

Subjective and Objective Quality Assessment of Stitched Images for Virtual Reality

A THESIS
SUBMITTED FOR THE DEGREE OF
Master of Technology (Research)
IN THE
Faculty of Engineering

BY
Pavan C M

under the Guidance of
Dr. Rajiv Soundararajan



Electrical Communication Engineering
Indian Institute of Science
Bangalore – 560 012 (INDIA)

July, 2018

© Pavan C M

July, 2018

All rights reserved

DEDICATED TO

My Parents

Veena and Madhusudana Rao

Acknowledgements

No thesis acknowledgement is complete without thanking one's advisor given the magnitude of things one gets to learn under their guidance. I consider myself fortunate as well as privileged to have worked under Dr. Rajiv Soundararajan without whom I wouldn't have known what research is all about. He has molded me from a naive undergrad who had very nascent idea about research, into taking up research as a career. His approach of giving a lot of exploratory freedom while making sure we are not lost by providing critical suggestions and new ideas was instrumental in arriving at the desired results. I have deep admiration for his commitment and tremendous patience, particularly in reviewing papers and presentations, and also his attention to detail is something which I have tried to imbibe, but it appears to be a long shot dream.

I am indebted to IISc for providing an amazing environment and resources for pursuing research. IISc's gorgeous campus is quite addictive and makes the IISc experience unforgettable as well as something which is hard to part away with. I would like to thank the faculty members at IISc, especially from ECE, EE and CSA departments, where I attended several interesting, intriguing as well as challenging courses.

I would like to thank my labmates Sameer, Biju, Deepankar and others for helping and supporting me and making the lab environment more enjoyable. The discussions we have had in the lab on wide ranging topics that come under the sun were something which made the lab experience memorable.

My experience at IISc outside lab was special because of the company of many wonderful friends. Thanks to Navaneet, Shreyas, Srinivas, Anantha, Prabhasa, Varadaraj, Vinith, Deekshith and the rest of A-mess gang for sharing quite memorable and entertaining moments in the mess. I would like to thank all the volunteers who took part in the subjective study by spending their valuable time for providing data for my work, which played a crucial role in evaluating our algorithm.

Last but not the least, I have deepest gratitude for my parents whose encouragement and support have made me what I am today.

Abstract

We consider the problem of quality assessment (QA) of image stitching algorithms used to generate panoramic images for virtual reality applications. Our contributions are two fold. We design the Indian Institute of Science Stitched Image QA (ISIQA) database consisting of 264 stitched images obtained by employing multiple stitching algorithms on images captured from 26 diverse scenes. The database has a wide perceptual quality spread and consists of a variety of artifacts due to stitching such as blur, ghosting, color and geometric distortions. We subjectively evaluate these images by acquiring 6600 human quality ratings by for these images viewed on a virtual reality head mounted device. We then devise an objective QA model called the Stitched Image Quality Evaluator (SIQE) using the statistics of steerable pyramid decompositions. In particular, we propose a Gaussian mixture model to capture the bivariate statistics of neighboring coefficients of steerable pyramid decompositions and show this to be effective in modeling the increased spatial correlation due to ghosting artifacts. An important characteristic of the proposed algorithm is that it requires no knowledge of the stitching algorithm used to obtain the panorama and only needs individual images and stitched image as input for predicting quality. We show through extensive experiments that our quality model outperforms the existing No-Reference and Full-Reference QA models and correlates very well with subjective scores in the ISIQA database. The ISIQA database as well as the software release of SIQE have been made available online for public use and evaluation purposes.

Publications based on this Thesis

1. Pavan Chennagiri Madhusudana and Rajiv Soundarajan, "Subjective and Objective Quality Assessment of Stitched Images for Virtual Reality," submitted, *IEEE Transactions on Image Processing*, June 2018.

Contents

- Acknowledgements i

- Abstract iii

- Publications based on this Thesis iv

- Contents v

- List of Figures vii

- List of Tables ix

- 1 Introduction 1**
 - 1.1 Related Work 2
 - 1.2 Contributions 4
 - 1.3 Organization 5

- 2 Subjective Quality Assessment 6**
 - 2.1 Database 6
 - 2.1.1 Feature Detection and Matching 8
 - 2.1.2 Image warping 9
 - 2.1.3 Image Blending 12

2.2	Subjective Test Methodology	15
2.3	Processing of Subjective Scores	16
2.4	Anchor Images	18
2.5	Analysis of Color Distorted Images	19
2.6	Analysis of Subjective Response Times	20
3	Stitched Image Quality Evaluator	21
3.1	Bandpass Analysis of Image Stitching	21
3.2	Marginal Statistics Model with Divisive Normalization	24
3.3	Bivariate Model	27
3.4	Patch Weighting	30
3.5	Prediction	32
4	Experiments and Results	33
4.1	Comparison of Models for Bivariate Distribution	33
4.2	Correlation With Human Judgments	35
4.3	Significance of Each Conceptual Feature	38
4.4	Application of Univariate GMM for capturing correlation	39
4.5	Effect of Quantization Levels in Patch Weighting	40
4.6	Time Complexity	41
5	Conclusion and Future Work	42
	Appendix A Decrease in Variance due to Ghosting	44
	References	46

List of Figures

1.1	Problem illustration	2
2.1	Image stitching pipeline with algorithms associated with each stage	7
2.2	Keypoint Detection	8
2.3	Keypoint Matching	8
2.4	Outlier removal	8
2.5	Illustration of stitching inconsistencies due to homography warp	9
2.6	Ghosting artifacts due to homography	10
2.7	Warping using local homography	10
2.8	Distortions in non-overlapping regions due to local homography	11
2.9	Ghosting distortion in SPHP warp	11
2.10	Illustration of various blending methods	13
2.11	Images illustrating various distortions due to stitching. The images shown above are cropped versions of the images present in the database.	14
2.12	Histogram of MOS	17
2.13	MOS of anchor images plotted with number of subjects along with 95% confi- dence intervals	18
2.14	Comparison of MOS of color distorted images (left column) with images having no artifacts (right column)	19
2.15	Average response time for various distortions	20

3.1	High Level Overview of SIQE Framework	22
3.2	Illustration of homography warp. Red boxes indicate overlapping regions	23
3.3	Illustration of orientation selectivity	24
3.4	Comparison of marginal statistics with divisive normalization in presence of ghosting	25
3.5	Comparison of marginal statistics in presence of geometric distortion	25
3.6	Joint and conditional statistics in presence of ghosting. The above histograms are plotted for $s_1^{60^\circ}$ for patches shown in Fig. 3.4	28
3.7	Illustration of GLCM energy values for patches with varying amounts of texture	31
3.8	Effect of σ on function g	31
4.1	Ghosting example	34
4.2	2D illustration of probability distributions for GMM and BGGD fits of undistorted patch. The above distributions represent subband $s_1^{0^\circ}$ for patch shown in Fig. 4.1a	34
4.3	2D illustration of probability density functions for GMM and BGGD fits of ghosted patch. The above distributions represent subband $s_1^{120^\circ}$ for ghosted patch shown in Fig. 4.1b	34
4.4	Comparison of median SROCC distribution across multiple QA algorithms	37
4.5	Effect of number of quantization levels of GLCM on the performance, median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets	40

List of Tables

2.1	Number of images present across categories	7
3.1	Summary of extracted features	30
4.1	Comparison of GMM and BGGD fits for pristine patches for horizontal neighbors	35
4.2	Comparison of GMM and BGGD fits for distorted patches for horizontal neighbors	35
4.3	Median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets (subjective MOS vs predicted MOS)	36
4.4	Standard deviation of SROCC and LCC values for 1000 iterations of randomly chosen train and test sets	37
4.5	Comparison of Median values of SROCC and LCC with full reference QA algorithms on a subset of 238 images in the ISIQA database	38
4.6	Each conceptual feature in isolation of other features, median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets	39
4.7	Performance comparison using univariate GMM, median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets	40
4.8	Complexity analysis in SIQE. Values reflect the percentage of time spent on each step in SIQE	41
4.9	Comparison of the amount of time taken to compute various quality measures for a 100×100 image patch	41

Chapter 1

Introduction

Virtual Reality (VR) refers to technologies that enable the virtual presence of an individual in a different environment. In recent days VR is gaining significant popularity due to its wide applications ranging from education, motion pictures, games to medicine and healthcare. VR aims to provide a realistic and immersive experience in virtual environments to the users through wide field of view images or panoramic images and videos viewed on head mounted displays (HMD). HMD devices like Google Cardboard, Oculus Rift etc. are increasingly becoming popular and are widely getting adopted by mainstream media and general public. Also the development of specialized camera rigs of multiple cameras such as Facebook Surround 360 [2], and handheld single cameras consisting of multiple lenses like Ricoh Theta S, have made it simpler to capture, stitch and broadcast. Panoramic images are typically obtained by capturing images with overlapping areas and stitching together the wide field of view image. Thus image stitching forms an important aspect of generating panoramic or wide field of view images. A large number of image stitching algorithms have been developed in literature [3, 4, 5, 6] including several commercial algorithms like Autostitch [7] and Microsoft Image Composite Editor [8]. Nevertheless, the quality assessment of panoramic images with respect to stitching algorithms has been much less studied. Such assessment will help benchmark, compare and

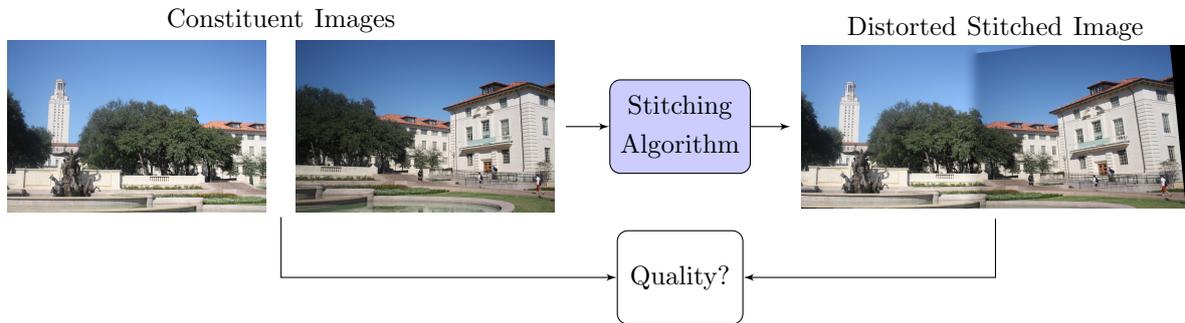


Figure 1.1: Problem illustration

tune the performance of various stitching algorithms. We focus on the problem of subjective and objective quality assessment (QA) of image stitching in the context of VR applications. Fig. 1.1 illustrates the high level block diagram of the problem we are addressing in this work. For the design of objective algorithms, we assume that the QA model has access to the stitched image in addition to the individual images that are stitched.

1.1 Related Work

Objective QA algorithms can be broadly classified as Full Reference (FR) algorithms and No Reference (NR) algorithms. FR algorithms are provided with a pristine unimpaired stimulus along with the stimulus whose quality is to be measured as input [9, 10]. NR methods predict the quality of the impaired stimulus without any information about the pristine stimulus [11, 12, 13]. Note that the problem setting we consider does not fall under the FR category due to the non-availability of a true undistorted panorama. Further, the problem cannot be strictly categorized under the NR setting as some information about the reference is available in the form of individual images (or constituent images) that undergo stitching. Thus the challenge in the QA of image stitching lies in how reference information from the individual images can be used to reliably assess the quality of the stitched image. Further, distortions that arise due to image stitching such as ghosting, geometric errors and color artifacts are very specific to panoramas

and existing generic QA algorithms do not effectively account for these distortions.

Existing stitched image QA algorithms [14, 15, 16] mainly account for the presence of specific artifacts like photometric and geometric distortions. In [15], the structural similarity index (SSIM) of high frequency information of the stitched and unstitched images in overlapping regions is used as a measure of geometric error. Quality indices proposed in [14, 16] only address color correction and intensity consistencies. Cheung *et al.* [17] construct a database of stitched images from virtual scenes and model geometric inconsistencies using optical flow field energy. In [18], the gradient of luminance differences between the stitched and reference images is used to measure geometric error. The frameworks proposed in [17, 18] differ from ours in the availability of a reference stitched image. In [17], each camera view is used as a full reference of the stitched view of its left and right adjacent images, which requires a careful arrangement of cameras. Since every stitched image is obtained via a stitching algorithm, a reference stitched image is generally not available in all situations.

Recently, a few QA algorithms for 360° images (omnidirectional images) have been proposed to capture compression artifacts arising from JPEG, JPEG2000, HEVC etc. [19, 20, 21, 22, 23, 24]. In [19] a spherical-PSNR (S-PSNR) was proposed by calculating PSNR values between corresponding pixels in the spherical domain. A craster parabolic projection based PSNR (CPP-PSNR) was proposed by Zakharchenko *et al.* [20] as a quality measure. Sun *et al.* [21] use a weighted spherical PSNR (WS-PSNR) to compare distorted images, where the weights are computed based on pixel locations on the spherical surface. In [22, 23] the performance of existing FR image QA (IQA) metrics are compared on a database of omnidirectional images suffering from compression distortions. In [24] an adversarial learning approach has been proposed to determine the quality of distorted images, where the generator and discriminator networks are learned using a database of distorted images, their corresponding human opinion scores as well as reference images. All of the above methods mainly address artifacts arising from compression and do not explicitly model stitching induced distortions.

1.2 Contributions

Our main contributions are in the design of a stitched image QA database and algorithms that capture stitching induced distortions and provide quality scores that correlate well with human judgments. We provide further details of our contributions as follows:

Subjective Quality Assessment

We design a dataset containing 264 panoramic images obtained using various stitching algorithms from images captured from 26 diverse scenes. We conduct a large scale subjective study to evaluate the images in the database by acquiring around 25 human opinion scores for each image. The subjective study involves obtaining human ratings when images are viewed on a HMD as applicable to virtual reality. The images in the dataset as well as their corresponding human ratings are used to train and evaluate objective QA models.

