No Reference Opinion-free Stereoscopic Image Quality Assessment

a thesis submitted by

Sai Sree MG

in partial fulfilment of requirements
for the award of the dual degree of

**BACHELORS AND MASTERS
IN ELECTRICAL ENGINEERING**

Department of Electrical Engineering
Indian Institute of Technology Madras
Chennai
12 June, 2020

THESIS CERTIFICATE

This is to certify that the thesis titled **No Reference Opinion-free Stereoscopic Image Quality Assessment**, submitted by **Sai Sree MG**, to the Indian Institute of Technology, Madras, for the award of the degree of **Bachelors and Masters**, is a bona fide record of the research work done by her under our supervision. The contents of this thesis, in full or in parts, have not been submitted to any other Institute or University for the award of any degree or diploma.

Dr. Kaushik Mitra
Dept. of Electrical Engineering
IIT Madras, 600036

Place: Chennai
Date: 12 June, 2020

ACKNOWLEDGEMENTS

My adviser, Dr. Kaushik Mitra, has been a source of encouragement at every stage, from my first inclination to study Computational Photography to the writing of this report. I am grateful for the inspiration and the guidance that I have received from him, and most of all, for his generosity with his time and his concern for my growth as a researcher and a person.

Dr. Balasubramanyam Appina selflessly provided large volumes of academic and non-academic advice, pointers to literature, and help with my code, before and throughout the course of this project.

I would like to thank my family for giving support and guidance through out my life. I would also like to thank all my friends and well-wishers for helping me in difficult times and being a good source of support and guidance.

Abstract

Over the past few years, there have been a lot of advancements in the field of image processing. But the Human Visual System hasn't changed over years. The way in which we perceive images has remained the same. Hence it is important for us to assess the visual quality of image.

With the increase in 3-D videos and images, a lot of research is being done on the stereoscopic and virtual reality videos. With increasing algorithms and systems, the need to evaluate the quality of 3-D images is also increasing. Hence, we chose our problem statement to be No-reference opinion free quality assessment of stereoscopic images.

We have proposed a No reference opinion-free stereoscopic image quality assessment metric. There are 3 major steps. Step 1 involves generation of stereoscopic image dataset and the corresponding quality maps with 4 different distortions. In step 2, a deep network has been trained that predicts the quality map from distorted images without a reference image. In the last step, we have a pooling algorithm which gives us the objective metric from the predicted quality maps.

Contents

Chapter 1

Introduction

Image Quality Assessment (IQA) is the process of determining the visual quality of an image. There are many ways in which we can evaluate the quality of an image.

Figure 1.1: Quality of an image?

There are several of metrics in literature to assess the quality of image, but there are only a few metrics to predict the quality of stereoscopic images. With increasing interest in Stereoscopic technology for both entertainment and scientific purposes, we focus on No-reference Opinion-free Stereoscopic Image Quality Assessment in this project.

Figure 1.2: Stereoscopic Image Quality Assessment

1.1 Applications

Applications of Image Quality Assessment:

- Dynamically monitor and control multi-media services
- Benchmark for different image processing algorithms
- Optimise image processing systems

1.2 Types of IQA

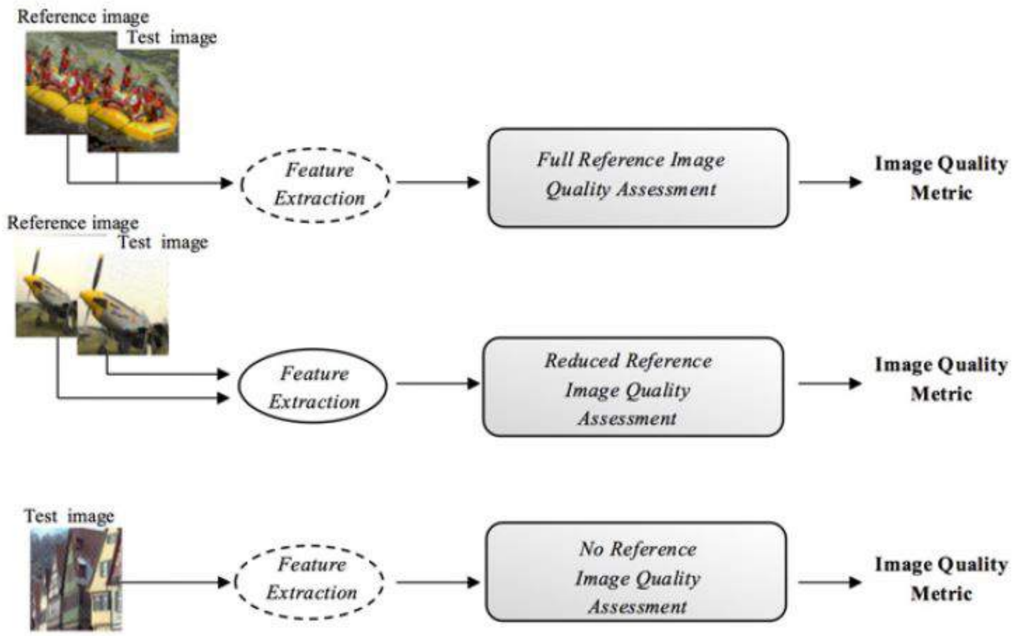
Image Quality Assessment is broadly classified into 2 types:

1.2.1 Subjective IQA

It is based on the way in which humans perceive quality. It is usually time consuming and inconvenient. Hence we need an algorithm based approach to predict visual quality of images.

1.2.2 Objective IQA

It is an algorithm based method to predict quality.



Objective IQA is can be of 3 types:

- **Full reference**: These methods use the test image and the reference (original) image to predict the quality of test image. It is usually more accurate, but the flip side is that the reference image is not available always.
- **Reduced reference**: These methods use extracted features from reference image and not the entire reference image to predict quality.
- **No reference**: These methods predict quality of an image by using only the test image, without the use of reference image. No reference IQA is further classified into 2 types:
    1. **Opinion aware**: Such methods require the subjective scores of training images to predict quality. It is usually a two stage framework where the first stage is to extract features from images and in next stage we learn a mapping function between extracted features and subjective scores.
    2. **Opinion free**: These methods use only test images to predict quality. They are usually expensive and time consuming. The correlation is subject to change with the quantity of images being used.

Our focus is on No-reference opinion free quality assessment of stereoscopic images.

Chapter 2

Prior work

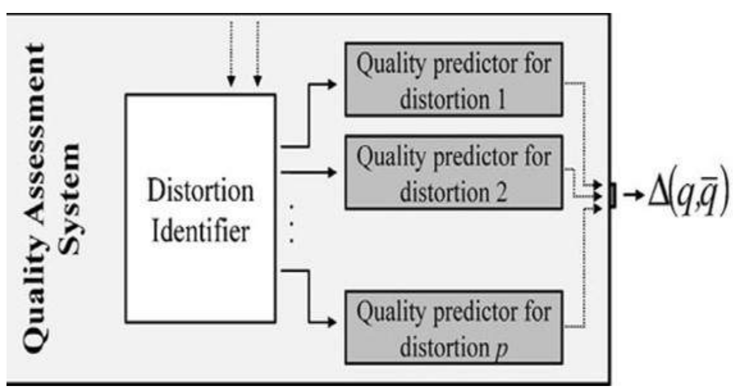
There are few state of the art models for no reference IQA. Two of them are discussed below:

2.1 BIQI

Blind Image Quality Index (BIQI):
In this paper, they have images with 5 different distortions: White Noise, Gaussian Blur, Fast Fading, JPEG and JPEG2000. The method that they proposed:

1. uses an SVM to classify the image into one of the five distortions. It predicts the probability distribution of distortion in the image

