CHAPTER 2
LITERATURE SURVEY ON STATISTICAL IMAGE SIMILARITY MEASURES FOR COMPARING TWO IMAGES

2.1 Introduction

For convenience and clarity of discussions, the following abbreviations will be used. Firstly image similarity measure will be denoted by ISM. Statistical-based similarity measure will be denoted by SISM. Full-reference image, Reduced-reference image and No-reference image are labelled as FR, RR and NR, respectively. This chapter consists of a brief survey on ISM, followed by a discussion of three image issues and finally some consideration of the properties of selected ISM.

This chapter surveys statistical measures of similarity between two images from years 1980 to 2010. In total 330 ISMs were found from 30 international published journals, including IEEE Transactions on Image Processing, IEEE Transactions on Pattern Analysis and Machine Intelligent and WSEAS Transactions on Signal Processing, and only 85 of these ISMs or approximately 25% are SISMs. The frequency of SISMs used and their applications over time are listed in Table 2.1 and Table 2.2. Most of the studies surveyed in this period compare a transformed image with a full-reference (original) image (see Section 1.4.1) mainly for computational ease. Comparisons are also carried out for the no-reference and reduced-reference cases.

Since the majority of SISMs were designed for comparing two images, namely the reference image and the transformed image (henceforth labelled as the full-reference approach), issues related to FR-SISM will be highlighted, firstly the imperfect reference image, secondly the number of image attributes, and finally combining the local image information or global image information into one measure. The success and limitation of FR-SISMs to address these issues will then be highlighted. These discussions are directed towards a proposal for a new similarity measure based on statistical correlation and denoted by R_P^2. The potential of R_P^2 as a performance indicator will be investigated in the next chapters together with its applications on a particular data compression problem and a character recognition problem to illustrate the virtues of R_P^2.

2.2 Other Surveys Done

Over the period of study, there were a number of surveys done and the majority of similarity measures considered were not SISM. At most 3 survey papers showed the applications of SISM, for example Rubner et al. (2001). Eckert & Bradley (1998) discussed quality measures applied to a still image compression problem. This paper classified the objective quality into four major categories, which are Mathematical metrics, metrics which incorporate the contrast sensitivity function (CSF) and luminance adaptation, metrics which incorporate observer preferences for suprathreshold artefact and Threshold perceptual metrics. Zhang (1996) discussed image segmentation using empirical goodness methods and empirical discrepancy methods. Eskiciogu & Fisher (1995) compared the performance of a set of 14 metrics for the image compression problems, amongst those considered were Average Distance, Structural Content, Laplacian Mean Square Error and Hosaka Plot. Avcibas et al. (2002) considered the objective image quality metrics using Pixel different, non statistical Correlation, Edge measurement, Spectral distance, Context-based and human visual system (HVS)-based measures. On the other hand, Rubner et al. (2001) divided the dissimilarity measures for colour and texture into four categories, which are (i) Heuristic histogram distances such as Minkowski-form distance and the Weighted mean variance, (ii) Non parametric test statistic such as Kolmogorov-Smirnov distance, Chi-square statistic and a statistic of the Cramer von Mises type, (iii) Information-theory Divergences such as Kullback-Leibler divergence and Jeffer-divergence, and (iv)

Ground distance measures such as the Quadratic Form and the Earth Movers Distance. Eskiciogu (2000) divided the subjective quality measures into absolute and comparative categories, while the objective quality measures are divided into numerical and graphical categories. Holden et al. (2000) compared eight voxel similarity measures for 3-D serial MR brain image registration. Kim et al. (2004) and Kinape & Amorim (2003) discussed generally on a selected quality measures. This is followed by Cadik & Slavik (2004) who evaluated two approaches to objective ISM which are visible differences predictor (VDP) and structural similarity measures (MSSIM). Comparison of digital video quality measures has also been emphasized by Winkler (2005). Winkler (2005) divided the video quality measures into four categories, i.e. pixel-based metrics, single-channel models, multi-channel models and specialized metrics. A list of visual quality measures together with its application is also provided. In 2006, Rix et al. (2006) reviewed some intrusive objective measures, nonintrusive measures and parametric methods for speech and audio quality. Lastly, Leontaris et al. (2007) compared fifteen similarity-based metrics, blocking-based metrics and blurring-based metrics in compressed video.

Most of the ISMs are problem-dependent, hence raising the issue of selecting the suitable ISMs. It would be desirable to have a single ISM which could be applied to several types of problems. This problem motivated the exploration of a statistical ISM. Based on these previous surveys, no thorough discussion on SISM has been carried out for the past 30 years. This probably explains why the statistical-based approach has not been well understood and the important role of SISMs has been overlooked.

2.3 Chronological Survey of SISMs and Their Applications

Historically, objective image similarity measures were constructed using the properties of human visual system in as early as 1950s (Baker & Carpenter, 1989). In 1959, Stultz and Zweig (1959) published the relationship between a magnified image and the scanning aperture size. The relationship yield root mean-square (RMS) granularity values, which correlated with the perceived noisiness. Similar progress was made in the definition of objective correlates of perceived sharpness by Crane in 1964. There after, image quality assessment and hence similarity measure became one of the main challenges for the image processing field. Two commonly used classical distance measures are Peak signal-to-noise ratio (PSNR) (Puttenstein et al., 2004) and Root mean square error (RMSE) (Rogerio et al., 2003). A large number of similarity measures based on various methodologies have been proposed and its number is still increasing.

The first FR-SISM found in this survey is a metric based on information theory proposed by Girod (1981) (see Table 2.1). Girod (1981) used mutual information rate as a quality index to measure Gaussian processes (see Equation (2.29) in Section 2.5.1(viii)) between two images. However, the studies of using statistical methodologies as a quality measure were not intensive during the period 1980 to 1989. There were six statistical-based measures proposed in this period, they were stated in Girod (1981), Yasnoff & Bacus (1984), Steinberg (1987), Haris et al. (1988), and two measures were given in Basseville (1989). In 1984, Yasnoff & Bacus (1984) proposed using the probability of Object Count Agreement as a quality measure for segmentation process. The probability value indicates that the number of objects of class I are having the same distribution in reference image and segmented image. Although, the concept of correlation has been widely used in this period, its' application was mainly limited to measuring the relationship between proposed objective metric and subjective assessment. The use of correlation as a compression quality measure was introduced in 1987 and 1988 by Steinberg (1987) and Harris et al. (1988), respectively. The former introduced normalized correlation as a ratio between the product of mean values of two images and the product of their root mean square. The latter proposed the peak correlation value as standardized cross-correlation.

There was a significant increase in the number of publications on the use of statistical-based measures in 1990 until 1999. Twenty-six measures for comparing two images were introduced in this period where the majority of them were published in the late 1990s. In 1990, a non-parametric method using Kolmogorov-Simirnov distance was introduced by Geman et al. (1990). It is defined as the maximal discrepancy between the cumulative distributions based on one-dimensional histogram. Moment functions such as mean and variance are well known for image feature description. Mean value describes the luminance level of an image, while variance represents the image contrast. Amongst the earliest statistical-based measure that utilized moment functions is the Weighted Mean Variance proposed by Manjunath & Ma (1996). This measure has worked well for texture-based image retrieval. In the same year, a new statistical-based measure called RED was proposed in speech coding by Erkelens & Broersen (1996). It is an absolute measure using a scaled sum of the squared differences between the true impulse response in speech signal and its estimate. The impulse responses can be computed with the ARMA Time Series processes.

After year 2000, the statistical approach started to be used widely by image processing researchers. Within a decade, there were 39 Full Reference image quality measures using statistical approach in comparing two images. During these period, the introduction of universal quality index or better known as mean structural similarity (MSSIM) measure showed a milestone progress in statistical approaches (see also Section 2.5.1). The MSSIM is the most popular and highly cited image similarity measure in these period. It was first introduced by Wang et al. (2002a) and Wang & Bovik (2002). The SSIM become popular there after. Many researchers started to further implement, modify or improve on the MSSIM usage for different image applications, such as Wang et al. (2002b), Toet & Lucassen (2003), Zhou et al. (2003), Wang et al. (2004), Piella (2004), Alparone et al. (2004), Bouzerdoum et al. (2004), Aja-Fernandez et al. (2006), Wang & Ma (2008), Brooks et al. (2008), Yang et al. (2008), Moorthy and Bovik (2009), Wirandi et al. (2009) and Garzelli & Nencini (2009). The MSSIM was designed using a new philosophy that does not treat image degradation as an error, but extracted structural information from the viewing field and adapted the human visual system (HVS) into the new metric. Wang et al. (2002a, 2004) defined their new quality measure as a product of the structural component, luminance component and contrast component values.