Objective Quality Assessment

We design the Stitched Image Quality Evaluator (SIQE) by devising statistical features using steerable pyramid decompositions. We propose a Gaussian mixture model to model the bivariate statistics of pairwise neighbors of steerable pyramid decompositions and show this to be an effective model. We then apply this model to capture the ghosting artifacts introduced due to image stitching. We design our model independent of the knowledge of the stitching algorithm that may have been used to stitch the images. We adopt this approach to enable our model to be robust to image key point mismatches inherent in the stitching algorithm as well as the lack of knowledge of the stitching algorithm itself. We evaluate our model on the database we develop and show that the estimated quality correlates very well with the Mean Opinion Scores (MOS) obtained from the subjective study.

1.3 Organization

The remainder of this thesis is organized as follows. In Chapter 2 we describe the detailed procedure employed for creation of the stitched image quality assessment database. In this Chapter, we also describe the methods employed for processing the human opinion scores obtained from the subjective study. In Chapter 3 we explain our proposed QA model, where we discuss the motivation and design of features employed in our proposed framework. In Chapter 4 we discuss the extensive experiments we conduct to evaluate our proposed model and, we report and analyze these experimental results. Finally the thesis is concluded in Chapter 5, with discussion on some future directions of the problem.

Chapter 2

Subjective Quality Assessment

In this chapter we provide details about the subjective quality assessment of stitched images. Firstly, we discuss the methods employed for creating the stitched image quality assessment database. Secondly, we provide details of the subjective study and methods used for processing the human opinion scores obtained from the subjective study. We next analyze the influence of number of subjects on the mean opinion score obtained for each image. We also analyze the response time of the subjects on the basis of various types of distortions present in the database.

2.1 Database

We create the Indian Institute of Science (IISc) Stitched Image Quality Assessment (ISIQA) database consisting of 264 stitched images obtained from 26 diverse scenes. Every scene consists of multiple images with overlapping view points obtained through horizontal panning motion of the camera. Images were captured using a Samsung Galaxy S7 Edge smartphone camera with a resolution of 4032×2268 for every viewpoint image. During image capture, care was taken to avoid object motion between successive images since the focus of this study is in evaluating the performance of stitching algorithms for static scenes. Recent studies on saliency in VR have

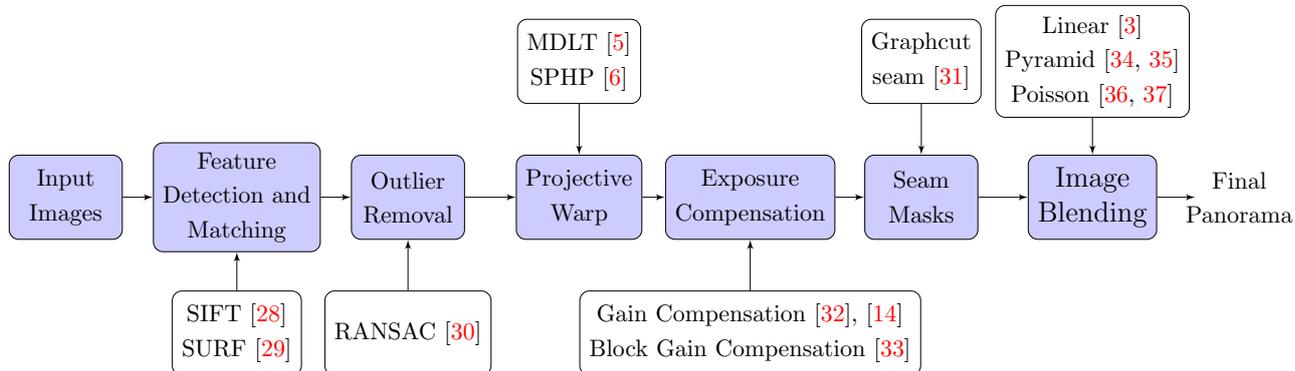


Figure 2.1: Image stitching pipeline with algorithms associated with each stage

	Buildings	Gardens	Indoor	Public Places	Total
Number of scenes	9	8	3	6	26
Number of stitched images	96	71	34	63	264

Table 2.1: Number of images present across categories

revealed a strong *equator bias* persistent among the subjects [19, 25, 26, 27], where viewers are more likely to observe content in equator adjacency rather than near the poles in a spherical projection. Since we focus on evaluating images obtained through horizontal image stitching, we ignore the top and bottom views in our study. Every scene consists of 4 – 5 images obtained along multiple viewpoints resulting in stitched images with horizontal field of view lying in the range $180^\circ - 270^\circ$ and having horizontal resolution around $8K - 10K$ (vertical resolution will be around $2K$ since the top and bottom views are absent and vertical field of view in the range $90^\circ - 120^\circ$). The 26 scenes consist of a wide variety of scenes including buildings, gardens, indoor and public places. Each set of overlapping viewpoint images corresponding to a scene is stitched using different algorithms to create the database of stitched images. The distribution of the stitched images present in each scene category is shown in Table 2.1.

Image stitching involves a series of operations from the constituent images with overlapping fields of view to the stitched image. A typical stitching pipeline shown in Fig. 2.1 illustrates the algorithms employed in each stage. By varying the choice of specific algorithms for each stage as well as the parameter options for these individual algorithms, we create the database of

stitched images having a wide perceptual quality spread. The effect of each stage present in the stitching pipeline on the quality of the stitched image is discussed in the following subsections.

2.1.1 Feature Detection and Matching

The first step in stitching procedure is to align the individual images by identifying the corresponding points in the overlapping regions. Keypoint detection algorithms such as SIFT [28], SURF [29] etc. are some of the widely used methods to detect the corresponding points. We observed that the quality of the final stitched image is not particularly sensitive to the choice of keypoint detection algorithm. Therefore we employ SIFT for all the images in the ISIQA database. Applying SIFT to each individual image results in keypoints over the entire image as shown in Fig. 2.2. However we are only interested in points lying in the overlapping regions. Therefore points corresponding to two consecutive image pairs are subjected to feature matching, where Euclidean distance between SIFT feature vectors are computed and only the points having distance lesser than a predefined threshold are retained. The procedure of feature detection and matching is illustrated on an example image pair in Fig. 2.2 and Fig. 2.3. It can be observed from Fig. 2.3 that even after feature matching there might exist some points in the

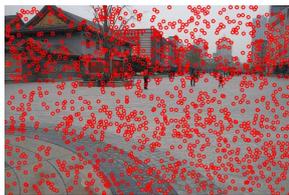
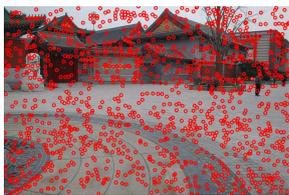


Figure 2.2: Keypoint Detection

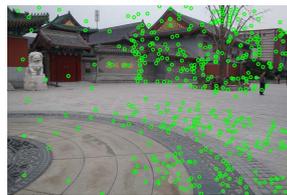


Figure 2.3: Keypoint Matching



Figure 2.4: Outlier removal

non-overlapping regions which are removed by an outlier rejection algorithms such as Random Sample Consensus (RANSAC) [30]. The effect of outlier rejection is shown Fig. 2.4, where after applying RANSAC the points only lie in the overlapping regions. The main contribution of feature detection stage is in identifying corresponding points in the individual images. In the case of inconsistencies associated in identifying the corresponding keypoints, it will show up in the form of distortions in the subsequent stages of stitching.

2.1.2 Image warping

Warping refers to the transformation applied on image co-ordinates in order to align the overlapping regions of the constituent images. Mathematically warping involves mapping co-ordinates by a suitable function and resampling. Various warping functions have been proposed in the literature for stitching, of which we only employ three of them in our database as we observed that they significantly influenced the quality of the final stitched image. The three warping functions are discussed below.

- *Homography* - Homography warp is one of the widely employed transformation for aligning images. Homography is calculated by an algorithm known as Direct Linear Transformation (DLT), by using the matched points obtained from the feature detection stage. The limitations associated with the above model is that it assumes that either the images

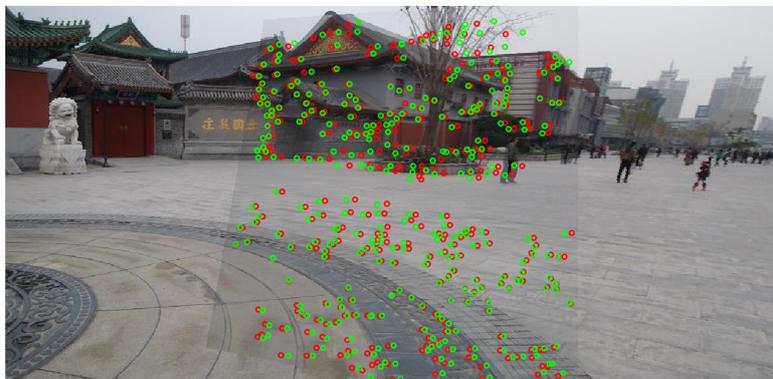


Figure 2.5: Illustration of stitching inconsistencies due to homography warp



Figure 2.6: Ghosting artifacts due to homography



Figure 2.7: Warping using local homography

taken by the rotating camera is parallax free/purely rotational (no translation should be present between two consecutive camera imaging positions) or imaging scene has to be planar [3]. Deviation from the above assumptions results in stitching inconsistencies and can manifest as ghosting artifacts. The above limitation is illustrated in Fig. 2.5 where the matching points are not perfectly overlapping due to violation of homography conditions, thereby resulting in ghosting artifacts. Another example of ghosting artifacts arising from applying homography is shown in Fig. 2.6.

- *Local Homography* - Moving DLT also known as local homography [5], was introduced to overcome the limitations of homography. In MDLT, homography transform is computed for every patch in the image instead of the entire image as is the case with homography. The effect of using MDLT is illustrated in Fig. 2.7 where ghosting artifacts which were



Figure 2.8: Distortions in non-overlapping regions due to local homography



Figure 2.9: Ghosting distortion in SPHP warp

observed in Fig. 2.6 have been eliminated. Drawbacks of local homography include the distortions such as blur, geometric etc. introduced in the non-overlapping parts of the stitched image as shown in Fig. 2.8.

- *Shape preserving warp* - In [6] a new warping method called shape preserving half projective (SPHP) warp is proposed by combining projective and similarity transformations. SPHP in general produces more natural looking panoramas when compared to homography and local homography, however some of the drawbacks associated with homography such as ghosting artifacts are still observable in SPHP as shown in Fig. 2.9.

From the above discussion on warping functions it is evident that the type and amount of distortions present in the final stitched image is dependent on the choice of the warping method

employed. Therefore the choice of warping method directly influences the quality of the stitched image.

2.1.3 Image Blending

Blending refers to the process of fusing multiple images to form a single composite image. The main aim of blending is to provide a smooth transition between images with no visible seams. Absence of suitable blending methods result in visible seams as observed in Fig. 2.10a where the warped images are shown without any blending. An exposure compensation stage precedes blending stage in the case of images having different exposure levels, and exposure correction algorithms are employed to compensate for these difference in exposure levels. We observed that the algorithms associated with exposure compensation stage [32, 14, 33] provide nearly similar results. Hence in our database we use only [33] for exposure correction. Multiple algorithms exist in the case of blending which are described below.

- *Linear Blending* - Linear blending or feathering employs a weighted averaging in the overlapping regions [3], where the weights are spatially varying in nature with pixels near the center of the image having higher weights than pixels near the edges. Feather blending applied to an exposure corrected image is shown in Fig. 2.10b.
- *Pyramid Blending* - Pyramid or multiband blending is carried out in band-pass domain by decomposing the image into a multiscale pyramid [34, 35]. The mask (valid pixel) image associated with each individual image is converted into a Gaussian (lowpass) pyramid. The masks are obtained by suitable seam cutting algorithms such as graphcut [31] etc. These blurred and subsampled masks become the weights used to perform a feathered blend of the band-pass images. Multiband blending is shown in Fig. 2.10c where the seams are visible (seams are relatively smooth due to blending) due to absence of exposure correction.

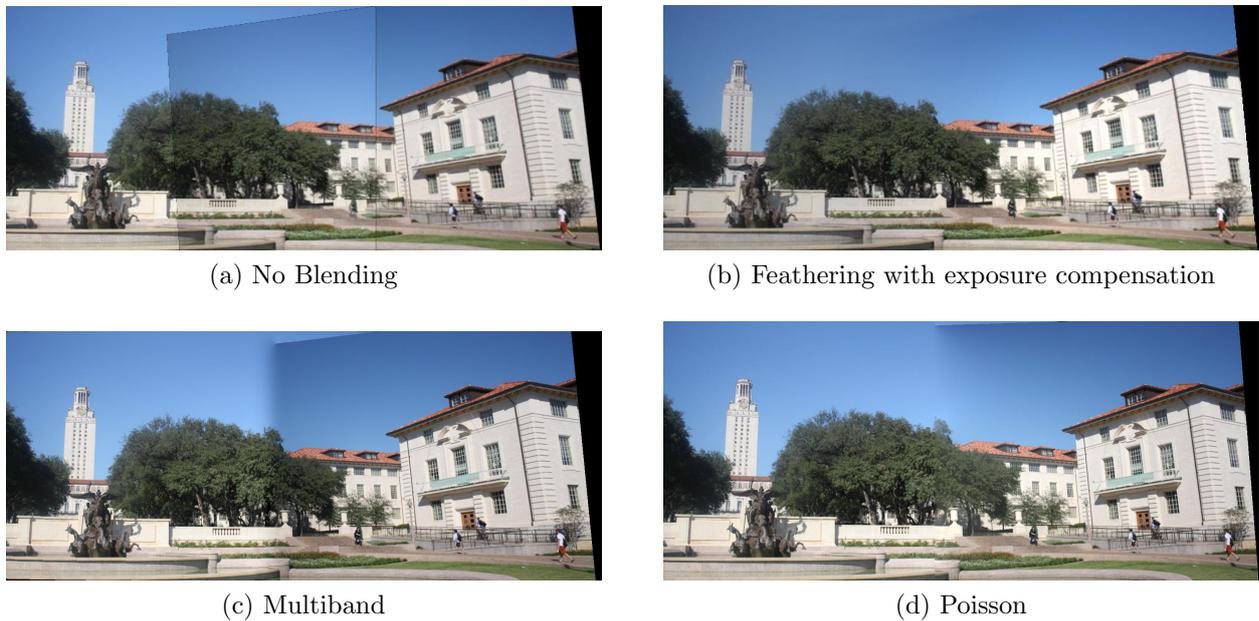


Figure 2.10: Illustration of various blending methods

- *Gradient domain blending* - An alternative to multiband image blending is to perform the operations in the gradient domain. In [36, 37] blending is done by solving an optimization function resulting in a Poisson equation. Poisson blending is shown in Fig. 2.10d where the seams (exposure correction is absent) are comparatively less visible when compared to multiband blending shown in Fig. 2.10c.