2. uses an SVR to predict quality index for each distortion separately



3. Quality index is computed as follows:

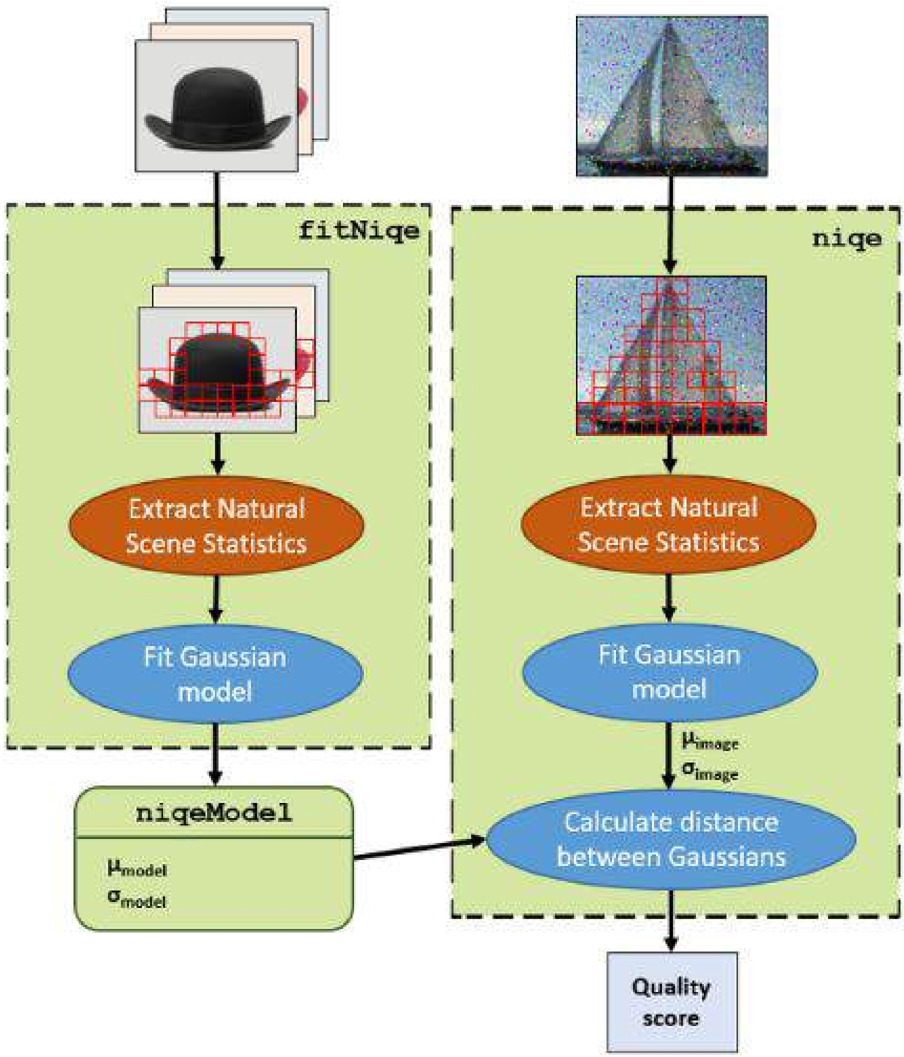
$$BIQI = \sum_{i=1}^{5} p_i \cdot q_i$$

2.2 NIQE

Natural Image Quality Evaluator (NIQE): This method is based on Natural Scene Statistics (NSS). We have a corpus of natural images. NSS features are extracted from each image by local mean removal and normalisation as follows:

$$\hat{I}(i,j) = \frac{I(i,j) - \mu(i,j)}{\sigma(i,j) + 1}$$

We add 1 in denominator to make sure if the variance in any of the patches is close to zero, the extracted features don't go to infinity.



Once we have extracted features, we fit a Bi-variate Gaussian model on those features and estimate mean and variance parameters for natural images. Repeat the same process for test images as well. The NIQE quality index is calculated as the distance between estimated mean and variance parameters for the natural image corpus and the given test image.

2.3 StereoQA-net

No reference Image Quality Assessment: This method is a No Reference opinion-aware Stereoscopic Image Quality Assessment. They have used Live Phase-1 stereo image dataset. Left and right image patches are extracted and the following network has been trained.