More recently, Wang et al. (2007) defined a performance metric for face recognition as an exponential function of the moment values mean and standard deviation. Almost at the same time, Mitra et al. (2007) used Bayesian inference to evaluate the performance of biometric authentication system. On the other hand, similarity between a given pattern and the query of home video is modelled by a probability value (Mei et al., 2007) and the mutual information combined with B-splines is used to evaluate and optimize the nonrigid medical image registration by Klein et al. (2007). Another recent application is Fronthaler et al. (2008) who use a modified correlation coefficient for fingerprint image quality measure.

While the other measures consider the reference image as fixed, however the Mutual Information Similarity (MIS) measure proposed by Chen et al. (2003) defined the images as random entities. This measure provides consistent result on multimodal remote sensing registration algorithms (Holden et al. 2000). The idea of random variable was extended by Chang et al. (2008a, 2008b) for JPEG compression problem where the reference image and JPEG codec image are both subjected to errors in their Functional Quality $\left(R_F^2\right)$ measure.

Initially the studies of quality metric or similarity measure between two images were concentrated on the Full Reference approach. The used of No Reference statistical-based measure only started after the late 1980s. One of the earlier NR measure was proposed in 1989 by Pal and Pal (1989) using higher order local entropy based on information theory. It was used to measure the region homogeneity in segmented images for the performance evaluation of segmentation processes. Another NR statistical-based measure found was proposed in 1996 by Berizzi and Corsini (1996) using negative image entropy. It was used to measure the contrast of an image obtained from 'through-the-wall' microwave imaging applications. Another two No-Reference statistical-based measures appeared in the period of 1990s were Mixed Effect Linear Model (Song et al., 1998) in 1998 and Sharpness metric (Zhang et al., 1999) in 1999. The Mixed Effect Linear Model applied regression ideas in assessing image compression while the Sharpness metric evaluates video quality using statistical moments.

The use of No-Reference statistical-based measures has increased accordingly after year 2000. There were seven NR statistical measures introduced; one in year 2000 by Hieu et al. (2000), four in year 2002 introduced by Sheikh et al. (2002), Wang et al. (2002a), Lu et al. (2002) and Marchant (2002), and two in year 2004 which proposed by Russo (2004) and Luo (2004). The Motion Statistics Based Region Similarity proposed by Hieu et al. (2000) was a probability-based metric and was used for video segmentation. Among the four measures introduced in 2002, two of them were applied to image compression using Principle Component Analysis (Sheikh et al., 2002) and Nonlinear Regression (Wang et al. 2002a). At the same year, the Entropy (Marchant, 2002) metric was applied to image acquisition problem and the Quantization Error (Lu et al., 2002) using probability was developed for video quality evaluation. Furthermore, moments-based Coefficient of Variation (Russo, 2004) and a probability-based Mixture of Gaussian (Luo, 2004) distribution were introduced in 2004 for segmentation and face detection problems, respectively. The Spearman rank order correlation was applied to video quality evaluation by Oelbaum et al. (2009).

Throughout this study, only one Reduced Reference method was found in the literature. Cheng & Cheng (2009) used the generalised Laplace distribution to model the natural images in their gradient domain. The parameters were then estimated by using the variance and kurtosis.

Table 2.1 and Table 2.2 show the Full Reference and No Reference statistical metrics respectively for quality assessment or similarity measurement from year 1980 – 2010. The reference lists were obtained from the study of 330 research publications for both statistical-based and non-statistical-based measures in various image applications. These tables showed the methods used in the cited publications and its related Statistics-fields. However, only the papers that had proposed new measure or had suggested some improvement on the existing measures are discussed here. It also indicates the number of quality measures proposed and the year they were introduced for the first time.

Table 2.1: Full Reference SISM from year 1980 to 2010.

1980-1989	**# Metrics**	**Reference**	**Proposed methods (Statistical area)**	**Applications**
1981	1	Girod	Mutual Information rate (Information theory)	General
1984	1	Yasnoff & Bacus	Object Count Agreement (Probability)	Segmentation
1987	1	Steinberg	Normalized correlation (Correlation)	Compression
1988	1	Harris et al.	Peak Correlation (Correlation)	SAR compression
1989	2	Basseville	Hellinger distance, Generalized Matusita distance (Probability)	Signal Processing, Pattern Recognition
Total	**6**			

Table 2.1: Continued.

1990-1999	# Metrics	Reference	Proposed methods (Statistical area)	Applications
1990	1	Lee et al.	Probability of error (Probability)	Segmentation
	1	Geman et al.	Kolmogorov Simirnov distance (Nonparametric)	Segmentation
1993	1	Pal & Bhandari	Symmetric divergence (Probability)	Segmentation
1996	1	Schiele & Crowley	Chi-square statistics (Nonparametric)	Image Correspondence
	1	Manjunath & Ma	Weighted Mean Variance (Moments)	Retrieval
	1	Erkelens & Broersen	RED Absolute Measure (Time Series)	Speech coding
	2	Berizzi & Gorsini	Rate Distortion Based Measure, Relative entropy (Information theory)	General
	1	Bhat & Nayar	Kappa Correlation (Correlation)	General
1997	2	Nielsen et al.	Pearson Correlation (Correlation), Normalized Covariance (Moments)	Signal Processing
	1	Puzicha et al.	Chi-square statistics (Nonparametric)	Retrieval
1998	1	Squire	Cohen Kappa Statistics (Non Parametric)	Retrieval
	1	Martens & Meesters	Multiple Correlation (Correlation)	Noise, Compression
	2	Vidal et al.	Entropy, PID error (Information theory)	Colour Image
	1	Yang	Scaling similarity (Moments)	Retrieval
	1	Howe	Percentile Blob-Based Similarity (Moments)	General
	1	Bhat & Nayar	Kappa Correlation (Correlation)	Image Correspondence
	1	Vidal et al.	Kullback Leibler divergence (Information theory)	Retrieval
	1	Andreutos et al.	Mean of the angular differences (Moments)	Retrieval
	1	Broersen	ME Relative Measure (Time Series)	General
1999	2	Comaniciu et al.	Bayes error, Bhattacharyya distance (Probability)	Retrieval
	1	Santani & Jain	Mahalanobis distance (Non Parametric)	Retrieval
	1	Robert et al.	Localized correlation (Correlation)	SAR compression
Total	**26**			
2000-2010	**# Metrics**	**Reference**	**Proposed methods (Statistical area)**	**Applications**
2000	2	Vasconcelos & Lippman	Maximum likelihood, Quadradic distance (Probability)	Retrieval
	1	Jia & Kitchen	Object based image similarity (Probability)	Retrieval

Table 2.1: Continued.

2000-2010	# Metrics	Reference	Proposed methods (Statistical area)	Applications
2000	1	Janssen & Blommaert	Measure for Partial Flexibility (Probability)	General
	2	Cramariuc et al.	Kendall's Tau Correlation, Spearman's Rho Correlation (Correlation)	General
2001	1	Ichalalene et al.	Subject contrast (Information theory)	X-Ray image
	2	Stevens	Pearson's R-Statistics (Correlation), Normalized Covariance (Moments)	3D-scene
	2	Rubner et al.	Cramer von Mises (Non Parametric), Jeffrey divergence (Information theory)	Retrieval
2002	1	Wang et al.	Mean Structural Similarity (Moments)	General
	1	Wang et al.	Weighted SSIM (Moments)	Video
	1	Avcibas	Block-based Spearman Rank Correlation (Correlation)	Colour image
2003	1	Toet & Lucassen	Colour Fidelity metric (Moments)	Colour image, General
	1	de Freitas Zampolo & Seara	Composed Quality Measure (Moments)	Image Restoration
	1	Goldberger et al.	Gaussian Mixture Kullback Leibler divergence (Probability)	Retrieval
	1	Chen et al.	Mutual Information similarity measure (Information theory)	Image Registration
2004	1	Ivkovic & Sankar	Average localized correlation (Correlation)	Noise
	1	de Freitas Zampolo & Seara	Bayesian Composed Quality Measure (Moments)	Noise
	1	Piella	Edge-dependent fusion quality index (Moments)	Image fusion
	1	Alparone & Baronti	Complex SSIM (Moments)	Image fusion
2005	1	Cates et al.	Sensitivities and Specificities (Probability)	Segmentation
2006	1	Tan & Chang	Canonical Correlation (Correlation)	Compression
	1	Loh & Chang	Modified SSIM (Moments)	Compression
	1	Chang & Tan	Multiple Correlation (Correlation)	Colour and Gray image
	1	Aja-Fernandez et al.	Quality Index based on Local Variance (Moments)	Medical image

Table 2.1: Continued.