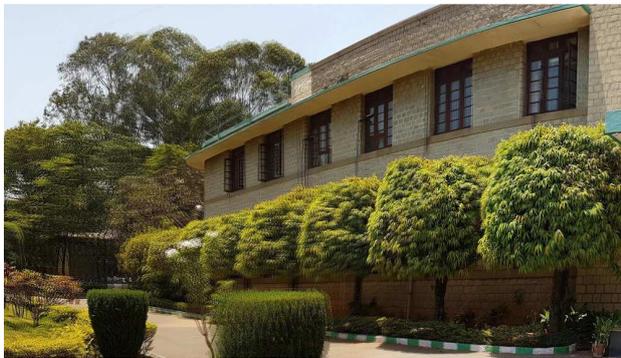
The above discussion on blending methods demonstrates the sensitivity of the quality of stitched image with respect to the choice of blending algorithms. Summarizing the above discussions, we showed that the choice of algorithms associated with warping and blending stage significantly influence the quality of stitched image. In particular, we only vary algorithms associated with the warping stage using [5] and [6], and the blending stage using [3, 34, 36, 35] and [37]. In addition to the above, a commercially available stitching algorithm [8] was also used to generate stitched images in the database. As a result of stitching we observe that ghosting, blur, color and geometric distortion are some of the prominently observed impairments in stitched images as described below.



(a) Ghosting



(b) Geometric Distortion



(c) Blur



(d) Color Distortion

Figure 2.11: Images illustrating various distortions due to stitching. The images shown above are cropped versions of the images present in the database.

- *Ghosting and blur* - occur due to inaccurate matching of feature points in the overlapping regions of constituent images resulting in stitched image with misalignments.
- *Color distortion* - occurs when constituent images have different exposure levels.
- *Geometric distortion* - occurs due to improper blending of multiple images (in the blending stage of the stitching process).

Fig. 2.11 illustrates the above types of distortions observed in stitched images.

2.2 Subjective Test Methodology

We employ the Single-Stimulus Continuous Quality Evaluation (SSCQE) [38] procedure to obtain human quality ratings for images in the ISIQA database. The images were viewed by the subjects on a Samsung Gear Virtual Reality HMD and rated on a scale of 0 – 100, 100 representing best quality and 0 corresponding to images suffering from severe distortions. In order to view the images on the HMD, the images were projected using equirectangular projection spanning the entire 360° with a 2 : 1 aspect ratio. Since the individual input images (or constituent images) in this dataset do not contain top and bottom views, black regions were augmented for areas corresponding to top, bottom and certain parts of horizontal view (as horizontal field of view is not completely 360°) such that the final aspect ratio was 2 : 1.

A total of 35 subjects were recruited for the study mostly from the student community of IISc. Each subject viewed a subset of 192 out of the 264 images from the database so that each image received a minimum of 25 user ratings. The 192 images were divided into a three viewing sessions consisting of 64 images each and the subjects took no more than 30 minutes to complete each viewing session. Images were shown in a randomized order with no two successive images belonging to the same scene.

Each subject was individually explained the goal of the experiment and a short training session was conducted to familiarize the rating procedure. The subjects were instructed to provide ratings based on the quality perceived rather than the aesthetics of the image. During the training session 5 images were shown, the first being a calibration image used to adjust the focus of the lens present in the HMD and the rest four images spanning the same range of quality as the test images, to provide the subject an idea of the quality of images present in the study. The training images are not part of the ISIQA database and their scores were not considered for evaluation. The training set of images were the same for all participants. The subjects were provided with a swivel chair to facilitate 360° viewing capability. Also there was no time constraint on viewing each image, the subjects were allowed to view them as long as

they wished. The subjects were allowed to pause in between their session if they experienced uneasiness or felt fatigued. To reduce the subjective fatigue, a minimum gap of 24 hours was provided between successive sessions. The subjects orally reported the ratings after viewing every image which were recorded by an individual conducting the study.

2.3 Processing of Subjective Scores

Let m_{ijk} denote the score provided by subject i to image j in session $k = \{1, 2, 3\}$. Since all images in the ISIQA database are not rated by every subject, let $\delta(i, j)$ be an indicator function such that

$$\delta(i, j) = \begin{cases} 1 & \text{if subject } i \text{ rated image } j \\ 0 & \text{otherwise.} \end{cases} \quad (2.1)$$

To normalize the scores received across multiple sessions of each subject, we calculate the Z-scores per session [39] as

$$\mu_{ik} = \frac{1}{N_{ik}} \sum_{j=1}^{N_{ik}} m_{ijk} \quad (2.2)$$

$$\sigma_{ik} = \sqrt{\frac{1}{N_{ik} - 1} \sum_{j=1}^{N_{ik}} (m_{ijk} - \mu_{ik})^2} \quad (2.3)$$

$$z_{ijk} = \frac{m_{ijk} - \mu_{ik}}{\sigma_{ik}}, \quad (2.4)$$

where N_{ik} is the number of images seen by subject i in session k . The Z-scores from all sessions are concatenated to form the matrix $\{z_{ij}\}$ denoting the Z-score assigned by subject i to image j with $j = \{1, 2, 3 \dots 264\}$ with the entries of $\{z_{ij}\}$ being empty at locations (i, j) where $\delta(i, j) = 0$. These Z-scores are then used in the subject rejection procedure detailed in the ITU-R BT 500.11 recommendation [38] to discard scores from unreliable subjects. In our

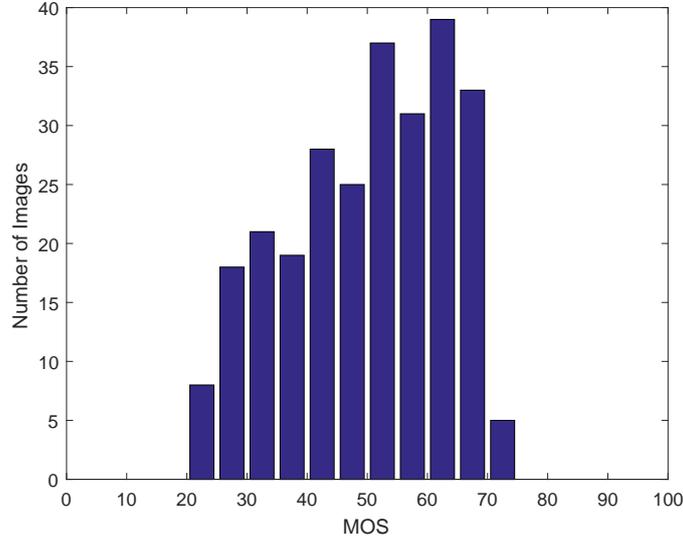


Figure 2.12: Histogram of MOS

study two subjects were rejected. Assuming z_{ij} to have a standard normal distribution, 99% of the Z-scores lie in $[-3, 3]$. A linear rescaling was used to map scores to the range $[0, 100]$ as

$$z_{ij} = \frac{100(z_{ij} + 3)}{6}. \quad (2.5)$$

Finally the Mean Opinion Score (MOS) of each image was calculated by averaging the ratings received for that image as

$$MOS_j = \frac{1}{N_j} \sum_{i=1}^N z_{ij} \delta(i, j), \quad (2.6)$$

where $N_j = \sum_{i=1}^N \delta(i, j)$ and $N = 264$. The MOS were found to lie in the range $[20.6, 72.5]$ and the mean of standard deviations of the rescaled Z-scores obtained from all subjects across all images is 9.7. The histogram of MOS scores is shown in Fig. 2.12 indicating a reasonably wide distribution of MOS values.

2.4 Anchor Images

Note that in our study, every image was not rated by all subjects and a unique subset of images were viewed by each subject. While this procedure of only having a unique subset of images viewed by each subject has been used in various studies [40, 41, 42], we did not find a discussion in literature on the impact on MOS of having a different subset of subjects view each image as opposed to the entire population. In order to analyze this, we choose a subset of 12 images from the ISIQA database referred to as *anchor images*. These anchor images were present in the viewing set for all subjects (in 192 images that were scored by the subjects, 12 images were viewed by all subjects and received 35 ratings while the other images received 25 ratings). To analyze the impact of the number of subjects on MOS, we randomly sampled subsets of the subject scores and recalculated MOS on these reduced subsets for these images as shown in Fig. 2.13. We see that these computed MOS values remain relatively constant across the number of subjects although the standard deviation tends to increase for number of subjects lesser than 22. The confidence intervals were calculated based on the variation of MOS over

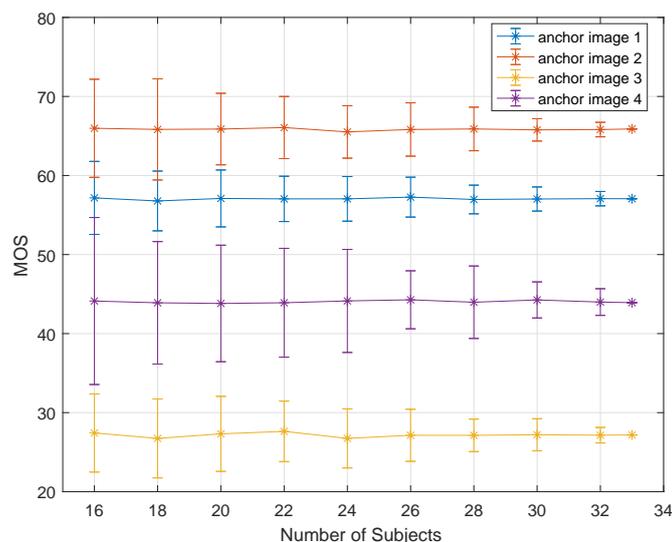


Figure 2.13: MOS of anchor images plotted with number of subjects along with 95% confidence intervals

25 trials. While Fig. 2.13 shows the results of 4 anchor images, similar results are observed for other anchor images as well.

2.5 Analysis of Color Distorted Images

As explained earlier, stitched images suffer from a variety of distortions. We now analyze the sensitivity of humans to color distortions in particular. The analysis of images suffering from color distortions reveals an interesting observation. In a majority of the scenes, the scores received for color distorted images are relatively close to the score received for images suffering from little or no distortion. This observation is shown in Fig. 2.14 where two color distorted images along with their MOS are compared with images having no artifacts. The relative closeness of the MOS potentially indicates that color distortion appears to be less annoying, particularly when viewed on a HMD. We believe that since the images were rated by viewing them on a HMD which provides approximately 90° field of view, there are instances where color distortion might not appear in the purported field of view resulting in higher ratings even in the presence of such impairments.



(a) $MOS = 61.7624$



(b) $MOS = 61.6771$



(c) $MOS = 59.7492$



(d) $MOS = 59.954$

Figure 2.14: Comparison of MOS of color distorted images (left column) with images having no artifacts (right column)

2.6 Analysis of Subjective Response Times

While conducting the subjective study, in addition to the subjective ratings, we also collected the time taken by each subject to rate every image. In Fig. 2.15 the average response time for various distortions is shown. It can be seen that images with ghosting artifacts require least amount of time for rating as they are easily recognized, while geometry and color distorted images require relatively longer time as they require careful observation for their identification. As expected, the longest response time corresponded to images with no distortions as they required more scrutiny to make sure the absence of any impairments. Note that the categorization of images into different distortion types has been done approximately by the authors. Often, these distortions occur concurrently, in which case the most dominant distortion type has been assigned for the purposes of this experiment.

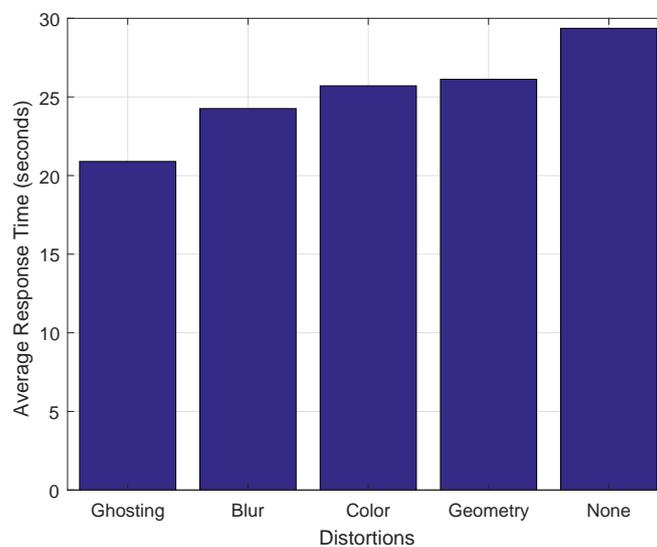


Figure 2.15: Average response time for various distortions

Chapter 3

Stitched Image Quality Evaluator

We now present an objective algorithm to automatically evaluate the quality of stitched images. The framework of our proposed method is shown in Fig. 3.1. We devise two sets of features to characterize the distortions and assess the quality of the stitched image. These features are designed to capture deviations in the statistics of images due to the presence of artifacts such as ghosting and blur. The first set of features captures structural changes (addition or deletion of edges and blur) occurring due to distortions. The second set of features extracts variations observed in spatial correlation due to the presence of ghosting in the stitched image. Feature extraction is carried out at a patch level, and a weighted average of the patch level features is computed to obtain image level features for both the stitched as well as constituent images. Feature differences are then fed to a support vector regressor for quality prediction. We describe further details of the framework in the following sections.