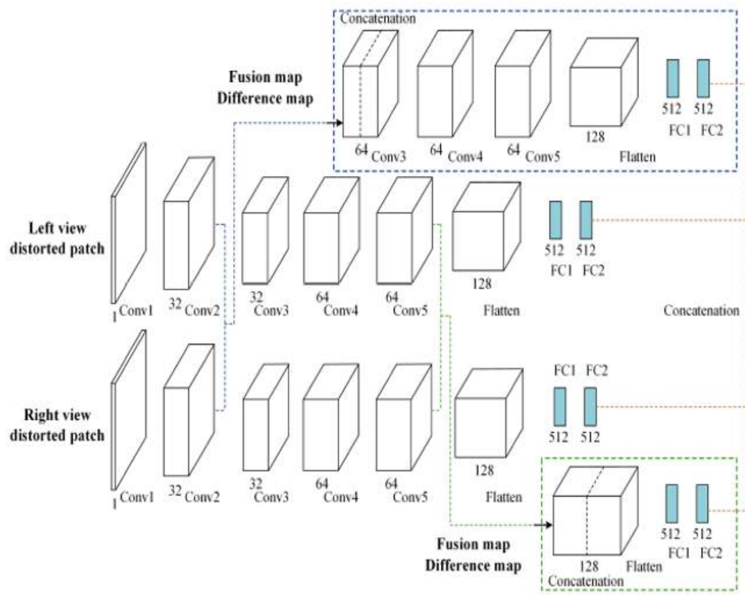


Fig. 1. The architecture of our proposed StereoQA-Net. The model takes as input left and right view distorted image patches, and conducts network interaction in multiple layers. The feature maps are subsampled by max pooling. After flattening convolutional layers and concatenating fully connected layers, the predicted quality score is regressed as a scalar output.

Fusion and difference maps are also computed using the features extracted from the network. Subjective scores have been used while training.

Chapter 3

Proposed method

The above methods, NIQE and BIQI are for single image. But for the case of stereoscopic images, the disparity between left and right images also comes into picture. StereoQA net is an opinion aware IQA metric. It requires a dataset with subjective scores, which is time consuming for huge data. Hence we tried to build a stereo IQA metric which is opinion-free.

3.1 Dataset

The dataset we used is LIVE Phase - I dataset with 120 pristine stereo image pairs. We have added 4 different distortions to pristine images, each image with 6 levels of distortion. We also considered asymmetric distortion, hence we have a total of 36 image pairs generated from each image pair per distortion.

Figure 3.1: 4 pairs of pristine images

(a) Left

(b) Right

(c) Left

(d) Right

(e) Left

(f) Right

(g) Left

(h) Right

Example image with distortion all levels of distortion applied on it is shown below.

Figure 3.2: Distorted images with increasing level of distortion

(a) Gaussian Blur (b) White Noise (c) JPEG (d) JP2K

3.2 Algorithm

The proposed method has 3 major steps:

1. Quality maps generation
2. Deep network to predict quality maps
3. Pooling algorithm to obtain quality score

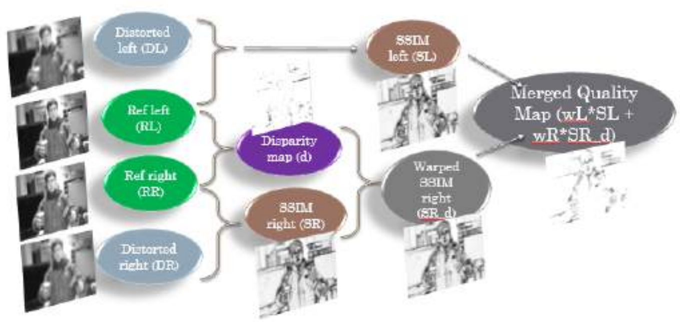
3.2.1 Quality Maps generation

There are two different ways in which we have generated the quality maps. Initially, the image format has been converted into *double* before computing the quality map. Later on, the image format has been kept as *unit8* for preciseness of the mapas computed.

- **Method - 1**

  In the first step, we generate quality maps of distorted stereo images generated with the use of reference image pairs.
  Structural SIMilarity (SSIM) map is one of the most used full reference image quality metric. It basically removes the effect of luminosity and contrast from the images and considers the impact of only structural similarity between right and left images which is what exactly human beings perceive while looking at an image.
  We compute the SSIM maps of left and right images separately. They can't be directly added since we have to also consider the impact of disparity between left and right images. Hence disparity maps are also computed between the left and right images. Left disparity map is computed pixel-wise by keeping right image as reference and horizontally shifting left image by different number of pixels and we choose that shift with least error between left and right images. Similarly, right disparity map is computed by keeping left image as reference and shifting right image. We have assumed that the disparity is only in horizontal direction.

  Figure 3.3: Quality map generation

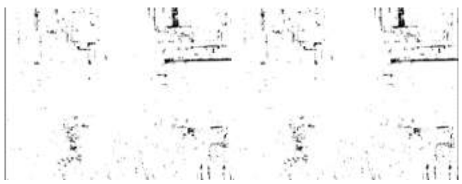
  

  - **Symmetric distortion:**
    With symmetric distortion, there are a lesser number of black pixels in disparity map. Number of black pixels in quality map is increasing with level of distortion. But there is no significant improvement in the number of black pixels in the disparity map. Since the distortion in both left and right images is the same, the disparity maps of both the image pairs are also similar. The images are shown below.