2000-2010	# Metrics	Reference	Proposed methods (Statistical area)	Applications
2007	1	Mitra et al.	Bayesian inference (Probability)	Face recognition
	1	Wang et al.	Perfect Recognition Similarity Scores (Moments)	Face recognition
	1	Klein et al.	B-Splines based measure (Information theory)	Registration, medical image
2008	1	Fronthaler et al.	Modified Correlation (Correlation)	Fingerprint
	1	Wang & Ma	Weighted Sum MSSIM (Moments)	3D object
	1	Brooks et al.	Complex Wavelet SSIM (Moments)	Image, video
	1	Yang et al.	Perceptual Frame Interpolation Quality Metric (Moments)	Video compression
	1	Blanc et al.	Empirical Mean and Variance (Moments)	Texture image
	1	Chang et al.	Functional Quality Metric or R_F^2 (Correlation)	Image compression
2009	1	Moorthy & Bovik	Combined Percentile-Fixation SSIM (Moments)	General
	1	Wirandi et al.	Neural Network based SSIM (Moments)	General
	1	Garzelli & Nencini	Generalized universal quality index (Moments)	SAR image
Total	**39**			

Table 2.2: No Reference and Reduced Reference SISM.

1980-1989	# Metrics	Reference	No Reference SISM (Statistical area)	Applications
1989	1	Pal & Pal	Entropy (Information theory)	Segmentation
Total	**1**			
1990-1999	**# Metrics**	**Reference**	**No Reference SISM (Statistical area)**	**Applications**
1996	1	Berizzi, & Corsini	Negative entropy (Information theory)	Through-the-wall imaging
1998	1	Song et al.	Mixed Effect Linear Model (Correlation)	Compression
1999	1	Zhang et al.	Sharpness metric (Moments)	Video
Total	**3**			
2000-2010	**# Metrics**	**Reference**	**No Reference SISM (Statistical area)**	**Applications**
2000	1	Hieu et al.	Motion Statistics Based Region Similarity (Probability)	Video Segmentation
2002	1	Sheikh et al.	Prinsipal Component Analysis	Compression

Table 2.2: No Reference and Reduced Reference SISM.

2000-2010	**# Metrics**	**Reference**	**No Reference SISM (Statistical area)**	**Applications**
2002	1	Wang et al. (c)	Nonlinear Regression (Correlation)	Compression
	1	Lu et al.	Quantization Error (Probability)	Video
	1	Marchant	Entropy (Information theory)	Image Acquisition
2004	1	Russo	Coefficient of Variation (Moments)	Segmentation
	1	Luo	Mixture of Gaussian (Probability)	Face detection
2007	2	Mei et al.	Rule-Based method (Moments), Learning Based method (Probability)	Video
Total	**9**			
1980-2010	**# Metrics**	**Reference**	**Reduced Reference SISM (Statistical area)**	**Applications**
2009	1	Cheng & Cheng	Dual Derivative measure (Moments)	Natural image
Total	**1**			

2.3.1 Summary Comments on FR-SISMs

It is found that the number of statistical-based measures has increased over the periods of interest as shown in Figure 2.1. There are a total of seventy-five publications on FR-SISM in which only seventy one of them were considered because some publications used the same measure for different applications. There are one paper on Reduced Reference and thirteen publications on No Reference measures. Studies on SISM for image similarity and performance assessment are comparatively active as compared with non-statistical measures (see Table 2.3). Statistical-based approach is slightly less popular compare to the pixel-based approach. There were a total of 246 non-statistical ISMs collected from the same period on a wide range of image applications.

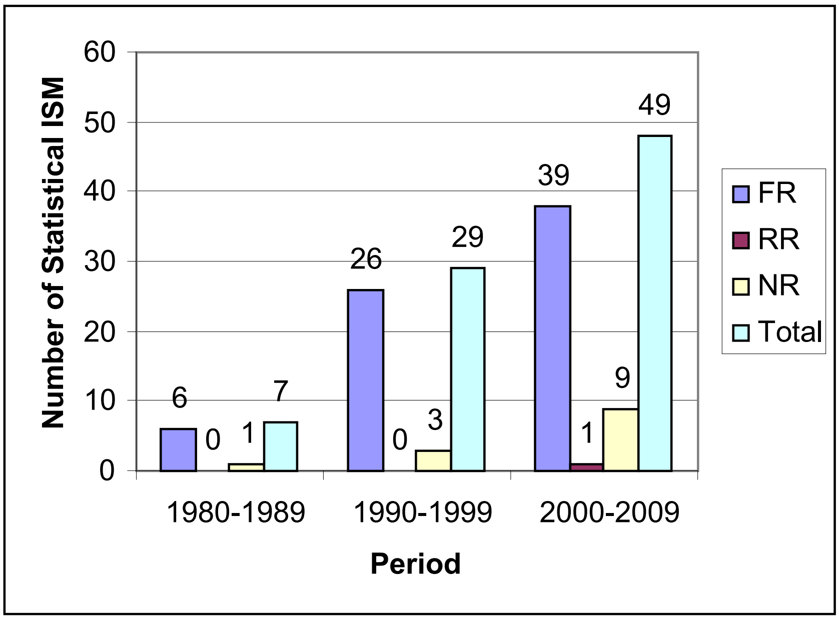


Figure 2.1: Summary of the FR, RR and NR-SISMs from year 1980 – 2010 for different original applications.

Applications of FR-SISM vary from image segmentation and pattern recognition to video quality assessment (see Table 2.1 and Table 2.2). Certain FR-SISM may have more than one application. Image retrieval is the most popular application for FR-SISM, especially on nonparametric approach. On the other hand, medical and video applications see an increasing application of FR-SISM after year 2000.

Table 2.3: Comparison of the number of Statistical based and Non-Statistical based ISMs for different applications from year 1980 to 2010.

ISM	FR	RR	NR	Total
Statistical based	71	1	13	85
Pixel based	86	2	9	97
Structural based	43	1	1	45
Neighborhood based	15	0	1	16
HVS based	19	0	1	20
Graphical based	11	0	0	11
Subjective	26	0	5	31
Others	19	4	3	26
Total	290	8	33	331

It is observed that most of the information theory-based measures were applied to benchmark the performance of image retrieval system. The probability approach was commonly used to assess the quality of image segmentation and image retrieval application. Also, most researchers applied correlations to evaluate the quality of image compression. Image comparison using moments tend to be more popular in general image quality problem regardless of their applications. The time series approach is the least frequently used method for FR-SISM. An illustration of these comparisons is given in Figure 2.2.