3.1 Bandpass Analysis of Image Stitching

As discussed earlier, ghosting is one of the prominent distortions that occur due to image stitching. Ghosting occurs when overlapping regions in the constituent images are not aligned perfectly. Fig. 3.2 illustrates the warping stage of the stitching process using homography

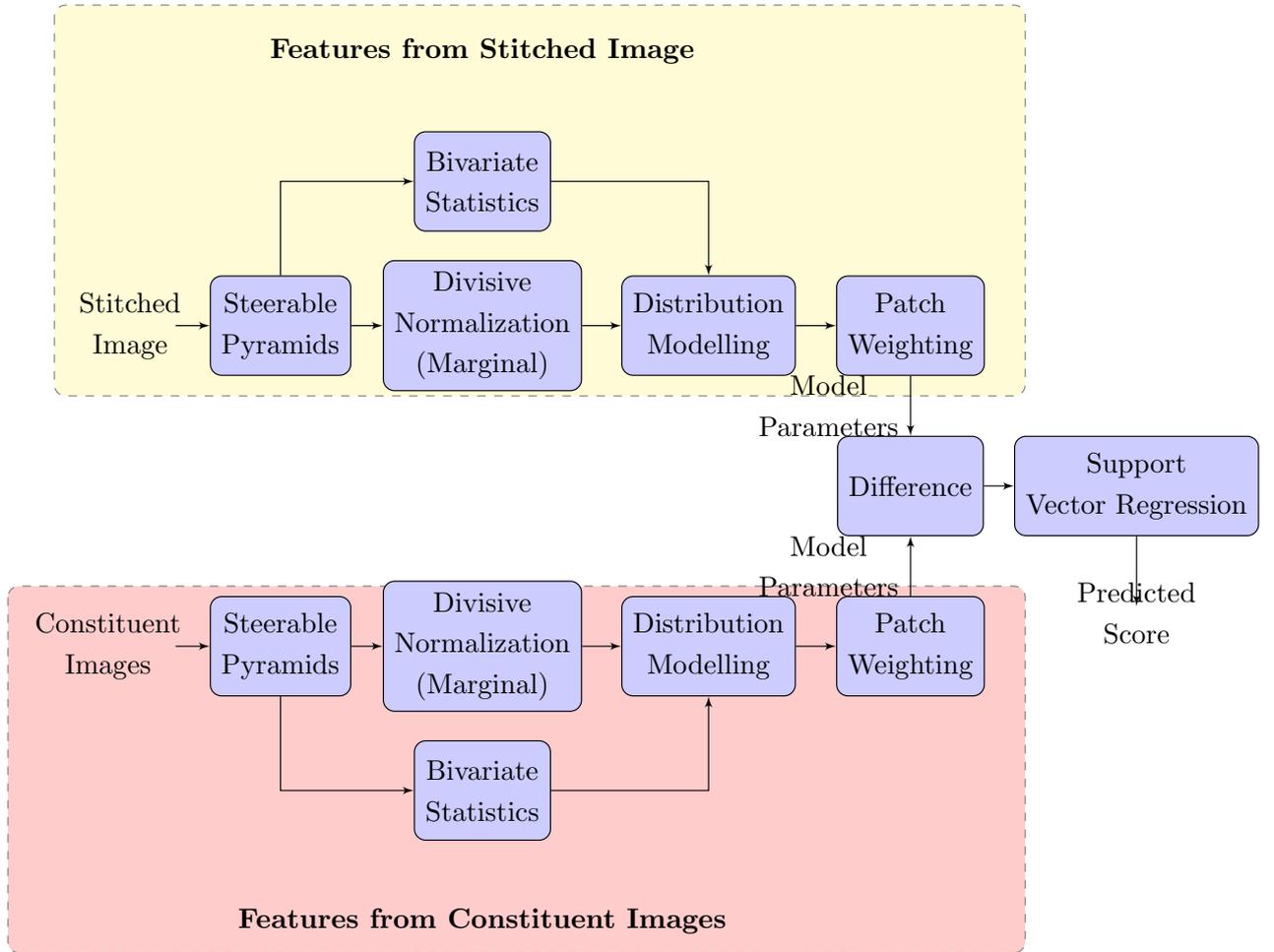


Figure 3.1: High Level Overview of SIQE Framework

transform. Let X denote the set of all coordinates representing the overlapping region in I_1 . Ghosting is observed when $I_1(x) \neq I_2(Hx)$ for any $x \in X$, where H denotes homography transformation between I_1 and I_2 . Mathematically ghosting can be represented as

$$I(x) = (1 - \alpha(x))I_1(x) + \alpha(x)I_2(Hx), \quad (3.1)$$

where $\alpha(x) \in (0, 1)$ denotes the weighting value at location x . Ghosting artifacts are characterized by the presence of additional edges as well as the increased spatial correlation due to the replication of certain regions by means of shifting and adding pixels from the original image (the

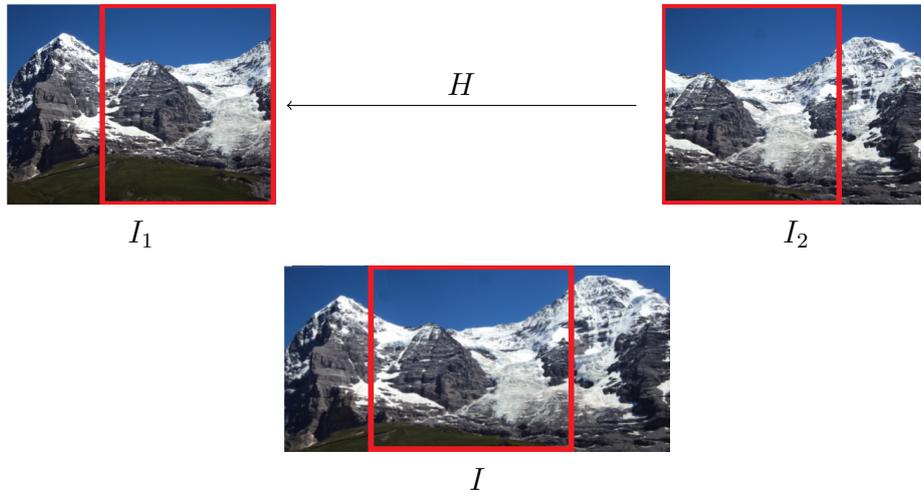


Figure 3.2: Illustration of homography warp. Red boxes indicate overlapping regions

shift can be non-linear with each pixel having different shifts) as shown in Equation 3.1. Thus features that capture ghosting need to detect the presence of additional edges as well as the amount of increased correlation between neighboring pixels for efficient quality prediction. In addition to ghosting, geometric distortions such as shape deformations also manifest themselves in the form of structural distortions in the stitched image. The quality features we design need to be sensitive to such distortions.

We notice that structural artifacts due to ghosting and geometric distortions primarily occur at specific orientations and these are not isotropic. Thus we subject the images to multi-scale, multi-orientation decompositions using steerable pyramids [43] to capture such distortions. Steerable pyramids have been used in multiple image QA algorithms effectively [44, 11, 45]. The significance of orientation selectivity is illustrated in Fig. 3.3, where the presence of additional edges due to ghosting in Fig. 3.3a is prominently visible in the orientation decomposition along 150° in Fig. 3.3c and not so much in the decomposition along 60° in Fig. 3.3b. Similarly, the geometric errors in Fig. 3.3d are more clearly visible in Fig. 3.3e when compared to Fig. 3.3f.

Given an image whose quality is to be assessed, we partition the image into $N \times N$ patches and subject each patch to a steerable pyramid decomposition across 2 scales and 6 orientations.

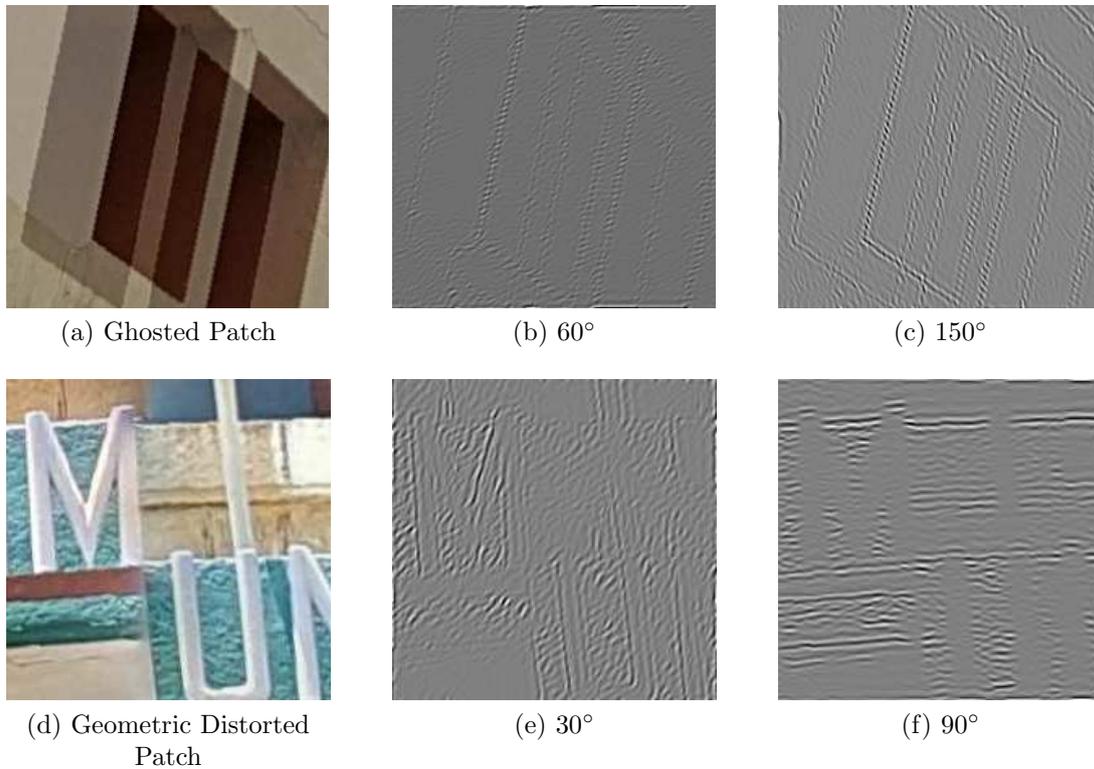


Figure 3.3: Illustration of orientation selectivity

The resulting decomposition results in 12 subbands denoted by s_{α}^{θ} , where $\alpha \in \{1, 2\}$ represents the scale index and $\theta \in \{0^{\circ}, 30^{\circ}, 60^{\circ}, 90^{\circ}, 120^{\circ}, 150^{\circ}\}$ represents the orientation. We now extract features in each of the resulting subbands as described in the subsections below.

3.2 Marginal Statistics Model with Divisive Normalization

Ghosting artifacts as well as geometric distortion are associated with the presence of structural deformations that can manifest as addition/deletion of edges when compared to the original undistorted patch. We propose to employ a marginal statistics model of divisively normalized subband coefficients to capture the quantum of additional edges observed due to distortions.

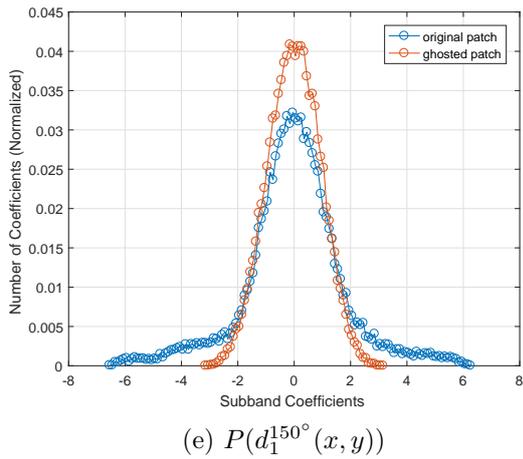
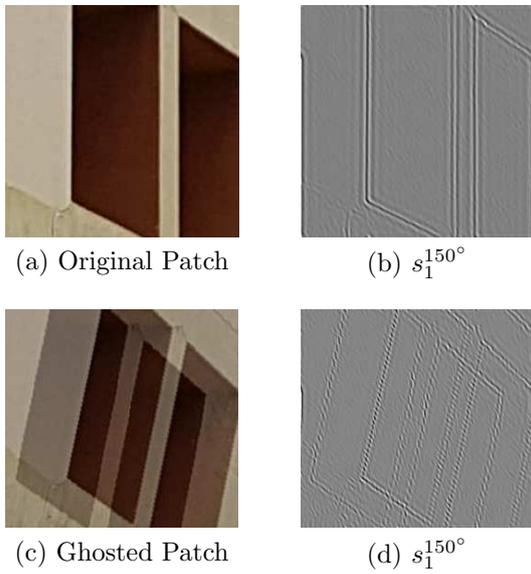


Figure 3.4: Comparison of marginal statistics with divisive normalization in presence of ghosting

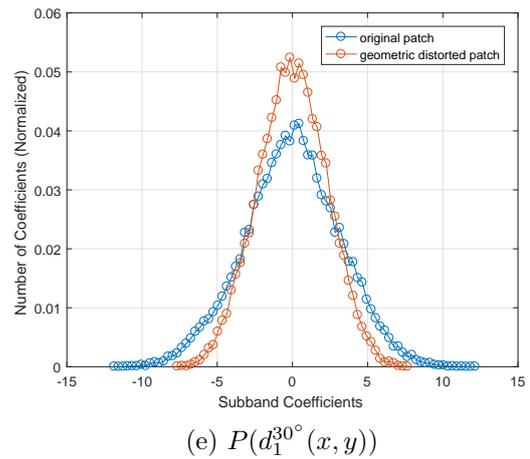
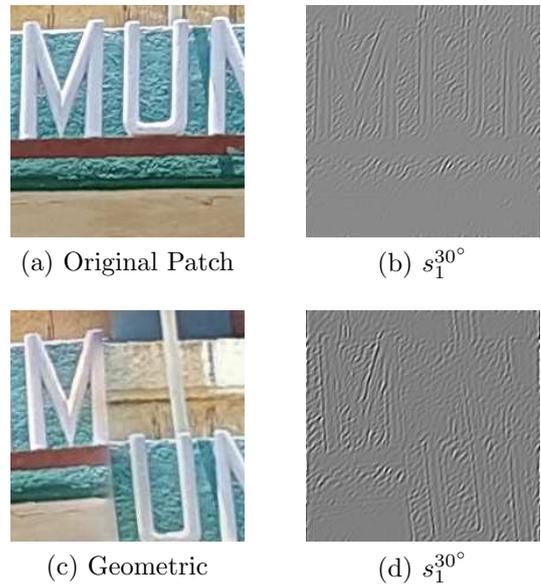


Figure 3.5: Comparison of marginal statistics in presence of geometric distortion

Divisive normalization performs the function of contrast gain control and also reduces the statistical dependencies between neighboring coefficients. In the past [44, 11, 45] have successfully used divisive normalization for quality prediction in the presence of various distortions such as blur, noise and compression artifacts.