Figure 3.4: Symmetric distortion with level 1 of distortion

(a) Image pair

(b) Disparity Maps



(c) Quality Map

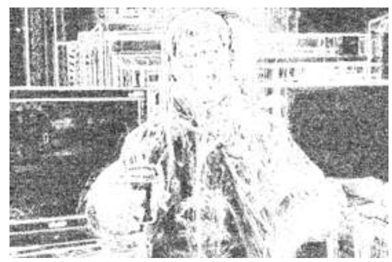
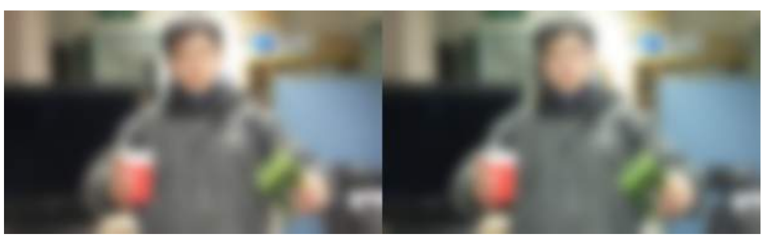
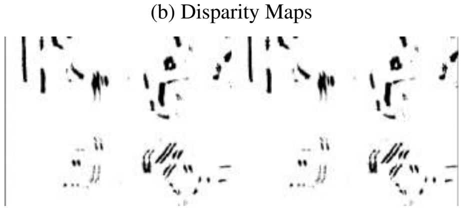


Figure 3.5: Symmetric distortion with level 6 of distortion

(a) Image pair



(b) Disparity Maps

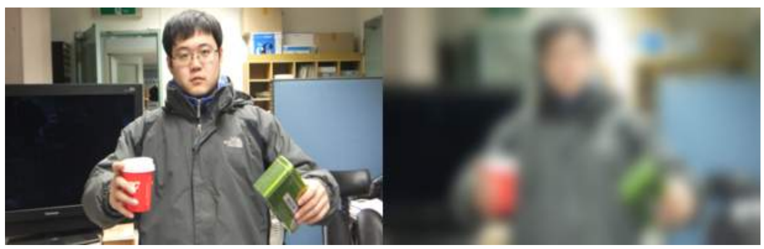


(c) Quality Map

- **Asymmetric distortion:**
  With asymmetric distortion, disparity map has a higher number of black pixels indicating higher disparity which is expected since the images are not similar, the distortion level is different for each image from the pair. But quality map still is considerably white since at least one image is clear and more weightage will be given to the image with lesser variance in its quality map.

Figure 3.6: Asymmetric distortion with level 1 of distortion in left image and level 6 of distortion in right image

(a) Image pair



(b) Disparity Maps

(c) Quality Map

- **Method - 2**

  This method is just the same as the previous method, but the format of images hasn't been changed. It is kept as *uint8*. The number of levels of distortion has been increased from 6 to 9.

  - **Symmetric distortion:**
    Similar results, there are a lesser number of black pixels in disparity map. Number of black pixels in quality map is increasing with level of distortion.