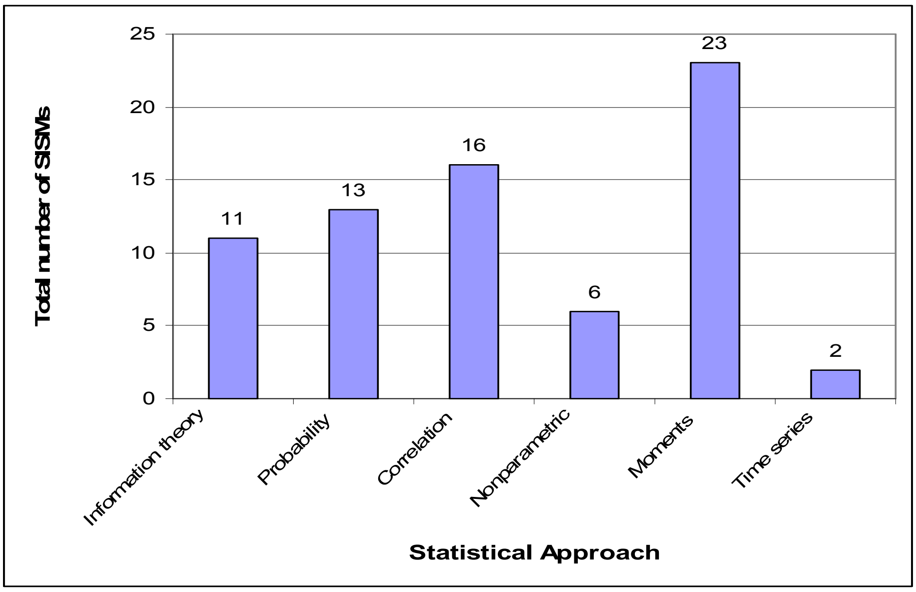


Figure 2.2: Number of FR-SISMs for various statistical approaches from year 1980 to 2010.

2.4 Issues Related to FR-SISM

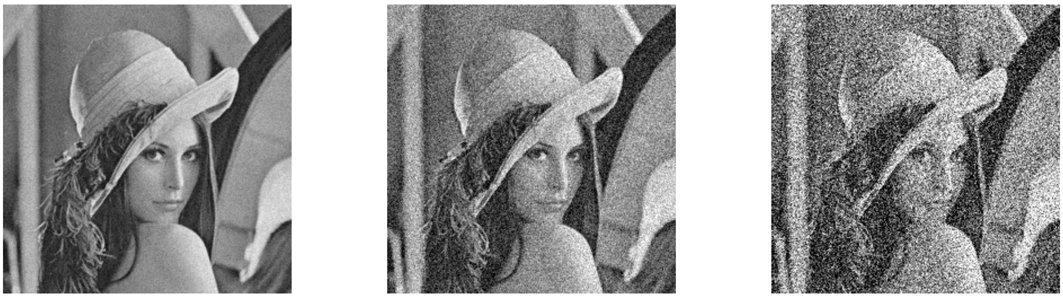
The ISMs surveyed were used in many applications and each of them works under certain specified constraints or conditions. Weighted Mean Variance (WMV) and MSSIM assumed the full reference image is of perfect quality (Wang et al., 2002a), while many of them considered single image attribute (Eskiciogu & Fisher, 1995), for example one image attribute extracted from each reference image and distorted image. Furthermore, advantages of the global measure and the localized measure were documented in many articles, but a solution that combines both of them together has not been found in this survey. MSSIM for example, proposed to compute the global measure by calculating the mean of local quality indices. Henceforth this section discusses three issues or image problems that have not been solved simultaneously. Firstly, full-reference image subject to error, secondly the need to compare images using multiple image attributes and finally the need to combine image local and global information. An example of JPEG compression is used to illustrate these issues.

2.4.1 Issue One: The Need to Consider Full Reference Image that Subject to Error

Most existing ISMs assume the full reference image is of perfect quality. In practice, it is not easy to obtain a perfect reference image because of pre-processing procedures such as image sampling, data transmission, data storage (such as changing the signal/image format to reduce storage size) and image enhancement always resulted in the existence of noise in the reference image used. This is true for the end-user of a digital image, which usually does not have the original image and has no idea about the level of image degradation. Wang et al. (2002a) has also pointed out this issue even though they argued that the assumption of perfect reference is reasonable for image/video coding and communication applications. This means that the assumption of perfect reference may not be true for many other conditions.

For example, to further illustrate this issue, a manufacturer of a camera brand located in Malaysia requests its headquarters from England to provide some standard reference pictures to optimise the algorithms and the parameter settings of a new camera product. Due to large image size and time factor, the engineer from the headquarters decided to send the reference pictures in jpeg format through the internet instead of posting the analogue image in its usual form. The above scenario explains two processes when artifacts are introduced in the reference images. The first process is when the images are converted to jpeg format. These compression artifacts include the blocking effect, blur, ringing and others (Winkler, 2005). The second process occurs during the transmission of images from England to Malaysia through the internet. Winkler (2005) pointed that this transmission error is often overlooked by researchers. In this example, the manufacturer would “optimize” the algorithms and parameter settings based on a lower quality reference image. This greatly affect the product quality manufactured by the manufacturer and may have negative financial consequence.

Figure 2.3 shows the effect of using imperfect reference image for image quality assessment. Selected Gaussian noise were added into the original reference image and compared to the JPEG codec image. The numerical values of three FR-ISMs and one non-statistical metric were calculated for selected compression factor ranging from 1 to 100 (see Figure 2.4), such that high metric value and large compression factor may be used as an indicator of quality. These measures are MSSIM, squared Pearson’s correlation (denoted Rs2 or R_S^2), Chi-Square measure (Chi2) and root mean squared error (RMSE). A non-statistical measure, RMSE is used for comparison since it is one of the most commonly used ISM. Note that the first two methods measure the similarity of the two images, while the last two methods measure their distance. Figure 2.4(a) to Figure 2.4(c) show that the performance of all metrics decline when the reference image is subjected to the increasing amount of errors.



(a) (b) (c)

Figure 2.3: Non-perfect Lena reference image. (a) with Gaussian noise N(0,0.001), (b) with Gaussian noise N(0,0.01), and (c) with Gaussian noise N(0,0.05).

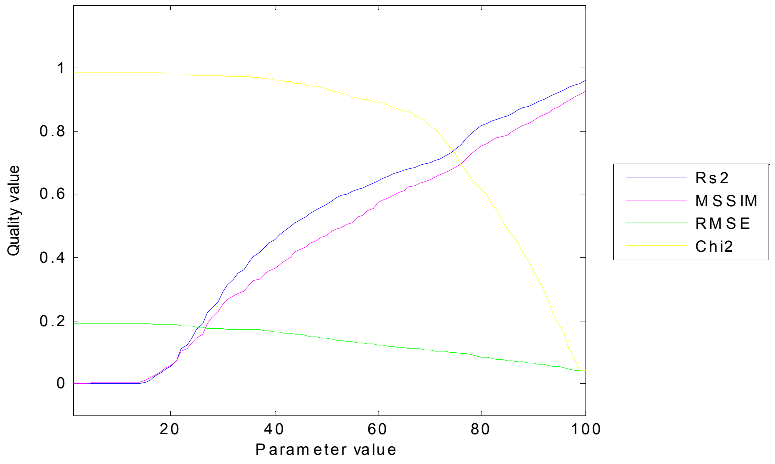


Figure 2.4(a): Image similarity values obtained from Fig. 2.3(a).

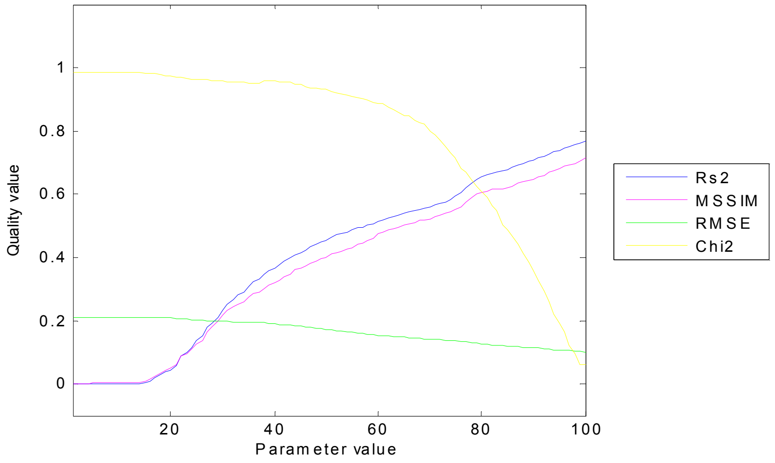


Figure 2.4(b): Image similarity values obtained from Fig. 2.3(b).