Given a subband coefficient y , a $P \times P$ neighborhood with y at the center is constructed to obtain vector Y . We implement divisive normalization as discussed in [11], where divisive

normalization is defined as $\hat{y} = y/p$. Here \hat{y} represents the normalized coefficient with p being the normalizing factor, calculated as $p = \sqrt{Y^T C_U^{-1} Y / N}$ [46], where C_U is the covariance of the neighborhood for the given subband and N representing the number of neighbors. The value of p is obtained at every location in each subband to obtain a set of divisively normalized coefficients denoted by d_α^θ , where $\alpha \in \{1, 2\}$ and $\theta \in \{0^\circ, 30^\circ, 60^\circ, 90^\circ, 120^\circ, 150^\circ\}$.

Divisive normalization helps in capturing orientation distortions where statistics of unnormalized coefficients fail. It plays a key role in determining the extent of additional edges present due to distortions. Intuitively p can be considered as a factor dependent on local variance, therefore the values of p in regions containing edges tend to be higher than the regions with no edges. The presence of additional edges due to distortions results in more number of locations having higher p values. This leads to a large number of divisive normalized coefficients having smaller values as $\hat{y} = y/p$, and a distribution with higher peak value at zero than that of the reference. The effect of divisive normalization on the distribution of subbands is illustrated in Fig. 3.4e where subbands of original and distorted patches along $\theta = 150^\circ$ are compared. The new edges due to ghosting that are prominently visible in $s_1^{150^\circ}$ (Fig. 3.4d), lead to a change in the shape of distributions $P(d_1^{150^\circ}(x, y))$ for the reference and distorted as shown in Fig. 3.4e. We also observe deviations in the statistics of divisively normalized coefficients in images with geometric distortions due to the presence of additional edges as illustrated in Fig. 3.5 where the distribution $P(d_1^{30^\circ}(x, y))$ of the original and distorted are compared (Fig. 3.5e). Divisive normalization also captures blur as previously shown by various QA indices [11, 45].

The normalized coefficients d_α^θ are statistically modeled using a generalized Gaussian distribution (GGD), where the pdf of a GGD with zero mean is given by:

$$f(x; \gamma, \sigma^2) = \frac{\gamma}{2\beta\Gamma(1/\gamma)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\gamma\right) \quad (3.2)$$

$$\beta = \sigma \sqrt{\frac{\Gamma(1/\gamma)}{\Gamma(3/\gamma)}}, \quad (3.3)$$

and $\Gamma(\cdot)$ is the gamma function

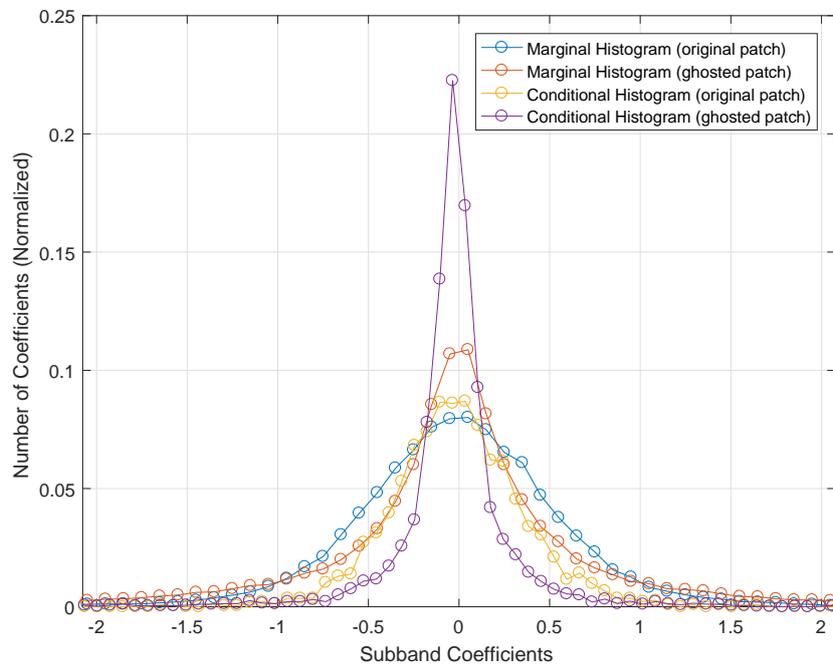
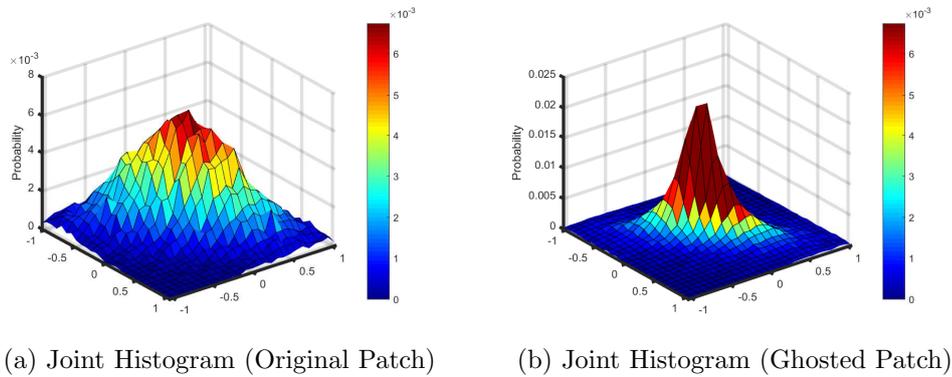
$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt. \quad (3.4)$$

The parameter γ controls the shape of the distribution and σ^2 controls the variance. The parameters of $\text{GGD}(\gamma, \sigma^2)$ are estimated using the moment matching technique proposed in [47]. The shape parameters¹ of the fitted model represent the features for a given patch. The above procedure is repeated for all patches and features are weighted averaged (patch weighting is discussed in Sec. 3.4) resulting in a 12 dimensional (6 orientations \times 2 scales) feature vector which we denote as f_{1-12}^s, f_{1-12}^c for the stitched and constituent images respectively.

3.3 Bivariate Model

Ghosting artifacts are also associated with an increase in spatial correlation (or decrease in variance in bandpass domain, a proof for decrease in variance in case of ghosting under some assumptions is provided in Appendix A) which we capture by studying the bivariate statistics of neighboring bandpass coefficients. Note that divisive normalization reduces statistical dependencies between adjacent coefficients [46, 48], while we wish to capture the correlation introduced in the adjacent coefficients due to ghosting. As a result, we focus on modeling the bivariate statistics of the unnormalized coefficients. In particular, we obtain the bivariate statistics by considering immediate neighboring subband coefficients of s_{α}^{θ} denoted by $P(s_{\alpha}^{\theta}(x, y), s_{\alpha}^{\theta}(x+1, y))$ and $P(s_{\alpha}^{\theta}(x, y), s_{\alpha}^{\theta}(x, y+1))$ for horizontal/vertical neighbors respectively. Before we present our model for the bivariate statistics, we try to understand its relevance in capturing ghosting artifacts. In Fig. 3.6a and 3.6b, we note the remarkable difference in the bivariate statistics of original and ghosted patches. To further understand this, we proceed by

¹The other parameters of GGD did not contribute for higher quality prediction. Hence only shape parameters were retained.



(c) Conditional Histogram

Figure 3.6: Joint and conditional statistics in presence of ghosting. The above histograms are plotted for $s_1^{60^\circ}$ for patches shown in Fig. 3.4

analyzing the conditional statistics $P(s_\alpha^\theta(x+1, y)/s_\alpha^\theta(x, y) \in (-\delta, \delta))$ (δ is a small positive constant close to 0, in our experiments we use $\delta = 0.0125$) as shown in Fig. 3.6c. The conditional histograms in Fig. 3.6c reveal that in the case of ghosting, when the neighboring coefficient is close to zero (i.e. $s_\alpha^\theta(x, y) \in (-\delta, \delta)$), the probability of the coefficient under consideration being close to zero is significantly high when compared to that of the original patch. Note the

large difference in peak values illustrated in Fig. 3.6c between the conditional distributions of the original and distorted. On the other hand, the marginal distributions of the original and ghosted patches are relatively closer to each other indicating the stronger nature of bivariate statistics in capturing increased correlation. Thus modeling the bivariate statistics of neighboring coefficient can help capture the correlation introduced due to ghosting.

Bivariate statistical models have recently gained attention in the modeling of natural scene statistics [49], where divisively normalized coefficients have been modeled using a bivariate Generalized Gaussian distribution (BGGD). Further, these have been used in blind QA in [50]. Another competing model to fit bivariate distributions is the Gaussian Scale Mixture (GSM) model [46]. A random vector Y is a GSM if $Y \equiv zU$ where \equiv denotes equality in distribution, U is a zero-mean Gaussian random vector, z is a scalar random variable; and U and z are independent. While a few models for the scalar random variable have been explored for natural scenes [46], we observe that employing simple priors on z such as *Jeffrey's prior* [51] does not fit the empirical distribution of original and distorted patches well.

We model the bivariate joint distribution using a bivariate Gaussian Mixture Model (GMM). While univariate GMMs have been used to model wavelet coefficients in denoising applications [52, 53], their application in QA has been limited. Let $s_\alpha^\theta(x, y) = a$, $s_\alpha^\theta(x + 1, y) = b$, then the probability density function (pdf) for a GMM $f(a, b)$ is given by

$$f(a, b) = \sum_{i=1}^M \omega_i \mathcal{N}(\mathbf{0}, \Sigma_i), \quad (3.5)$$

where ω_i is the weighting parameter with $\sum_{i=1}^M \omega_i = 1$ and $\mathcal{N}(\mathbf{0}, \Sigma_i)$ is a zero mean bivariate normal distribution with covariance Σ_i . We use zero mean normal distributions in the mixture since subband coefficients s_α^θ are zero mean. The parameters ω_i and Σ_i are calculated through maximum likelihood (ML) estimation via Expectation Maximization (EM) algorithm [54]. The

Feature ID	Feature Description	Computation Procedure
f_{1-12}^s, f_{1-12}^c	Shape parameters of subband coefficients	Fitting a GGD to subband coefficients
f_{13-24}^s, f_{13-24}^c	Eigen values of bivariate distribution obtained from horizontal neighbors	Fitting a bivariate GMM to horizontally adjacent subband coefficients
f_{25-36}^s, f_{25-36}^c	Eigen values of bivariate distribution obtained from vertical neighbors	Fitting a bivariate GMM to vertically adjacent subband coefficients

Table 3.1: Summary of extracted features

covariance C for the pair of wavelet coefficients is given by

$$C = \sum_{i=1}^M \omega_i \Sigma_i. \quad (3.6)$$

C is computed for each patch and the eigen values of C are obtained as features. For each subband, a bivariate GMM with $M = 4$ is employed for horizontal as well as vertical neighbors. Since C has dimensions 2×2 , 2 eigen values are obtained for each subband. Weighted averaging is carried across all patches (patch weighting is discussed in Sec. 3.4) to obtain a 24 dimensional (6 orientations \times 2 eigen values \times 2 horizontal and vertical neighbors²) feature vector which we denote as f_{13-36}^p, f_{13-36}^c for the stitched and constituent images respectively. The extracted features are summarized in Table 3.1.

3.4 Patch Weighting

We observe that ghosting, blur and geometric artifacts are not perceived when patches are smooth (*e.g.* sky) and majorly occur in textured regions. Therefore while combining patch level feature vectors to obtain feature vectors at an image level, we weight the textured patches more than smooth regions. We employ a simple texture estimation algorithm based on the Gray Level Co-occurrence Matrix (GLCM) [55]. We use the energy measure of GLCM to quantify the amount of texture present. Energy values of GLCM lie in the range $[0, 1]$, with constant

²We only employ features from first scale as we observed that including features from second scale did not result in higher quality prediction.