Figure 3.7: Symmetric distortion with level 2 of distortion

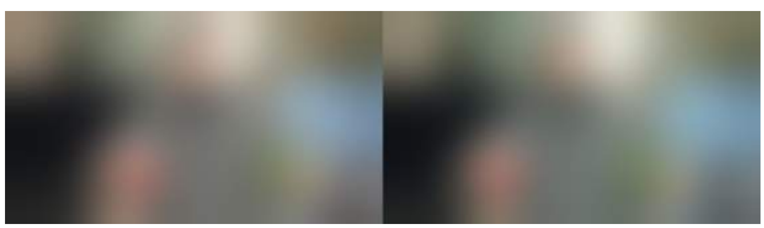
(a) Image pair

(b) Disparity Maps

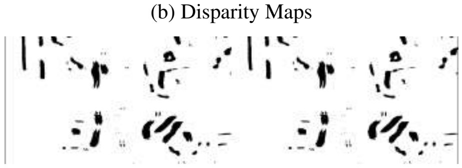
(c) Quality Map

Figure 3.8: Symmetric distortion with level 9 of distortion

(a) Image pair



(b) Disparity Maps

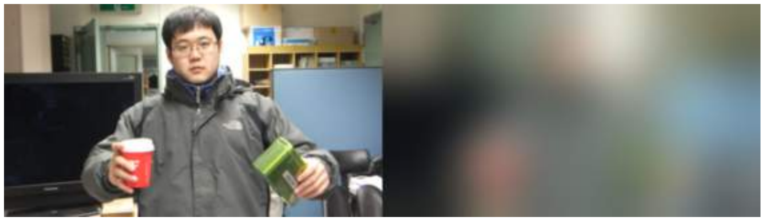


(c) Quality Map

- **Asymmetric distortion:**
  Same as before, disparity map has a higher number of black pixels indicating higher disparity which is expected since the images are not similar. But quality map still is considerably white since at least one image is clear.

Figure 3.9: Asymmetric distortion with level 2 of distortion in left image and 9 of distortion in right image

(a) Image pair



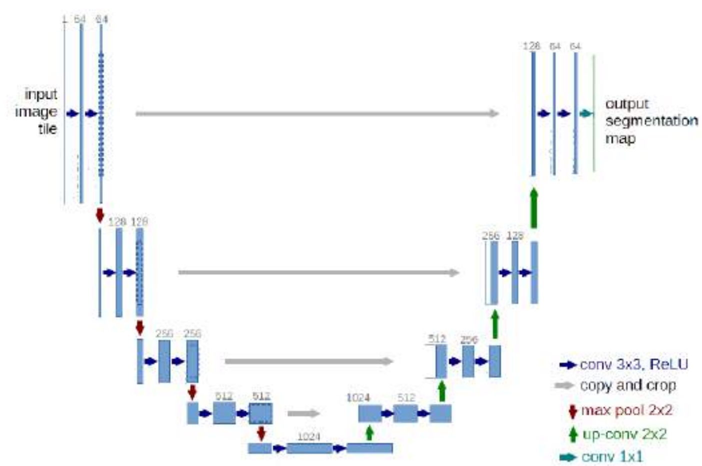
(b) Disparity Maps

(c) Quality Map

The same pattern is being followed in both the methods, but the maps generated from the latter method better explain the distorted images.

3.2.2 Deep network architecture

The following deep network has been implemented in Pytorch to predict quality maps left and right images without the use of reference image.



- **Training the network**

  The network has been initially trained over the data generated in PyTorch with Adam optimizer, L1 norm has been used as loss function.

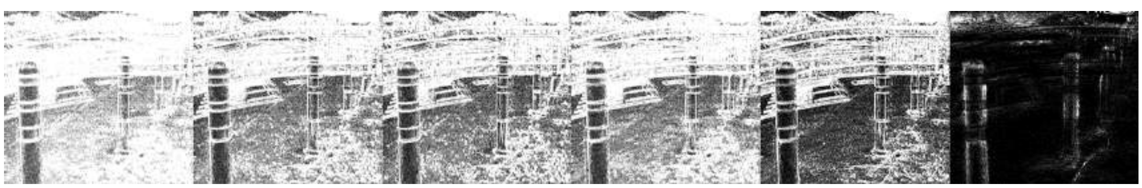
  These are the results for a few of the images from validation data, the distortion in all of the right images is level 6, with the distortion in left image varying from level 1 to level 6.