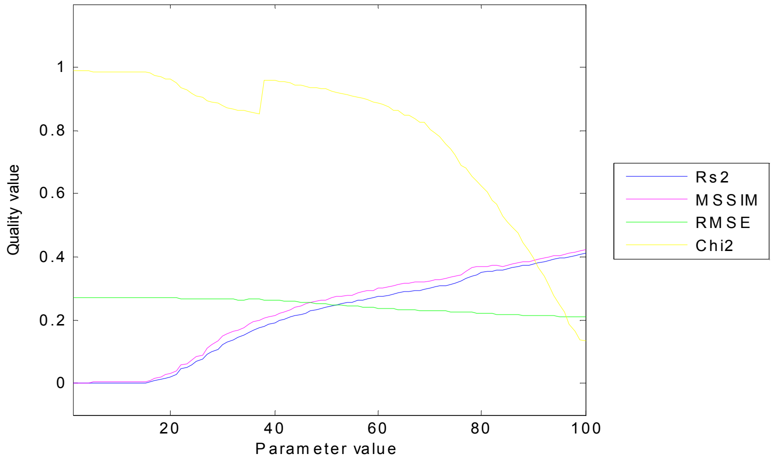


Figure 2.4(c): Image similarity values obtained from Fig. 2.3(c).

2.4.2 Issue Two: The Need to Compare Images Using Multiple Image Attributes

Image quality is one of those concepts that combine many interrelated factors to create what people perceive (Clarity's Product Marketing Department, 2002), suggesting that a combination of several image attributes is required for determining image quality. Further, Sprawls (1993) indicates that every image quality attribute counts. Among the most frequently used image quality attributes are brightness, luminance, and contrast.

The need to consider multiple image attributes is further supported by remarks from Keelan (2002). Only the artifactual attributes and preferential attributes can be measured objectively due to their objective tractability, experimental accessibility and pertinence to the imaging system. Artifactual attributes include un-sharpness, graininess, noisiness and other digital artefacts, while preferential attributes refer to image contrast, saturation and colour balance. These two types of attributes occurred from the transmission process, compression process, devices used and other image manipulation activities.

To further illustrate the need to use multiple image attributes, some results, (from Chapter 7) on a study of the JPEG codec Lena image using R_S^2 and Chi2 measures separately on luminance and contrast are shown in Figure 2.5. Figure 2.5(b) suggests that it is safer to make inferences with individual attribute if the Chi2 measure is used. If R_S^2 is to be used (Figure 2.5(a)), the indication is that both luminance and contrast should be considered simultaneously.

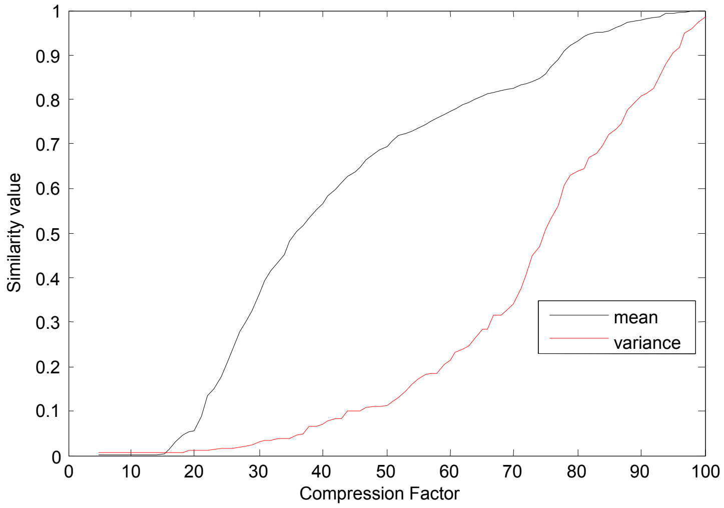


Figure 2.5(a): R_S^2 measured of mean and variance for the compressed Lena image.

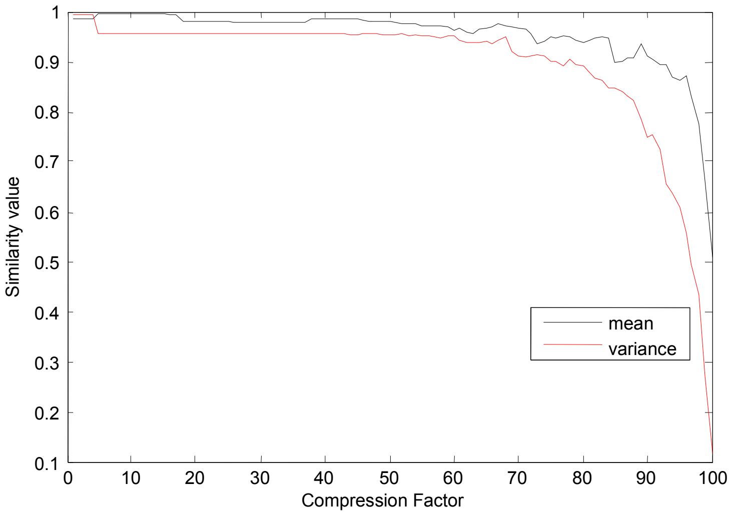


Figure 2.5(b): Chi-square measured of mean and variance for the compressed Lena image.

2.4.3 Issue Three: The Need to Combine Local and Global Image Information

It has been agreed that image quality is a measure that reflects the degree to which the entire image can be successfully exploited by the observer in term of usefulness and naturalness (Janssen, 1999). Furthermore, Keelan (2002) and Burningham et al. (2002) defined an objective image quality measure as a single number that is correlated with this overall perceived attribute of quality, accounting for its viewing conditions and the properties of the human visual system. Traditionally, many of the dissimilarity measures such as PSNR and mean square error (MSE) measure are the global image quality, where the local features are not considered.

More recently the application of localized similarity measures has gained popularity. This is because human eyes usually view images only on a specific part at a given time (Eskiciogu & Fisher, 1995). An image is divided into a set of disjoint windows and its localized features are measured for each disjoint window. Unfortunately, this localized measure failed to display an overall quality for the entire image. In order to overcome this problem, one may calculate the average of the local indexes. One example of this kind of similarity measure is MSSIM. However, this average index is sensitive to extreme local information.

Figure 2.6 explains the effect of extreme local values to the overall image quality. Figure 2.6(a) is the original image. Figure 2.6(b) and Figure 2.6(c) are distorted Lena images by the same Gaussian noise with mean 0 and variance 0.001. However, some extreme values (white area in the red circle) are added to Figure 2.6(c). Two ISMs, namely MSSIM and R_S^2 measured the similarity between the reference image and these distorted images locally with window of size 8×8. The indicated global similarity values are obtained from the average local similarity values for the entire image. It is shown that the global similarity values from both measures dropped from 0.9448 and 0.9719 to 0.9444 and 0.9570, respectively when there is an extreme value in Figure 2.6(c), even though the overall perceived attribute of quality remain good.

Since both global and localized measures are essential for image similarity measure, it is of interest to propose a new similarity measure that is able to reflect the image features locally and globally. A global-localized measure may be defined as an overall quality value obtained from the localized features. Most correlation-based measures are easily adapted to this condition.

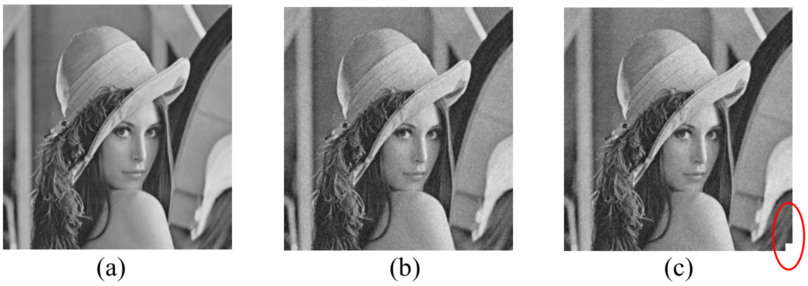

(a) (b) (c)

Figure 2.6: Original Lena image (Left). Distorted Lena image with Gaussian noise *N*(0,0.001) (Middle). Distorted Lena image with Gaussian noise *N*(0,0.001) and extreme values.