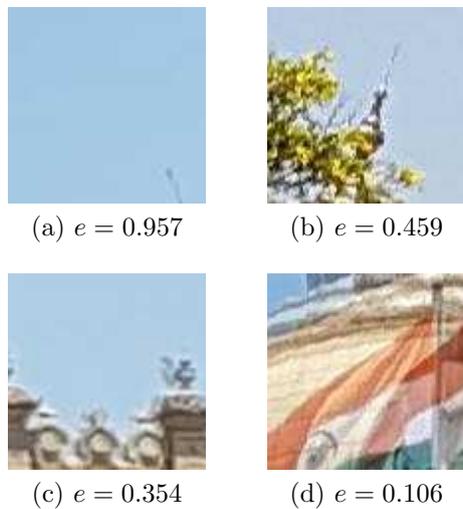


Figure 3.7: Illustration of GLCM energy values for patches with varying amounts of texture

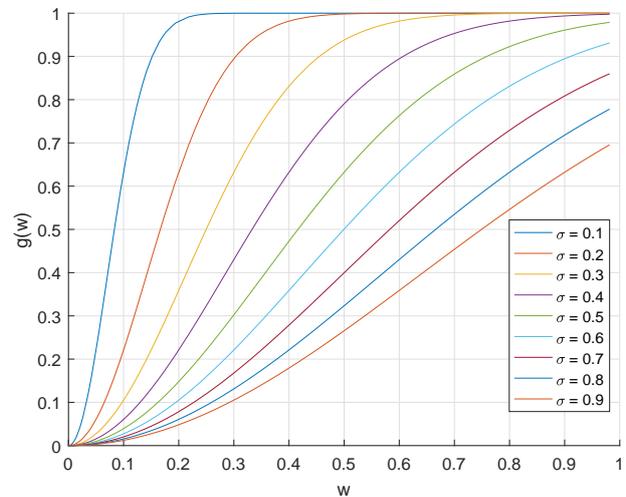


Figure 3.8: Effect of σ on function g

images having energy value equal to 1. This is illustrated in Fig. 3.7 where energy values are represented by e . Smooth patch shown in Fig. 3.7a has energy value closer to 1 while textured patch in Fig. 3.7d has value closer to 0. Since we require larger weights for textured patches, we compute the weights as $w = 1 - e$. Further, we apply a non-linearity on the weight w given by

$$g(w) = 1 - \exp\left(-\left(\frac{w}{\sigma}\right)^2\right), \quad (3.7)$$

where σ parameter controls the point of saturation for the non-linear function g . The above non-linearity provides a saturation behavior where all patches above a particular amount of texture achieve roughly equal weights. The effect of parameter σ is shown in Fig. 3.8 where values of function g is plotted for different σ values. σ values very close to zero result in equal weights for almost all patches (including smooth patches) while sigma values away from zero approach linear variation, indicating little or no effect of the non-linearity. In our model we use $\sigma = 0.1$.

3.5 Prediction

Given a database of distorted stitched images and corresponding human opinion scores, we obtain the difference of features extracted from the stitched image and those from the corresponding to constituent images, $f_{1-36} = f_{1-36}^c - f_{1-36}^s$, and a mapping is learned from feature space to quality scores by means of a regression model. Our proposed framework is generic enough to be employed with any regressor. In our implementation we use a Support Vector Regressor (SVR) [56] as the SVR has been successfully used in various quality assessment problems [11, 12, 57, 58, 59]. We use the LIBSVM implementation [60] of a radial basis function (RBF) kernel SVR to train and test our models.

Chapter 4

Experiments and Results

In this chapter we report the results of various experiments we perform in order to evaluate the performance of SIQE. Firstly, we compare the quality of fits of BGGD and GMM in modeling bivariate distributions. Secondly, we assess the performance of SIQE by examining its correlation with human judgments and comparing with existing algorithms. We next analyze the significance of each conceptual feature present in SIQE framework. We also report the time complexity associated with SIQE.

4.1 Comparison of Models for Bivariate Distribution

We now evaluate the quality of fit of our bivariate GMM for adjacent subband coefficients on both original and stitched image patches of size 100×100 . We compare our model with the recently developed BGGD model [49, 61] for bivariate statistics of subband decompositions. In Table 4.1, we evaluate the quality of GMM and BGGD fits with respect to the empirical bivariate distribution using the L_1 -norm error and Kullback Liebler (KL) divergence for patches from the original constituent images. We employ the procedure detailed in [61] for estimating BGGD parameters. The L_1 -norm error and KL divergence are calculated between empirical and fitted histograms for every patch and the mean over all patches is reported. The values reported

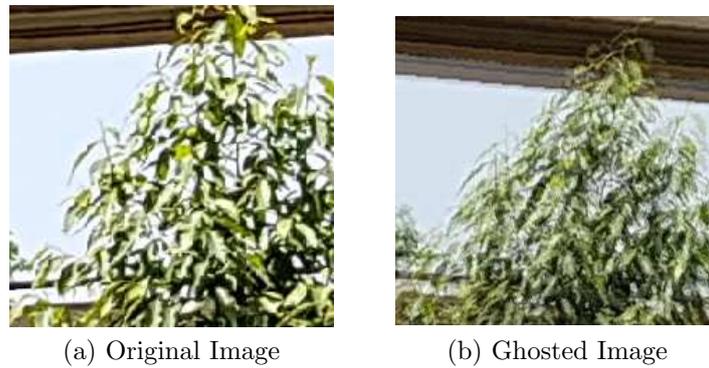


Figure 4.1: Ghosting example

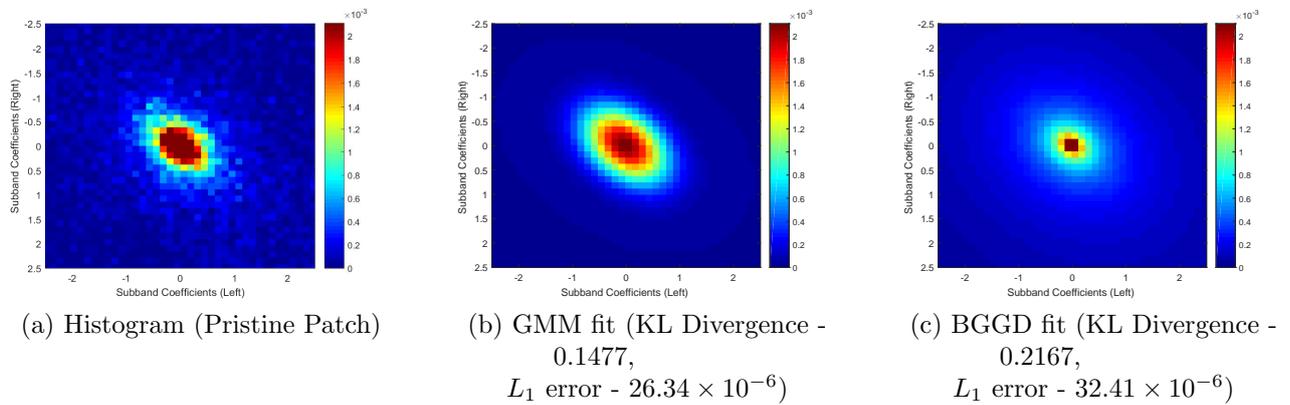


Figure 4.2: 2D illustration of probability distributions for GMM and BGGD fits of undistorted patch. The above distributions represent subband $s_1^{0^\circ}$ for patch shown in Fig. 4.1a

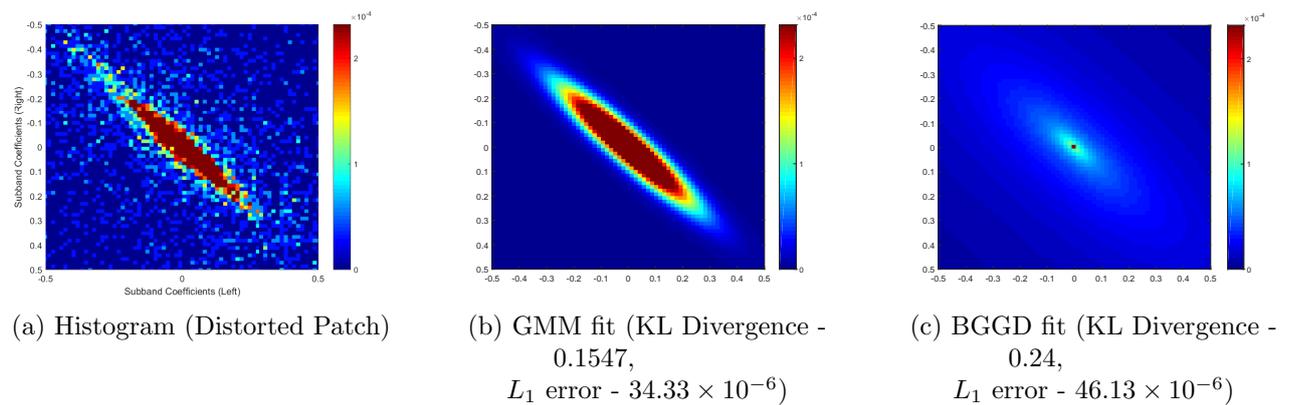


Figure 4.3: 2D illustration of probability density functions for GMM and BGGD fits of ghosted patch. The above distributions represent subband $s_1^{120^\circ}$ for ghosted patch shown in Fig. 4.1b

Subbands	L_1 norm error (10^{-6})		KL Divergence	
	GMM	BGGD	GMM	BGGD
$s_1^{0^\circ}$	33.3291	34.8282	0.1715	0.2034
$s_1^{30^\circ}$	33.4749	34.9788	0.1581	0.1842
$s_1^{60^\circ}$	29.0116	30.5941	0.1856	0.2026
$s_1^{90^\circ}$	24.6289	27.0378	0.2069	0.2205
$s_1^{120^\circ}$	29.1184	30.6145	0.1858	0.2025
$s_1^{150^\circ}$	33.4337	34.7679	0.1582	0.1850

Table 4.1: Comparison of GMM and BGGD fits for pristine patches for horizontal neighbors

Subbands	L_1 norm error (10^{-6})		KL Divergence	
	GMM	BGGD	GMM	BGGD
$s_1^{0^\circ}$	38.9041	40.6054	0.2305	0.2413
$s_1^{30^\circ}$	39.3184	41.4931	0.2004	0.2161
$s_1^{60^\circ}$	33.6191	37.6948	0.2132	0.2263
$s_1^{90^\circ}$	27.7381	32.4613	0.2180	0.2455
$s_1^{120^\circ}$	33.8411	35.7571	0.2107	0.2242
$s_1^{150^\circ}$	39.4030	42.0467	0.1991	0.2150

Table 4.2: Comparison of GMM and BGGD fits for distorted patches for horizontal neighbors

in Table 4.1 correspond to distributions of horizontal neighbors and represent the mean of all the patches obtained from the constituent images of 26 scenes present in the ISIQA database. It can be inferred from Table 4.1 that GMM fits across all subbands are superior to corresponding BGGD fits. The same is observed in Fig. 4.2 where we see that the visual illustration using contour plots of the GMM fits are closer to the empirical statistics when compared to the BGGD fits. Similar observations hold true for distorted patches in the stitched images, as shown in Table 4.2 and Fig. 4.3. The values reported in Table 4.2 are obtained from patches of 26 images suffering from ghosting artifacts present in the ISIQA database and correspond to distributions of horizontal neighbors. Similar observations were made when distributions involving vertical neighbors were considered for both pristine as well as distorted images. Due to the superior modeling capability of GMM for both undistorted as well as ghosted images, we use GMM in our SIQE framework. We also observe that the time taken to estimate BGGD parameters is significantly high when compared to that of GMM, prohibiting its usage in our framework.

4.2 Correlation With Human Judgments

Before analyzing the performance of SIQE, we provide further details of its implementation. The patch size is fixed to 100×100 . For patch weighting, a 13×13 GLCM is constructed for every patch of quantized image as explained in Section 3.4 [55]. This choice of quantization levels

	SROCC	LCC
BRISQUE (trained on ISIQa)	0.6225	0.5954
NIQE	0.1524	0.1051
DIIVINE (trained on ISIQa)	0.5683	0.5880
SIQE	0.8318	0.8395

Table 4.3: Median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets (subjective MOS vs predicted MOS)

was determined by experimentation on the training set. We also note that the performance is not very sensitive to the exact choice of the number of levels.