Figure 3.10: True vs predicted quality maps of test images from same dataset

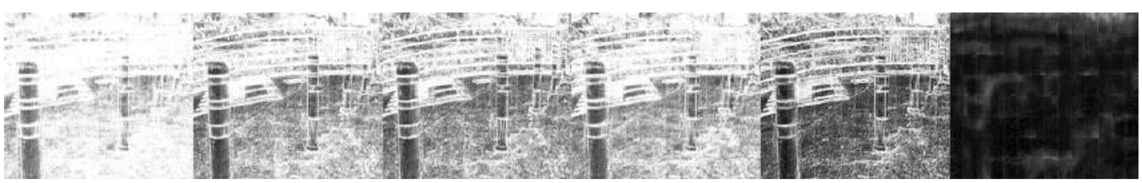
(a) Right image

(b) Left images

(c) True Maps

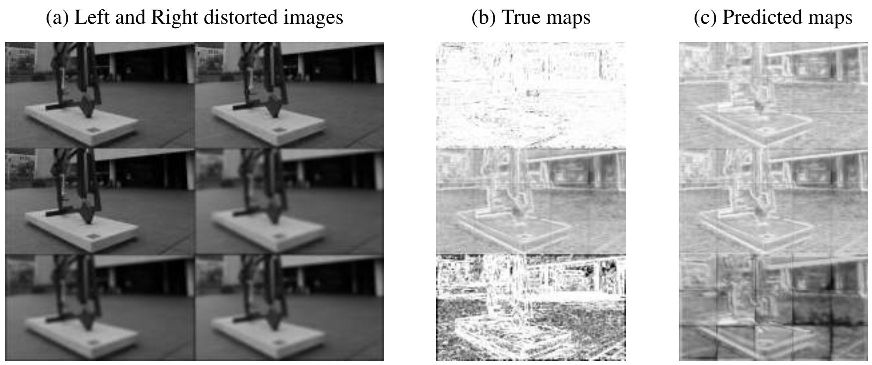


(d) Predicted maps



These are the results for a few of the images from LIVE Phase 2 data.

Figure 3.11: True vs predicted quality maps of test images from LIVE Phase 2 dataset



3.2.3 Pooling Algorithm

Once we have predicted quality maps, we'll then apply a pooling algorithm on predictions to get an objective score. Our goal is to improve correlation between subjective and objective scores.

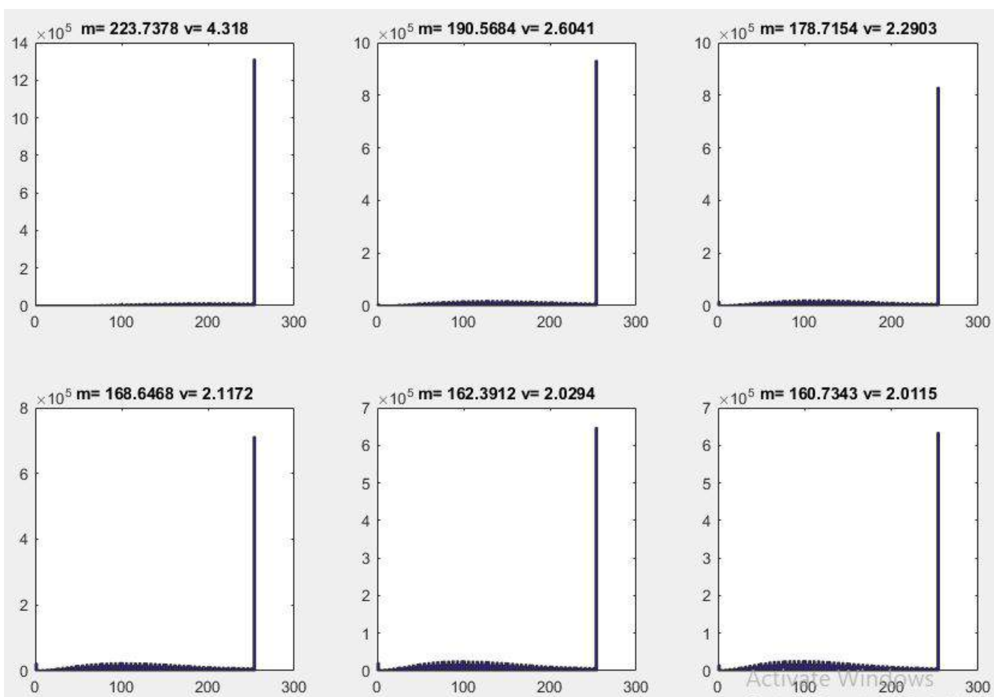


Figure 3.12: Histograms of quality maps with increasing distortion

As we can clearly see from the picture, mean and variance are reducing with distortion. Also, the number of pixels with 255 is also reducing with increasing level of distortion. Hence the possible ways of pooling could be:

- **Mean**: It is the simplest possible metric that can be used as score. The correlation achieved by using mean is **0.44**.
- **Variance**: The correlation obtained by using variance is **0.4**
- **Ratio of mean and variance**: correlation obtained is **0.39**
- **Weighted mean**: Image has been divided into patches and weighted mean has been used with entropy of each patch to be the weight. Correlation of **0.54** has been achieved.
- **NIQE**: The NIQE index of the quality map can be used, correlation of **0.3** has been achieved.