2.5 Properties of the Selected SISMs and Their Strengths and Limitations

2.5.1 Properties of the Selected SISMs

(i) Structural Similarity Measure (SSIM) and its related measures

The structural similarity measure proposed by Wang et al. (2002a) is based on the philosophy that the human eyes are the main means to extract structural information from its viewing field. Thus, a measurement of structural distortion should be a good approximation of perceived image distortion (Wang et al., 2002a). SSIM models distortion in a combination of three different factors: the loss of correlation, mean distortion and variance distortion. The quality index is defined as follows:

$$Q(x,y)=\frac{2\mu_x\mu_y}{\mu_x^2+\mu_y^2}\cdot\frac{2\sigma_x\sigma_y}{\sigma_x^2+\sigma_y^2}\cdot\frac{\sigma_{xy}}{\sigma_x\sigma_y}=\frac{4\sigma_{xy}\mu_x\mu_y}{\left(\sigma_x^2+\sigma_y^2\right)\left[\mu_x^2+\mu_y^2\right]} \tag{2.1}$$

The first term models the mean distortion by measuring the closeness between the brightness of the two images. The second term models the variance distortion, which can be defined as the measure of similarity between the contrasts of both images. The last term refers to the linear correlation coefficient (loss of correlation factor) between the two images *X* and *Y* and is of the dynamic range $\left[-1,1\right]$. According to Wang et al. (2004), the MSSIM satisfies the symmetry and reflexivity (the similarity metric has output one for two identical images) conditions, but it has a dynamic range of $\left(-\infty,1\right]$.

The algorithm gives a local measure of quality in the regions of 8×8 pixels to parallel the way human eyes view images. To have an overall image quality measure, one can combine the sum of various local quality indices (Q_j) and average it over *W* windows to acquire a mean structural similarity measure (MSSIM) as given by the formula in (Wang & Bovik, 2002):

$$Q(X,Y)=\frac{1}{W}\sum_{j=1}^{W}Q_j\left(x,y\right) \tag{2.2}$$

However, Wang et al. (2004) noted that the quality index stated in Equation (2.1) is not stable for very low correlation, mean and variance values. To avoid these instabilities, some constant values C_1, C_2 and C_3 are incorporated where

$$Q(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}\cdot\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}\cdot\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{2.3}$$

Wang et al. (2004) proposed $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$ and $C_3 = {}^{C_2}\!/_2$, where L is the dynamic range of the pixel values (255 for 8-bit greyscale images), $K_1 = 0.01$ and $K_2 = 0.03$.

Wang et al. (2004) advocates the use of $Q(x, y)$ due to its simplicity where a statistical function only needs the mean and variance of an image to calculate its quality. Another advantage of it is that it is 'universal', which means to say that it is independent of the image being tested, the viewing condition as well as the individual observers. More importantly, it is applicable to various image processing applications.

The MSSIM is also able to differentiate between the various distortions. If the distortion is related to the structure of the image, the quality given should be lower. On the other hand if the distortion between the images is only in terms of contrast stretching, and the structure is preserved, the quality given to the image should be higher. This is a plus point compared to the Mean Square Error (MSE) (Battaglia, 1996) where picture with similar MSE sometimes have very different perceptual quality.

Despites $Q(x, y)$ has range $(-\infty, 1]$, the main drawback of the MSSIM comes from its pre-defined parameters in which the proposed constant values display inconsistent results for different image problems.

The use of MSSIM expanded to video application in 2002 by the same author Wang et al. (2002b). Let Q_{ij}^{Y}, $Q_{ij}^{C_b}$ and $Q_{ij}^{C_r}$ denote the quality index values of the Y, C_b and C_r components of the j-th sampling window in the i-th video frame, respectively. Wang et al. (2002b) defined the local video quality index by $SSIM_{ij} = 0.8Q_{ij}^{Y} + 0.1Q_{ij}^{C_b} + 0.1Q_{ij}^{C_r}$ and combined them into a frame-level quality index $Q_i = \dfrac{\sum_{j=1}^{R_s} w_{ij} SSIM_{ij}}{\sum_{j=1}^{R_s} w_{ij}}$. Finally, the overall quality of the entire video sequence is given by

$$Q = \frac{\sum_{i=1}^{F} W_i Q_i}{\sum_{i=1}^{F} W_i} \tag{2.4}$$

where F is the number of frames, and w_{ij} and W_i are the corresponding assigned weighting values that will be determined from the video.

The Weighted SSIM (WSSIM) enjoys the advantage from its computational simplicity enabling computation to be done more efficiently and hence, used for real time implementations. It also has effective normalization for various image structures and distortions. However, the WSSIM failed to explain certain HVS characteristics such as why the vertical distortion is more significant than the horizontal distortions. Besides, the weighting of the frames usually does not improve the performance of the measure significantly. Its value differs with the quality measures that take an average of the burst of error, as this may occurs if some frames are badly damaged.

A simple way in which to define fidelity of a colour image is also given by Toet & Lucassen (2003) in 2003. The colour image is transformed into $l\alpha\beta$ colour mode and the MSSIM is applied on each colour band. Then, the colour fidelity metric is given by

$$Q_{color} = \sqrt{w_l\left(Q_l\right)^2 + w_\alpha\left(Q_\alpha\right)^2 + w_\beta\left(Q_\beta\right)^2} \tag{2.5}$$

where w_l, w_α and w_β are weighting values to be determined.

Another two generalisations of MSSIM are Complex SSIM (Alparone et al., 2004) and Edge-dependent fusion quality index (Piella, 2004), both were introduced in 2004. These SISM are used to measure the performance of image fusion method. The former measure derived from the theory of hyper-complex numbers, in particular of 'quaternions' and has the form:

$$Q4 = \frac{\left|\sigma_{z_1 z_2}\right|}{\sigma_{z_1}\sigma_{z_2}} \cdot \frac{2\sigma_{z_1}\sigma_{z_2}}{\sigma_{z_1}^2 + \sigma_{z_2}^2} \cdot \frac{2\left|\overline{z}_1\right|\left|\overline{z}_2\right|}{\left|\overline{z}_1\right|^2 + \left|\overline{z}_2\right|^2} \tag{2.6}$$

where $\boldsymbol{z}_1 = a_1 + \boldsymbol{i}b_1 + \boldsymbol{j}c_1 + \boldsymbol{k}d_1$ and $\boldsymbol{z}_2 = a_2 + \boldsymbol{i}b_2 + \boldsymbol{j}c_2 + \boldsymbol{k}d_2$ are complex numbers. The advantage of using hyper-complex numbers is that both spectral and radiometric distortions, as well as correlation are incorporated. Thus, it is able to show a subtle discrimination capability of spectral distortion and of inaccuracies in spatial enhancement. On the other hand, this may also limited the availability of the measure to images with a number of bands not greater than four.

While, the Edge-dependent fusion quality index (Piella, 2004) takes into account some aspect of the HVS, namely the edge information. The edge images X', Y' and Z' are computed from the original greyscale images *X*, *Y* and their composite image *Z*, respectively. Then the edge-dependent fusion quality index is given by

$$Q_E\left(X,Y,Z\right) = Q_W\left(X,Y,Z\right)^{1-\alpha} \cdot Q_W\left(X',Y',Z'\right)^{\alpha} \tag{2.7}$$

where $Q_W\left(X,Y,Z\right) = \sum_{w \in W} c\left(w\right)\left[\lambda_X\left(w\right)Q\left(X,Z \mid w\right) + \lambda_Y\left(w\right)Q\left(Y,Z \mid w\right)\right]$ is called the weighted fusion quality measure. Note that $c\left(w\right)$ is a weight to be defined, *Q* is the MSSIM and the parameter $\alpha \in \left[0,1\right]$ is the contribution of the edge images compared to the original images. This measure has a dynamic range of $\left[-1,1\right]$ and it takes into account the locations and the magnitude of the distortions. For other type of MSSIM, the computation of the quality value involved different parameters that need to be optimized. This increases the computational complexity of the measure.

