 67–76.

[92] Masayuki Tanimoto et al., “Reference softwares for depth estimation and view synthesis”, *in*: *ISO/IEC JTC1/SC29/WG11 MPEG* 20081 (2008), p. M15377.

[93] Gerhard Tech et al., “Overview of the multiview and 3D extensions of high efficiency video coding”, *in*: *IEEE Transactions on Circuits and Systems for Video Technology* 26.*1* (2016), pp. 35–49.

[94] Alexandru Telea, “An image inpainting technique based on the fast marching method”, *in*: *Journal of graphics tools* 9.*1* (2004), pp. 23–34.

[95] *The Future of Video: Enabling Immersion*, `https://developer.att.com/blog/shape-future-of-video`.

[96] Dong Tian et al., “View synthesis techniques for 3D video”, *in*: *SPIE Optical Engineering+ Applications*, International Society for Optics and Photonics, 2009, 74430T–74430T.

[97] Dong Tian et al., “View synthesis techniques for 3D video”, *in*: *SPIE Optical Engineering+ Applications*, International Society for Optics and Photonics, 2009, 74430T–74430T.

[98] Shishun Tian et al., “A benchmark of DIBR Synthesized View Quality Assessment Metrics on a new database for Immersive Media Applications”, *in*: *IEEE Transactions on Multimedia* (2018).

[99] Shishun Tian et al., “A full-reference Image Quality Assessment metric for 3D Synthesized Views”, *in*: *Image Quality and System Performance Conference, at IS&T Electronic Imaging 2018*, Society for Imaging Science and Technology, 2018.

[100] Shishun Tian et al., “NIQSV: A NO REFERENCE IMAGE QUALITY ASSESSMENT METRIC FOR 3D SYNTHESIZED VIEWS”, *in*: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2017.

[101] Shishun Tian et al., “NIQSV+: A No-Reference Synthesized View Quality Assessment Metric”, *in*: *IEEE Transactions on Image Processing* 27.*4* (2018), pp. 1652–1664.

[102] Shishun Tian et al., “Performance comparison of objective metrics on free-viewpoint videos with different depth coding algorithms”, *in*: *Applications of Digital Image Processing XLI*, vol. 10752, International Society for Optics and Photonics, 2018, 107520O.

[103] Shishun Tian et al., “SC-IQA: Shift compensation based image quality assessment for DIBR-synthesized views”, *in*: *IEEE International Conference on Visual Communications and Image Processing*, 2018.

[104] Yu Tian et al., “CR-GAN: learning complete representations for multi-view generation”, *in*: *arXiv preprint arXiv:1806.11191* (2018).

[105] Anthony Vetro, Thomas Wiegand, and Gary J Sullivan, “Overview of the stereo and multiview video coding extensions of the H. 264/MPEG-4 AVC standard”, *in*: *Proceedings of the IEEE* 99.*4* (2011), pp. 626–642.

[106] Jiheng Wang et al., “Quality prediction of asymmetrically distorted stereoscopic 3D images”, *in*: *IEEE Transactions on Image Processing* 24.*11* (2015), pp. 3400–3414.

[107] Zhou Wang and A. C. Bovik, “A universal image quality index”, *in*: *IEEE Signal Processing Letters* 9.*3* (2002), pp. 81–84, ISSN: 1070-9908, DOI: 10.1109/97.995823.

[108] Zhou Wang and Alan C Bovik, “Modern image quality assessment”, *in*: *Synthesis Lectures on Image, Video, and Multimedia Processing* 2.*1* (2006), pp. 1–156.

[109] Zhou Wang, Alan C Bovik, and BL Evan, “Blind measurement of blocking artifacts in images”, *in*: *International Conference on Image Processing, 2000.* Vol. 3, IEEE, 2000, pp. 981–984.

[110] Zhou Wang and Qiang Li, “Information content weighting for perceptual image quality assessment”, *in*: *IEEE Transactions on Image Processing* 20.*5* (2011), pp. 1185–1198.

[111] Zhou Wang, Eero P Simoncelli, and Alan C Bovik, “Multiscale structural similarity for image quality assessment”, *in*: *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, 2004.* Vol. 2, IEEE, 2003, pp. 1398–1402.

[112] Zhou Wang et al., “Image quality assessment: from error visibility to structural similarity”, *in*: *IEEE Transactions on Image Processing* 13.*4* (2004), pp. 600–612, ISSN: 1057-7149, DOI: 10.1109/TIP.2003.819861.

[113] Soo Sung Yoon et al., “Inter-view consistent hole filling in view extrapolation for multi-view image generation”, *in*: *Image Processing (ICIP), 2014 IEEE International Conference on*, IEEE, 2014, pp. 2883–2887.

[114] Michael Yuen and HR Wu, “A survey of hybrid MC/DPCM/DCT video coding distortions”, *in*: *Signal processing* 70.*3* (1998), pp. 247–278.

[115] Z. Wang, *IW-SSIM Source Code*, https://ece.uwaterloo.ca/~z70wang/research/iwssim/, last accessed Aug. 30th 2017, [Online].

[116] Bo Zhao et al., “Multi-view image generation from a single-view”, *in*: *2018 ACM Multimedia Conference on Multimedia Conference*, ACM, 2018, pp. 383–391.

[117] Yin Zhao and Lu Yu, “A perceptual metric for evaluating quality of synthesized sequences in 3DV system”, *in*: *Visual Communications and Image Processing 2010*, International Society for Optics and Photonics, 2010, pp. 77440X–77440X.