In order to train SIQE, we divide the ISIQa database into two randomly chosen subsets - 80% training and 20% testing - ensuring that there exists no overlap of scenes between train and test content. This is carried out to ensure that the reported results are not dependent on features extracted from known scene content. Specifically, each training set consists of images from 21 scenes while the testing set contains the remaining 5 scenes. Further we repeat this random train-test split over 1000 iterations and report the median performance. The Spearman’s rank order correlation coefficient (SROCC) and Pearson’s linear correlation coefficient (LCC) between between the predicted quality scores from SIQE and the subjective MOS are used to assess the performance of SIQE. Before computing LCC, the scores predicted from the objective algorithm were passed through a five parameter logistic non-linearity described in [40] as

$$\text{Quality}(x) = \beta_1 \text{logistic}(\beta_2, (x - \beta_3)) + \beta_4 x + \beta_5 \quad (4.1)$$

$$\text{logistic}(\tau, x) = \frac{1}{2} - \frac{1}{1 + \exp(\tau, x)}. \quad (4.2)$$

We compare the performance of SIQE against existing NR IQA indices such as DIIVINE [11], BRISQUE [12] and NIQE [13] in Table 4.3. We report performances of BRISQUE and DIIVINE features when trained on the ISIQa database. We see from Table 4.3 that SIQE significantly outperforms the NR-IQA methods owing to the design of features that capture distortions

	SROCC STD	LCC STD
BRISQUE (trained on ISIQ)	0.0828	0.0747
NIQE	0.1887	0.1753
DIIVINE (trained on ISIQ)	0.1398	0.1314
SIQE	0.0591	0.0519

Table 4.4: Standard deviation of SROCC and LCC values for 1000 iterations of randomly chosen train and test sets

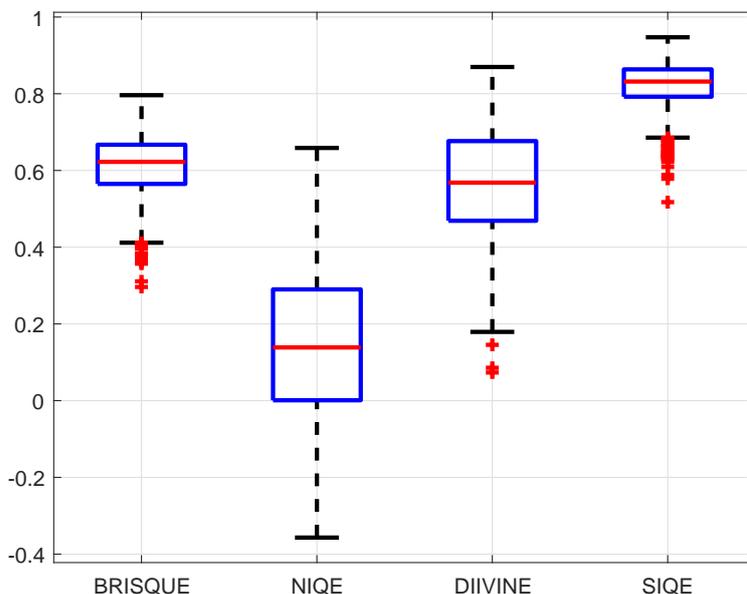


Figure 4.4: Comparison of median SROCC distribution across multiple QA algorithms

arising from stitching and also the use of reference information in the form of constituent images. Fig. 4.4 shows spreads of SROCC values for each algorithm over 1000 iterations. The plot indicates that SIQE has a much better confidence interval when compared to other methods underlining the robustness of the algorithm. The smaller standard deviation values of correlation coefficients for SIQE in Table 4.4 also imply ruggedness of the features used in SIQE. Note that since multiple distortions due to stitching occur concurrently, we only present results on the entire dataset and a fine grained analysis on individual distortion types appears difficult to perform.

We also compare with the performance of FR IQA indices proposed in [14] and [15] in

	SROCC	LCC
Xu [14] (PSNR)	0.1795	0.2341
Xu [14] (SSIM)	0.3383	0.4077
Qureshi [15]	0.3238	0.3627
SIQE	0.7848	0.8032

Table 4.5: Comparison of Median values of SROCC and LCC with full reference QA algorithms on a subset of 238 images in the ISIQA database

Table 4.5. Xu *et al.* [14] use PSNR for color similarity and SSIM for structural similarity, whereas Qureshi *et al.* [15] use SSIM along high pass components for quantifying the quality of stitched images. Note that FR IQA indices are only employed in the overlapping regions and they require knowledge of stitching algorithm for establishing pointwise correspondences. However commercially available stitching algorithms such as [8], do not provide point wise correspondences between the constituent and stitched images, which renders the application of FR IQA indices infeasible. Therefore we report the performances of [14] and [15] on a subset of the ISIQA database which does not include stitched images obtained using [8]. It can be observed from the Table 4.5 that the performance of SIQE is superior when compared with FR-IQA indices as well.

4.3 Significance of Each Conceptual Feature

To study the effect of each conceptual feature present in SIQE framework, each feature is tested in isolation for quality prediction and the correlation values are reported in Table 4.6. It can be inferred from the table that while the marginal model individually provides a slightly higher performance than the bivariate model, their performance greatly improves when employed together. We also conduct an experiment to study the benefit of constituent images where features from constituent images (f_{1-36}^c) are omitted and only features from the stitched image (f_{1-36}^s) are used to train the regressor. The drop in performance emphasizes the necessity of reference information in constituent images for efficient quality prediction. Employing only

stitched image features f^s for training is in effect an NR-IQA approach. Comparing with the performance of NR-IQA metrics listed in Table 4.3 indicates that features of SIQE when employed under an NR framework still outperform the existing NR-IQA methods. Table 4.3 also depicts the consequence of patch weighting where a drop in performance is observed when all patches are equally weighted.

Feature	SROCC	LCC
Marginal statistics model (f_{1-12})	0.7825	0.7853
Bivariate model (f_{13-36})	0.7085	0.7157
Features from stitched image (f_{1-36}^s) (when constituent image features are omitted)	0.6650	0.6929
SIQE (f_{1-36}^s and f_{1-36}^c)	0.8318	0.8395
SIQE (without patch weighting)	0.8041	0.8161

Table 4.6: Each conceptual feature in isolation of other features, median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets

4.4 Application of Univariate GMM for capturing correlation

In Sec 3.3 we proposed a bivariate GMM for capturing increased correlation arising due to ghosting. We also illustrated in Fig. 3.6c the large difference in peak values of conditional distribution compared to marginal distribution of unnormalized coefficients for ghosted patches. Here we perform an experiment where a univariate GMM is used in place of the bivariate GMM and the results are reported in Table 4.7. Let f^{c_u} and f^{s_u} represent the features obtained from univariate GMM for constituent and stitched images respectively. Since a univariate model is used, the variance is single dimensional resulting in a 12 dimensional (6 orientations \times 2 horizontal and vertical neighbors) feature vector $f_{13-24}^{c_u}, f_{13-24}^{s_u}$. These features are concatenated with features from marginal statistics ($f_{1-12}^{c_u} = f_{1-12}^c$ and $f_{1-12}^{s_u} = f_{1-12}^s$) and subtracted to obtain $f_{1-24}^u = f_{1-24}^{c_u} - f_{1-24}^{s_u}$. The values in the Table 4.7 indicate that bivariate statistics provide

Feature	SROCC	LCC
Univariate model (f_{13-24}^u)	0.5773	0.6208
Bivariate model (f_{13-36})	0.7085	0.7157
Univariate SIQE (f_{1-24}^u)	0.8089	0.8163
SIQE (f_{1-36})	0.8318	0.8395

Table 4.7: Performance comparison using univariate GMM, median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets

superior performance in both scenarios, when used in conjunction with marginal statistics as well as when used individually. These results reinforce our claim that bivariate modeling is stronger than univariate for capturing stitching induced distortions.

4.5 Effect of Quantization Levels in Patch Weighting

GLCM based patch weighting discussed in Sec. 3.4 is dependent on the number of quantization levels used in constructing GLCM. The performance variation for changes in number of quantization levels of GLCM is shown in Fig. 4.5. It can be inferred from the table that performance is relatively stable having very low variance across quantization levels.

Number of Levels	SROCC	LCC
No patch weighting	0.8041	0.8161
5	0.8315	0.8395
10	0.8237	0.8362
13	0.8318	0.8395
20	0.8226	0.8315
24	0.8260	0.8351
29	0.8158	0.8253
43	0.8087	0.8177
64	0.8058	0.8171
128	0.8039	0.8180
256	0.8045	0.8199

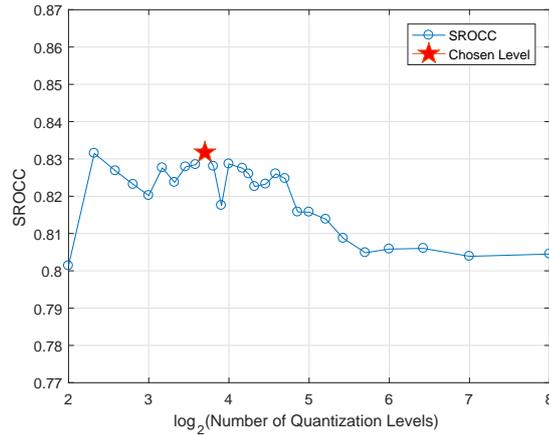


Figure 4.5: Effect of number of quantization levels of GLCM on the performance, median values of SROCC and LCC for 1000 iterations of randomly chosen train and test sets

4.6 Time Complexity

The computational complexity of SIQE is mainly dependent on computations of steerable pyramid decompositions, parameter estimation of the GGD model and the EM algorithm for estimating GMM parameters. In Table 4.8 we list the relative time taken at each stage for quality computation (once trained). It can be noted that bulk of the time is spent in estimating GMM parameters owing to the iterative nature of the EM algorithm.

We also compare the overall time taken by SIQE with BRISQUE, NIQE and DIIVINE in Table 4.6. The values reported are for a 100×100 image patch run on a 4 GHz PC with 32 GB RAM. We use unoptimized MATLAB code for all algorithms for ensuring a fair comparison. As Table 4.6 demonstrates, while SIQE takes more time than BRISQUE and NIQE, the time taken is lesser than DIIVINE which also uses a multi-orientation decomposition.

Step	Percentage of time
Steerable Pyramid Decomposition	2.92
GGD	3.21
GMM	93.8

Table 4.8: Complexity analysis in SIQE. Values reflect the percentage of time spent on each step in SIQE

Algorithm	Time (seconds)
BRISQUE	0.1092
NIQE	0.0116
DIIVINE	1.2369
SIQE	1.0534

Table 4.9: Comparison of the amount of time taken to compute various quality measures for a 100×100 image patch

Chapter 5

Conclusion and Future Work

We presented a subjective study to evaluate the effects of stitching induced distortions in the context of VR applications. This study consists of 264 stitched images derived from 26 scenes and evaluated by 35 subjects. The resulting IISc Stitched Image Quality Assessment (ISIQA) Database is unique in terms of content as well as types of distortions present and is publicly available for research purposes [1].

We proposed a steerable pyramid based model for quality assessment of stitched images - the Stitched Image Quality Evaluator (SIQE) and evaluated its performance on the ISIQA database. An important characteristic of SIQE is that it does not require the knowledge of the stitching algorithm. We detailed the SIQE framework and the features extracted, in particular the features obtained from the bivariate model and its significance in capturing additional correlations arising due to distortions. We performed a holistic evaluation of SIQE in terms of correlation with human perception and established that the performance of SIQE is superior and more robust than the existing NR algorithms. We conducted ablation studies where we analyzed each conceptual feature individually to gauge its contribution and demonstrated the importance of employing features from both marginal statistics of divisive normalized coefficients as well as bivariate statistics of unnormalized coefficients. We also substantiated the consequence of

using features from constituent images in conjunction with features from stitched images in enhancing the performance when compared to employing features only from stitched images. A software release of SIQE has been made available online [1].

In our proposed model we did not explicitly design features that account for color distortions. In Sec. 2.5 we made an important observation that in many of the images with color distortions the MOS values are relatively closer to images which have little or no distortion. This raises an important question about the relevance of color distortion in the context of VR applications. The possibility of improving the performance by designing features that capture color artifacts remains an open question. Another aspect we have not addressed in this work concerns the problem of geometric shape changes that occur due to stitching. This again requires a more careful consideration as shape changes need not always be a perceivable distortion and hence its relevance in stitched image QA needs to be investigated further.

Appendix A

Decrease in Variance due to Ghosting

Rewriting Equation 3.1 associated with blending stage with reference to Fig. 3.2, where X denotes the set of all coordinates representing the overlapping region in I_1 and for any $x \in X$

$$I(x) = (1 - \alpha(x))I_1(x) + \alpha(x)I_2(Hx), \quad (\text{A.1})$$

where $\alpha(x) \in (0, 1)$ denotes the weighting value at location x . Ghosting is observed when $I_1(x) \neq I_2(Hx)$ for any $x \in X$. Also in the case of perfect alignment $I_1(x) = I_2(Hx)$ holds true resulting in $I(x) = I_1(x)$ when substituted in Equation A.1, thus implying absence of ghosting. Let L represent the kernel associated with steerable pyramid decompositions. Applying L on Equation A.1 from Equation A.1 we obtain

$$L * I(x) = L * ((1 - \alpha(x))I_1(x)) + L * (\alpha(x)I_2(Hx)). \quad (\text{A.2})$$

For simplicity we neglect the notation x and represent $L * I(x)$ as LI . Following similar convention for $I_1(x)$ and $I_2(Hx)$ in Equation A.2 we get

$$LI = \alpha(LI_2 - LI_1) + LI_1. \quad (\text{A.3})$$

In order to show that $\text{Var}(LI) \leq \text{Var}(LI_1)$ or $\text{Var}(LI) \leq \text{Var}(LI_2)$, we assume that $\mathbb{E}[LI_1] = 0$, $\mathbb{E}[LI_2] = 0$, $\mathbb{E}[LI] = 0$ (\mathbb{E} denotes expectation operator) as steerable decomposition results in zero mean subbands. Also we assume that $\sqrt{\mathbb{E}[LI_1^2]} = \sqrt{\mathbb{E}[LI_2^2]}$, which is reasonable as overlapping regions have similar content in them resulting in similar statistics. Under these assumptions and α being a deterministic quantity, we obtain

$$\begin{aligned} \mathbb{E}[LI^2] &= \alpha^2 \mathbb{E}[LI_2^2] + \alpha^2 \mathbb{E}[LI_1^2] - 2\alpha^2 \mathbb{E}[LI_1 LI_2] + \mathbb{E}[LI_1^2] + 2\alpha \mathbb{E}[LI_1 LI_2] - 2\alpha \mathbb{E}[LI_1^2], \\ \implies \mathbb{E}[LI^2] - \mathbb{E}[LI_1^2] &= \alpha(\mathbb{E}[LI_1^2](\alpha - 2) + \alpha \mathbb{E}[LI_2^2] + 2\mathbb{E}[LI_1 LI_2](1 - \alpha)). \end{aligned} \quad (\text{A.4})$$

Applying Cauchy-Schwarz inequality $\mathbb{E}[AB] \leq \sqrt{\mathbb{E}[A^2]} \sqrt{\mathbb{E}[B^2]}$ with $A = LI_1, B = LI_2$ to Equation A.4 we get

$$\mathbb{E}[LI^2] - \mathbb{E}[LI_1^2] \leq \alpha(\mathbb{E}[LI_1^2](\alpha - 2) + \alpha \mathbb{E}[LI_2^2] + 2\sqrt{\mathbb{E}[LI_1^2]} \sqrt{\mathbb{E}[LI_2^2]}(1 - \alpha)). \quad (\text{A.5})$$

As $\mathbb{E}[LI_1] = 0$, $\mathbb{E}[LI_2] = 0$, we obtain $\mathbb{E}[LI^2] = \text{Var}(LI)$ and $\mathbb{E}[LI_1^2] = \text{Var}(LI_1)$. Denoting $\mathbb{E}[LI_1^2] = \sigma^2(LI_1)$ and $\mathbb{E}[LI_2^2] = \sigma^2(LI_2)$ in Equation A.5 leads to

$$\text{Var}(LI) - \text{Var}(LI_1) \leq \alpha(\sigma^2(LI_1)(\alpha - 2) + \alpha \sigma^2(LI_2) + 2\sigma(LI_1)\sigma(LI_2)(1 - \alpha)), \quad (\text{A.6})$$

$$\implies \text{Var}(LI) - \text{Var}(LI_1) \leq \alpha(\alpha \sigma(LI_2) - (\alpha - 2)\sigma(LI_1))(\sigma(LI_2) - \sigma(LI_1)) \quad (\text{A.7})$$

Using the assumption $\sqrt{\mathbb{E}[LI_1^2]} = \sqrt{\mathbb{E}[LI_2^2]}$ which is equivalent to $\sigma(LI_1) = \sigma(LI_2)$, in Equation A.7 we get

$$\text{Var}(LI) - \text{Var}(LI_1) \leq 0, \quad (\text{A.8})$$

$$\implies \text{Var}(LI) \leq \text{Var}(LI_1) \quad (\text{A.9})$$

Therefore variance decreases due to ghosting under the assumption $\sqrt{\mathbb{E}[LI_1^2]} = \sqrt{\mathbb{E}[LI_2^2]}$.