Loh & Chang (2006) proposed a modified SSIM that is not only able to show the overall quality of an image, but also show the individual information on luminance, contrast and structural index on a single number. The measure is given by

$$ModSSIM\left(X,Y\right) = A_1^*\left(X,Y\right) + B_1^*\left(X,Y\right) + C_1^*\left(X,Y\right) \tag{2.8}$$

where A_1^* is the rounded value into an integer of $1000\left(1-\frac{2\mu_X\mu_Y}{\mu_X^2+\mu_Y^2}\right)$, B_1^* is the rounded value into 2 decimal point of $\left(1-\frac{2\sigma_X\sigma_Y}{\sigma_X^2+\sigma_Y^2}\right)$, and C_1^* is the rounded value into 4 decimal point of $\frac{1}{100}\left[1-\left(\frac{\sigma_{XY}}{\sigma_X\sigma_Y}\right)^2\right]$. The dynamic range for the ModSSIM is $\left[0.0000,1000.9999\right]$. Index 0.0000 shows that both of the images are identical and 1000.9999 means that the distortion is at maximum level. At the same time, the integer part 0 to 1000 indicates the changes of luminance from the best to the worst conditions. The first two decimal values 0.00 to 0.99 show the changes of contrast from the lowest to the highest degree. Lastly, the last two decimal values 0.0000 to 0.0099 indicate structural information of the images. Note that the modified SSIM cannot apply to local window as the quality index is not a number. Unlike MSSIM and all other generalisation, the ModSSIM is not a symmetrical measure.

There are many more modifications and adaptations of the MSSIM in various image applications. Instead of the pixel mean, variance and correlation, Aja-Fernandez et al. (2006) calculates the statistics on local variance and define the Quality Index based on Local Variance (QILV) as

$$QILV\left(X,Y\right)=\frac{2\mu_{V_X}\mu_{V_Y}}{\mu_{V_X}^2+\mu_{V_Y}^2}\cdot\frac{2\sigma_{V_X}\sigma_{V_Y}}{\sigma_{V_X}^2+\sigma_{V_Y}^2}\cdot\frac{\sigma_{V_XV_Y}}{\sigma_{V_X}\sigma_{V_Y}} \tag{2.9}$$

where μ_{V_X}, μ_{V_Y}, σ_{V_X}, σ_{V_Y} and $\sigma_{V_XV_Y}$ are the respective statistic values of the local variance calculated from image *X* and image *Y*.

Wang & Ma (2008) is the first paper to use MSSIM approach to measure the quality of 3D object. They introduce the weighted sum of three individual distances as

$$D=k_1D_1+k_2D_2+k_3D_3 \tag{2.10}$$

where $k_i > 0$ are constants, D_1 is the kullback-leibler distance, D_2 is the distance for selective wavelet coefficients and $D_3 = \sqrt{1.0 - (MSSIM + 1.0)/2.0}$ in the wavelet domain. On the other hand, Brooks et al. (2008) transformed the image patches *X* and *Y* into complex wavelet coefficients $\boldsymbol{c}_X$ and $\boldsymbol{c}_Y$ and their similarity is measured by the complex wavelet structural similarity (CWSSIM) as

$$CWSSIM\left(\boldsymbol{c}_X, \boldsymbol{c}_Y\right) = \frac{2\left|\sum_{i=1}^{N} \boldsymbol{c}_{X,i}\boldsymbol{c}_{Y,i}^{*}\right| + K}{\sum_{i=1}^{N}\left|\boldsymbol{c}_{X,i}\right|^2 + \sum_{i=1}^{N}\left|\boldsymbol{c}_{Y,i}\right|^2 + K} \tag{2.11}$$

where *K* is a small positive constant set to 0.03.

In many video applications, motion compensated from interpolation is adopted to improve video quality by increasing the frame rate (Yang et al, 2008). In such, the Perceptual Frame Interpolation Quality Metric (PFIQM) based on MSSIM is used to assess the spatial quality degradation from frame interpolation. Details of the PFIQM are available in Yang et al. (2008). Main disadvantage of this metric is that prior knowledge about the frame interpolation is needed such as type of artifacts, possible regions of quality degradation and the occurrence of highly conspicuous local distortion.

Most ISM has been developed to assess the quality of monochrome images. They are not suitable for multiband images such as hyperspectral remote sensing images. Hence, Garzelli & Nencini (2009) introduces the measure

$$Q2^n = E\left[\left|\boldsymbol{Q2}^{\boldsymbol{n}}_{N\times N}\right|\right] \tag{2.12}$$

by averaging the magnitudes of all $\boldsymbol{Q2}^{\boldsymbol{n}}_{N\times N} = \frac{\sigma_{z,v}}{\sigma_z\sigma_v} \cdot \frac{2\overline{\boldsymbol{z}}\overline{\boldsymbol{v}}}{\overline{\boldsymbol{z}}^2 + \overline{\boldsymbol{v}}^2} \cdot \frac{2\sigma_z\sigma_v}{\sigma_z^2 + \sigma_v^2}$.

Krishna Moorthy & Bovik (2009) introduces the Combined Percentile and Fixation Based SSIM (PF-SSIM), in which the values of $F - SSIM(X,Y) = \frac{\sum_{i=1}^{P}\sum_{j=1}^{Q} w_{ij} SSIM\left(X_{ij}, Y_{ij}\right)}{\sum_{i=1}^{P}\sum_{j=1}^{Q} w_{ij}}$ are sorted and weighted by the procedure P-SSIM where P, Q are the image dimensions and w_{ij} are the SSIM weights.

Lastly, Wirandi et al. (2009) adapted the SSIM into a neural network system when human assessment is involved. This SSIM based neural network is useful in handling the quantitative and qualitative (subjective) factors. However, this method requires a large training set.

(ii) Weighted Mean Variance (WMV)

The second moment's method has been proposed by Manjunath & Ma (1996) in 1996. It is defined by

$$D^r\left(X,Y\right)=\frac{\left|\mu_r\left(X\right)-\mu_r\left(Y\right)\right|}{\left|\sigma\left(\mu_r\right)\right|}+\frac{\left|\sigma_r\left(X\right)-\sigma_r\left(Y\right)\right|}{\left|\sigma\left(\sigma_r\right)\right|} \tag{2.13}$$

which is called Weighted Mean Variance. The empirical means $\mu_r\left(X\right)$, $\mu_r\left(Y\right)$ and standard deviations $\sigma_r\left(X\right)$, $\sigma_r\left(Y\right)$ are applicable only to marginal distributions along channel r. The overall dissimilarity can be obtained by combining marginal value from all channels. Although its range is not $\left[0,1\right]$, but the dissimilarity measure have the advantages of symmetrical and less computational complexity. It has been applied to texture-based image retrieval.

(iii) Scaling Similarity (SS)

The scaling similarity was originally proposed by Yang (1998) for texture retrieval. It is defined on a set of scaling features

$$Q=\sum W_1\frac{\left|\Delta s\right|}{\max\left(\left|\Delta s\right|\right)}+\sum W_2\frac{\left|\Delta\Omega\right|}{\max\left(\left|\Delta\Omega\right|\right)} \tag{2.14}$$

where $\Delta\Omega$ is the difference of a 2-component vector consisting of mean and standard deviation, Δs is the difference of the scaling features of 2 images, and $\max\left(\cdot\right)$ is the maximum differences of the respective features in the whole test database. The measure has many useful properties such as it is invariant under scaling, translation and rotation of image lattice and it is robust against random uncorrelated noise. The noise has very small effect on the models at higher levels of the scale-downed images. Some parameters $\left(W_1, W_2, \alpha\right)$ are required prior to computation processes. Moreover, the scaling processes caused higher computational complexity. The quality value can only be computed after the whole test database has been considered.

(iv) Percentile blob-based Similarity (PBSIM)

The Percentile blob-based similarity (Howe, 1998) is defined by

$$D\left(X,Y\right)=\sum_{i=1}^{36}\left|f_i\left(X\right)-f_i\left(Y\right)\right|+\sum_{i=37}^{44}\left[\min_{b_1\in f_i(X),b_2\in f_i(Y)}\Delta\left(b_1,b_2\right)\right] \tag{2.15}$$

where b_1, b_2 refer to blob descriptive statistics such as mean, standard deviation, skewness, kurtosis, percentile and range of the compared images, $f_i\left(X\right)$ and $f_i\left(Y\right)$ are the simple numeric feature value of the respective image. The measure provides a consistent result and not depends on the number of images being considered. Thus, it is good for rapid search through large sets of natural scenes. On the other hand, there are some drawbacks on this measure; (i) it is reliable only if the differences between the categories are large, (ii) it is less sensitive to a bad match on any single feature which will result in a certain types of similarity not detected, and (iii) it is not suitable for every types of similarity problems such as face recognition and scenes that are cluttered with many small objects.