[118] Tinghui Zhou et al., “Stereo Magnification: Learning view synthesis using multiplane images”, *in*: *arXiv preprint arXiv:1805.09817* (2018).

[119] Zhiqiang Zhou, Bo Wang, and Jinlei Ma, “Scale-aware edge-preserving image filtering via iterative global optimization”, *in*: *IEEE Transactions on Multimedia* (2017).

[120] Ce Zhu and Shuai Li, “Depth image based view synthesis: New insights and perspectives on hole generation and filling”, *in*: *IEEE Transactions on Broadcasting* 62.*1* (2016), pp. 82–93.

[121] Xiang Zhu and Peyman Milanfar, “A no-reference sharpness metric sensitive to blur and noise”, *in*: *2009 International Workshop on Quality of Multimedia Experience, QoMEx 2009.*. IEEE, 2009, pp. 64–69.

LIST OF FIGURES

List of Tables

INSA de RENNES
Service des Formations

AVIS DU JURY SUR LA REPRODUCTION DE LA THESE SOUTENUE

Titre de la thèse:
Image Quality Assessment of 3D Synthesized Views

Nom Prénom de l'auteur : TIAN SHISHUN

Membres du jury :
- Monsieur DEVERNAY Frédéric
- Madame MORIN Luce
- Madame ZHANG Lu
- Monsieur LE CALLET Patrick
- Monsieur LARABI Chaker
- Monsieur CAGNAZZO Marco
- Monsieur PUECH William

Président du jury : PATRICK LE CALLET

Date de la soutenance : 22 Mars 2019

Reproduction de la these soutenue

☑ Thèse pouvant être reproduite en l'état
□ Thèse pouvant être reproduite après corrections suggérées

Fait à Rennes, le 22 Mars 2019

Signature du président de jury

Le Directeur,

M'hamed DRISSI

INSTITUT NATIONAL DES SCIENCES APPLIQUEES
- RENNES -

UNIVERSITE BRETAGNE LOIRE / MATHSTIC

INSA RENNES

Titre : Évaluation de la qualité des images obtenues par synthèse de vues 3D

Mots clés : Depth-image-based-rendering (DIBR), View synthesis, Distortion, Image quality assessment, Quality of Experience (QoE)

Résumé: Depth-Image-Based Rendering (DIBR) est une technologie fondamentale dans plusieurs applications liées à la 3D, telles que la vidéo en mode point de vue libre (FVV), la réalité virtuelle (VR) et la réalité augmentée (AR). Cependant, l'évaluation de la qualité des vues synthétisées par DIBR a également posé de nouveaux problèmes, car ce processus induit de nouveaux types de distorsions, qui sont intrinsèquement différentes des distorsions provoquées par le codage vidéo. Ce travail est destiné à mieux évaluer la qualité des vues synthétisées par DIBR en multimédia immersif.

Au chapitre 2, nous proposons deux métriques complètements sans référence (NR). Le principe de la première métrique NR NIQSV consiste à utiliser plusieurs opérations morphologiques d'ouverture et de fermeture pour détecter et mesurer les distorsions, telles que les régions floues et l'effritement. Dans la deuxième métrique NR NIQSV+, nous améliorons NIQSV en ajoutant un détecteur de "black hole" et une détection "stretching".Au chapitre 3, nous proposons deux métriques de référence complète pour traiter les distorsions géométriques à l'aide d'un masque de désocclusion et d'une méthode de correspondance de blocs multi-résolution. Au chapitre 4, nous présentons une nouvelle base de données d'images synthétisée par DIBR avec ses scores subjectifs associés. Ce travail se concentre sur les distorsions uniquement induites par différentes méthodes de synthèse de DIBR qui déterminent la qualité d'expérience (QoE) de ces applications liées à DIBR. En outre, nous effectuons également une analyse de référence des mesures d'évaluation de la qualité objective de pointe pour les vues synthétisées par DIBR sur cette base de données. Le chapitre 5 conclut les contributions de cette thèse et donne quelques orientations pour les travaux futurs.

Title : Image quality assessment of 3D synthesized views

Keywords : Depth-image-based-rendering (DIBR), View synthesis, Distortion, Image quality assessment, Quality of Experience (QoE)

Abstract: Depth-Image-Based Rendering (DIBR) is a fundamental technology in several 3D-related applications, such as Free viewpoint video (FVV), Virtual Reality (VR) and Augmented Reality (AR). However, new challenges have also been brought in assessing the quality of DIBR-synthesized views since this process induces some new types of distortions, which are inherently different from the distortions caused by video coding. This work is dedicated to better evaluate the quality of DIBR-synthesized views in immersive multimedia.

In chapter 2, we propose a completely No-reference (NR) metric. The principle of the first NR metrics NIQSV is to use a couple of opening and closing morphological operations to detect and measure the distortions, such as "blurry regions" and "crumbling". In the second NR metric NIQSV+, we improve NIQSV by adding a "black hole" and a "stretching" detection. In chapter 3, we propose two Full-reference metrics to handle the geometric distortions by using a dis-occlusion mask and a multi-resolution block matching methods.In chapter 4, we present a new DIBR-synthesized image database with its associated subjective scores. This work focuses on the distortions only induced by different DIBR synthesis methods which determine the quality of experience (QoE) of these DIBR related applications. In addition, we also conduct a benchmark of the state-of-the-art objective quality assessment metrics for DIBR-synthesized views on this database. The chapter 5 concludes the contributions of this thesis and gives some directions of future work.