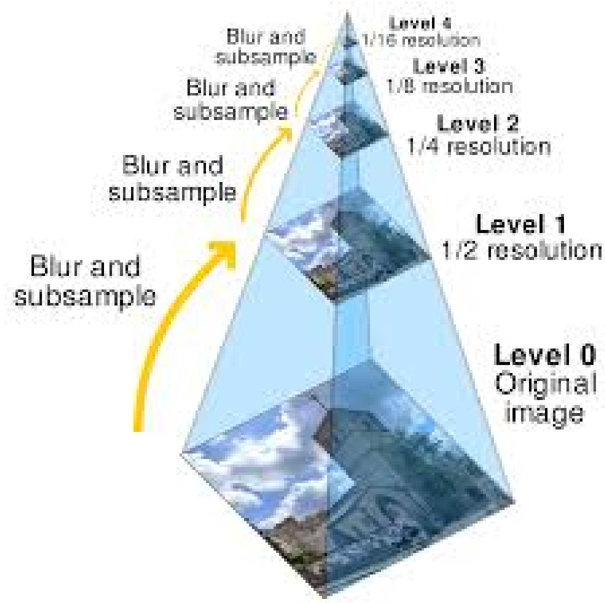
The table below shows the correlation values with different pooling methods.

Feature	Correlation
Mean	0.4
Variance	0.54
Weighted mean	0.44
NIQE	0.3
Entropy	0.58

Figure 3.13: Correlation values

Steerable Pyramid Decomposition

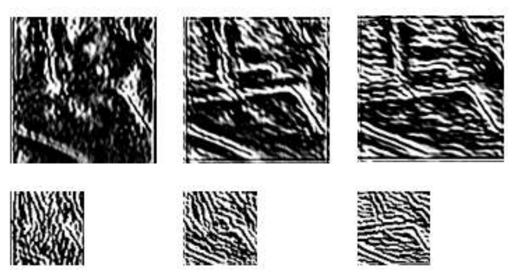
Multiscale decomposition has been performed before computing the objective score. Steerable pyramid decomposition is a linear multi resolution decomposition method, which involves blurring and subsampling an image. An image is divided into a collection of sub bands localized at different scales and orientations. The images below illustrate steerable pyramid decomposition.



(a) Steerable pyramid decomposition

Steerable pyramid decomposition with 2 scales and 3 orientations has been applied on the predicted maps, an example is shown below:

(b) Predicted Map



(c) Sub bands

Figure 3.14: Predicted quality map and its sub bands

The patch-wise variance and entropy have been calculated for each of the sub bands and a weighted mean of of entropy (with variance as the weight) has been calulated as objective score for each of the quality maps. The table below shows the correlation achived with steerable pyramid decomposition:

Feature	Mean	Variance	Entropy	<Variance,Entropy>
Correlation	**0.42**	**0.71**	**0.59**	**0.82**

Figure 3.15: Correlation values with steerable pyramid decomposition

Therefore, a correlation of **0.8** has been achieved with a No-reference opinion-free method that we have proposed.

Future work

- Extend the same method to other distortions as well
- Extend this to stereoscopic videos

References

1. A Two-Step Framework for Constructing Blind Image Quality Indices, Anush Krishna Moorthy and Alan Conrad Bovik, Fellow, IEEE

2. Making a 'Completely Blind' Image Quality Analyzer, Anish Mittal, Rajiv Soundararajan and Alan C. Bovik, Fellow, IEEE

3. Dual-Stream Interactive Networks for No-Reference Stereoscopic Image Quality Assessment, Wei Zhou , Zhibo Chen , Senior Member, IEEE, and Weiping Li, Fellow, IEEE

4. Predictive Auto-Encoding Network for Blind Stereoscopic Image Quality Assessment, Jiahua Xu*, Wei Zhou*, Zhibo Chen, Senior Member, IEEE, Suiyi Ling, Student Member, IEEE, and Patrick Le Callet, Fellow, IEEE

5. Image courtesy:

   (a) 3-D viewer: https://3dthis.com/stereosample.htm

   (b) http://ivc.univ-nantes.fr/en/pages/view/43/

   (c) https://www.researchgate.net/publication/313234607_Knowledge-based_Taxonomic_Scheme_for_Full-Reference_Objective_Image_Quality_Measurement_Models/figures?lo=1

   (d) https://in.mathworks.com/help/images/train-and-use-a-no-reference-quality-assessment-model.html

   (e) https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

   (f) https://en.wikipedia.org/wiki/Pyramid_(image_processing)