References

- [1] “(2018) IISc Stitched Image Quality Assessment Database,” <https://pavancm.github.io/stitched-qa/>. 42, 43
- [2] “Facebook Surround 360,” <https://code.facebook.com/posts/1755691291326688/introducing-facebook-surround-360-an-open-high-quality-3d-360-video-capture-system/>. 1
- [3] R. Szeliski, “Image alignment and stitching: A tutorial,” *Foundations and Trends® in Computer Graphics and Vision*, vol. 2, no. 1, pp. 1–104, 2006. 1, 7, 10, 12, 13
- [4] M. Brown and D. G. Lowe, “Recognising panoramas,” in *Proceedings Ninth IEEE International Conference on Computer Vision*, vol. 2, Oct 2003, pp. 1218–1225. 1
- [5] J. Zaragoza, T. J. Chin, Q. H. Tran, M. S. Brown, and D. Suter, “As-projective-as-possible image stitching with moving DLT,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 7, pp. 1285–1298, July 2014. 1, 7, 10, 13
- [6] C. H. Chang, Y. Sato, and Y. Y. Chuang, “Shape-preserving half-projective warps for image stitching,” in *IEEE Conference on Computer Vision and Pattern Recognition*, June 2014, pp. 3254–3261. 1, 7, 11, 13
- [7] “Autostitch,” <http://matthewalunbrown.com/autostitch/autostitch.html>. 1

-
- [8] “Microsoft research image composite editor,” <http://research.microsoft.com/en-us/um/redmond/groups/ivm/ice/>. 1, 13, 38
- [9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004. 2
- [10] Z. Wang and A. C. Bovik, “Mean squared error: Love it or leave it? a new look at signal fidelity measures,” *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, Jan 2009. 2
- [11] A. K. Moorthy and A. C. Bovik, “Blind image quality assessment: From scene statistics to perceptual quality,” *IEEE Trans. Image Process.*, vol. 20, no. 12, Dec. 2011. 2, 23, 25, 26, 32, 36
- [12] A. Mittal, A. K. Moorthy, and A. C. Bovik, “No-reference image quality assessment in the spatial domain,” *IEEE Trans. Image Process.*, vol. 21, no. 12, pp. 4695–4708, Dec. 2012. 2, 32, 36
- [13] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” *IEEE Sig. Process. Letters*, vol. 20, no. 3, pp. 209–212, Mar. 2013. 2, 36
- [14] W. Xu and J. Mulligan, “Performance evaluation of color correction approaches for automatic multi-view image and video stitching,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, June 2010, pp. 263–270. 3, 7, 12, 37, 38
- [15] H. S. Qureshi, M. M. Khan, R. Hafiz, Y. Cho, and J. Cha, “Quantitative quality assessment of stitched panoramic images,” *IET Image Processing*, vol. 6, no. 9, pp. 1348–1358, December 2012. 3, 37, 38

-
- [16] P. Paalanen, J.-K. Kämäräinen, and H. Kälviäinen, “Image based quantitative mosaic evaluation with artificial video,” *Scandinavian Conference on Image Analysis*, pp. 470–479, 2009. [3](#)
- [17] G. Cheung, L. Yang, Z. Tan, and Z. Huang, “A content-aware metric for stitched panoramic image quality assessment,” in *IEEE International Conference on Computer Vision Workshops (ICCVW)*, Oct 2017, pp. 2487–2494. [3](#)
- [18] M. Solh and G. AlRegib, “MIQM: A novel multi-view images quality measure,” in *2009 International Workshop on Quality of Multimedia Experience*, July 2009, pp. 186–191. [3](#)
- [19] M. Yu, H. Lakshman, and B. Girod, “A framework to evaluate omnidirectional video coding schemes,” in *IEEE International Symposium on Mixed and Augmented Reality*, Sept 2015, pp. 31–36. [3](#), [7](#)
- [20] V. Zakharchenko, K. P. Choi, and J. H. Park, “Quality metric for spherical panoramic video,” in *Proc. SPIE, Optics and Photonics for Information Processing X*, vol. 9970, Sept 2016, p. 99700C. [3](#)
- [21] Y. Sun, A. Lu, and L. Yu, “Weighted-to-spherically-uniform quality evaluation for omnidirectional video,” *IEEE Signal Processing Letters*, vol. 24, no. 9, pp. 1408–1412, Sept 2017. [3](#)
- [22] W. Sun, K. Gu, G. Zhai, S. Ma, W. Lin, and P. L. Calle, “CVIQD: Subjective quality evaluation of compressed virtual reality images,” in *IEEE International Conference on Image Processing (ICIP)*, Sept 2017, pp. 3450–3454. [3](#)
- [23] E. Upenik, M. Rerabek, and T. Ebrahimi, “On the performance of objective metrics for omnidirectional visual content,” in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–6. [3](#)

-
- [24] H.-t. Lim, H. G. Kim, and Y. M. Ro, “VR IQA NET: Deep virtual reality image quality assessment using adversarial learning,” in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018. 3
- [25] E. Upenik, M. ebek, and T. Ebrahimi, “Testbed for subjective evaluation of omnidirectional visual content,” in *2016 Picture Coding Symposium (PCS)*, Dec 2016, pp. 1–5. 7
- [26] A. D. Abreu, C. Ozcinar, and A. Smolic, “Look around you: Saliency maps for omnidirectional images in vr applications,” in *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, May 2017, pp. 1–6. 7
- [27] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein, “Saliency in vr: How do people explore virtual environments?” *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 4, pp. 1633–1642, April 2018. 7
- [28] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004. 7, 8
- [29] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded up robust features,” *Computer vision—ECCV 2006*, pp. 404–417, 2006. 7, 8
- [30] M. A. Fischler and R. C. Bolles, “Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography,” *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981. 7, 9
- [31] V. Kwatra, A. Schödl, I. Essa, G. Turk, and A. Bobick, “Graphcut textures: image and video synthesis using graph cuts,” in *ACM Transactions on Graphics (ToG)*, vol. 22, no. 3. ACM, 2003, pp. 277–286. 7, 12
- [32] M. Brown and D. G. Lowe, “Automatic panoramic image stitching using invariant features,” *International journal of computer vision*, vol. 74, no. 1, pp. 59–73, 2007. 7, 12

- [33] M. Uyttendaele, A. Eden, and R. Szeliski, “Eliminating ghosting and exposure artifacts in image mosaics,” in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, 2001, pp. II-509–II-516 vol.2. [7](#), [12](#)
- [34] P. J. Burt and E. H. Adelson, “A multiresolution spline with application to image mosaics,” *ACM Transactions on Graphics (TOG)*, vol. 2, no. 4, pp. 217–236, 1983. [7](#), [12](#), [13](#)
- [35] Z. Farbman, R. Fattal, and D. Lischinski, “Convolution pyramids.” *ACM Trans. Graph.*, vol. 30, no. 6, pp. 175–1, 2011. [7](#), [12](#), [13](#)
- [36] P. Pérez, M. Gangnet, and A. Blake, “Poisson image editing,” in *ACM Transactions on graphics (TOG)*, vol. 22, no. 3. ACM, 2003, pp. 313–318. [7](#), [13](#)
- [37] M. Tanaka, R. Kamio, and M. Okutomi, “Seamless image cloning by a closed form solution of a modified poisson problem,” in *SIGGRAPH Asia 2012 Posters*. ACM, 2012, p. 15. [7](#), [13](#)
- [38] ITU-R Recommendation BT.500-11, “Methodology for the subjective assessment of the quality of television pictures,” *International Telecommunication Union*, 2000. [15](#), [16](#)
- [39] A. M. van Dijk, J.-B. Martens, and A. B. Watson, “Quality assessment of coded images using numerical category scaling,” in *Proc. SPIE - Advanced Image and Video Communications and Storage Technologies*, 1995. [16](#)
- [40] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, “A statistical evaluation of recent full reference image quality assessment algorithms,” *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3440–3451, Nov 2006. [18](#), [36](#)
- [41] K. Seshadrinathan, R. Soundararajan, A. C. Bovik, and L. K. Cormack, “Study of subjective and objective quality assessment of video,” *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1427–1441, June 2010. [18](#)

-
- [42] D. Jayaraman, A. Mittal, A. K. Moorthy, and A. C. Bovik, “Objective quality assessment of multiply distorted images,” in *2012 Conference Record of the Forty Sixth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, Nov 2012, pp. 1693–1697. [18](#)
- [43] E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, “Shiftable multiscale transforms,” *IEEE Transactions on Information Theory*, vol. 38, no. 2, pp. 587–607, March 1992. [23](#)
- [44] H. R. Sheikh and A. C. Bovik, “Image information and visual quality,” *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430–444, Feb. 2006. [23](#), [25](#)
- [45] Q. Li and Z. Wang, “Reduced-reference image quality assessment using divisive normalization-based image representation,” *IEEE Journal of Selected Topics in Signal Processing: Special issue on Visual Media Quality Assessment*, vol. 3, pp. 202–211, Apr. 2009. [23](#), [25](#), [26](#)
- [46] M. J. Wainwright and E. P. Simoncelli, “Scale mixtures of Gaussians and the statistics of natural images,” in *Adv. Neural Information Processing Systems*, vol. 12, May 2000, pp. 855–861. [26](#), [27](#), [29](#)
- [47] K. Sharifi and A. Leon-Garcia, “Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, no. 1, pp. 52–56, 1995. [27](#)
- [48] M. J. Wainwright, O. Schwartz, and E. P. Simoncelli, “Natural image statistics and divisive normalization: modeling nonlinearities and adaptation in cortical neurons,” *Statistical theories of the brain*, pp. 203–222, 2002. [27](#)
- [49] C. C. Su, L. K. Cormack, and A. C. Bovik, “Oriented correlation models of distorted natu-

- ral images with application to natural stereopair quality evaluation,” *IEEE Transactions on Image Processing*, vol. 24, no. 5, pp. 1685–1699, May 2015. [29](#), [33](#)
- [50] Z. Sinno, C. Caramanis, and A. C. Bovik, in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018. [29](#)
- [51] G. E. Box and G. C. Tiao, *Bayesian inference in statistical analysis*. John Wiley & Sons, 2011, vol. 40. [29](#)
- [52] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, “Wavelet-based statistical signal processing using hidden Markov models,” *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 886–902, Apr 1998. [29](#)
- [53] J. K. Romberg, H. Choi, and R. G. Baraniuk, “Bayesian tree-structured image modeling using wavelet-domain hidden markov models,” *IEEE Transactions on Image Processing*, vol. 10, no. 7, pp. 1056–1068, Jul 2001. [29](#)
- [54] A. P. Dempster, N. M. Laird, and D. B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977. [29](#)
- [55] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov 1973. [30](#), [35](#)
- [56] B. Schölkopf, A. J. Smola, R. C. Williamson, and P. L. Bartlett, “New support vector algorithms,” *Neural computation*, vol. 12, no. 5, pp. 1207–1245, 2000. [32](#)
- [57] M. A. Saad, A. C. Bovik, and C. Charrier, “Blind prediction of natural video quality,” *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1352–1365, March 2014. [32](#)
- [58] M. Narwaria and W. Lin, “Objective image quality assessment based on support vector

- regression,” *IEEE Transactions on Neural Networks*, vol. 21, no. 3, pp. 515–519, March 2010. 32
- [59] M. Narwaria and W. Lin, “SVD-based quality metric for image and video using machine learning,” *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 42, no. 2, pp. 347–364, April 2012. 32
- [60] C.-C. Chang and C.-J. Lin, “LIBSVM: A library for support vector machines,” *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, May 2011. [Online]. Available: <http://doi.acm.org/10.1145/1961189.1961199> 32
- [61] C.-C. Su, L. K. Cormack, and A. C. Bovik, “Bivariate statistical modeling of color and range in natural scenes,” vol. 9014, 2014, pp. 9014 – 9014 – 10. 33