(v) Normalized Correlation (NC)

Normalized correlation is the first statistical correlation used to measure image quality. It was proposed by Steinberg (1987) using quantized aperture data in a microwave imaging system. The normalized correlation coefficient between the reference image $\left(I_f\right)$ and the distorted image $\left(I_g\right)$ is

$$\rho = \frac{\overline{I}_f \overline{I}_g}{I_{frms} I_{grms}} = \frac{\left|\overline{F(f)}\right|\left|\overline{F(g)}\right|}{\left[F(f)\right]_{rms}\left[F(g)\right]_{rms}} \tag{2.16}$$

where the overbar means average over all pixels in the images, *rms* represents root mean square, and $F\left(f\right)$ and $F\left(g\right)$ denote the Fourier Transform for the proper signal (*f*) and the quantized signal (*g*), respectively.

Given two independently identically distributed random variables and the number of pixels is large enough, Liang et al. (1989) has shown that the normalized correlation coefficient is uniquely specified when the data distribution is known. It is well preserved with highly compressed aperture data in which 80% of image fidelity is achieved with only 1 bit of aperture data (Liang et al., 1989). Although the normalized correlation coefficient is nearly scene-independent for a well scene classification, the theory is also applicable to other type of images and it is insensitive to variations in the distribution of phase.

Stevens (2001) has also discussed a normalized correlation using Pearson's-R statistic in 2001 for three-dimensional scene interpretation. It is defined as

$$\varepsilon_{correlation}\left(f_s, f_p\right) = 1.0 - \frac{\left(R\left(f_s f_p\right)+1\right)}{2} \tag{2.17}$$

where $R\left(f_s, f_p\right) = \frac{1}{N-1}\left[\sum_{j}^{N}\left(\frac{f_s(j)-\overline{f}_s}{\sigma_{f_s}}\right)\left(\frac{f_p(j)-\overline{f}_p}{\sigma_{f_p}}\right)\right]$ is the Pearson correlation. This measure preserves negative relationship of the two feature vectors f_s and f_p corresponding to sensor input image and output image respectively. In which, $\varepsilon \in \left(0.5, 1\right]$ indicates negative relationship, $\varepsilon \in \left[0, 0.5\right)$ denotes positive relationship and $\varepsilon = \left\{0\right\}$ when the two images has no relationship. However there is a pitfall in the case of multiple bands of data, i.e. RGB colour image. Problem can happen when computing the mean and standard deviation of the feature vectors for each band separately. For instance, take a region in an image that is mostly white and a mostly red region. The mean value for each band in both white and red regions would be shifted toward zero independently. This is not the case for the mean value over all three bands. Furthermore, it is known that different feature vectors re-act differently for certain image configuration.

(vi) Squared Pearson Correlation (Rs2 or R_S^2)

Pearson correlation has been widely used in measuring the relationship between subjective image quality and human observer performance. However, its application as an objective quality metric only started in 1997 where Nielsen et al. (1997) used it to detect the different impairment in a mobile radio channel signal processing. R_S^2 defined as follows,

$$R_S^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} = \frac{\left[\sum(x-\bar{x})(y-\bar{y})\right]^2}{\sum(x-\bar{x})^2\sum(y-\bar{y})^2} \tag{2.18}$$

can be used to detect the noise and time dispersion introduced by the mobile radio channel and a possible timing error in the receiver (Nielsen et al., 1997). It is also used to evaluate the shape of the discriminator output compared to the ideal output signal. However, Cramariuc et al. (2000) pointed out that this measure is sensitive to outliers and nonlinear increasing transformations.

Instead of calculating a single correlation value for the entire image, Robert et al. (1999) computes the Pearson correlation for each pixel location throughout the image. The correlation coefficient at a pixel is calculated based on its neighbourhood pixels. These localized correlation coefficients provide a correlation map, which displays visually the locations in the image where quality is the best and the worst. Then a global measure is calculated as an average of the localized correlation (C_{avg}) values

$$C_{avg} = \frac{1}{MN} \sum_{i} \sum_{j} R^2(i, j) \tag{2.19}$$

where $R^2(i, j)$ is the correlation coefficient at pixel location (i, j) over an image of size $M \times N$.

In 2004, Ivkovic & Sankar (2004) proposed a new algorithm to improve the average Pearson correlation for image quality assessment. The new measure takes into account of two HVS properties: (i) nonlinear relationship between intensity and perceived brightness, and (ii) presence of spatial filtering in HVS. It is defined as:

$$Q = \text{sign}\left\{\rho_{xy\_avg}\right\} \left|\rho_{xy\_avg}\right|^{f\left(\rho_{xy\_avg}\right)} \tag{2.20}$$

where $f\left(\rho_{xy\_avg}\right) = 1.2 + 0.5 \tanh\left(\frac{\left|\rho_{xy\_avg}\right| - 0.3}{0.15}\right)$ and ρ_{xy_avg} is the average of the localized Pearson's correlation. The quality measure Q is closer to human observer with localized measure and it is able to differentiate between random distortion and signal-dependent distortion, which have different effects on human observer. Random distortion presence if the magnitude of average correlation coefficient is close to zero, and it indicates signal dependent noise if the coefficient is close to one.

(vii) Probability of Error (PE)

Apart from the number of objects, the performance of image segmentation procedures can also be measured based on the number of mis-segmented pixels. Under the assumption that the image consists of objects and background each having a specified distribution of grey level, one can compute the probability of misclassifying an object pixel as background, or vice versa (Zhang, 1996). Based on this principal, Lee et al. (1990) has defined a PE for two-class problem by

$$PE = P(O)\times P(B\,|\,O) + P(B)\times P(O\,|\,B) = 2P(O\cap B) \tag{2.21}$$

where $P(O)$ and $P(B)$ are *a priori* probabilities of object and background in images, respectively. $P(B\,|\,O)$ is the probability of error in classifying objects as background and $P(O\,|\,B)$ is the probability of error in classifying background as objects. Although PE is relatively general for different types of segmentation algorithm, there is no quantitative measure for *a prior*i knowledge about images that can be incorporated into segmentation algorithms. So, various types of knowledge can hardly to be computed.

Some efforts to study the relations between various similarity functions that applied to image or video retrieval problem have also been carried out by Vasconcelos & Lippman (2000). A content-based retrieval system tries to minimize the probability of retrieving error, $P(g(\boldsymbol{x})\neq y)$ which means the probability of a set of feature vectors $\boldsymbol{x}$ drawn from class y has been classified as an image from a class $g(\boldsymbol{x})$ different than y. This optimal map formulation is best known as Bayes classifier.

$$g^*(\boldsymbol{x}) = \arg\max_i P(\boldsymbol{x}\,|\,y=i)P(y=i)$$

where $P(\boldsymbol{x}\,|\,y=i)$ is the likelihood function for the ith class and $P(y=i)$ is the prior probability. Meanwhile, Vasconcelos & Lippman (2000) pointed that the smallest achievable probability of error is the Bayes error (Comaniciu et al., 1999)

$$L^* = 1 - E_{\boldsymbol{x}}\left[\max_i P(y=i\,|\,\boldsymbol{x})\right] \tag{2.22}$$

Vasconcelos & Lippman (2000) also demonstrated that most of the similarity functions applied to image retrieval are special cases of this Bayes error. If an upper bound on the Bayes error of a collection of two-way classification problems is minimized instead of the probability of error of the original problem, then Vasconcelos & Lippman (2000) shown that the Bayesian criteria reduces to the Bhattacharyya distance (BD) (Comaniciu et al., 1999)

$$g(\boldsymbol{x}) = \arg\max_{i} \int \sqrt{P(\boldsymbol{x} \mid q) P(\boldsymbol{x} \mid y = i)} \qquad (2.23)$$

where $P(\boldsymbol{x} \mid q)$ is the density of the query. Equation (2.23) finds the lowest upper bound on the Bayes error for the collection of two-class problems involving the query and each of the database classes.

On the other hand, the Bayes error can be reduced to the Maximum likelihood (ML) method if the different image classes are uniformly distributed when the original criterion is minimized. The ML takes the form

$$g(\boldsymbol{x}) = \arg\max_{i} \frac{1}{N} \sum_{j=1}^{N} \log P(\boldsymbol{x}_j \mid y = i) \qquad (2.24)$$

when the query consists of a collection of N independent query features $\boldsymbol{x} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N\}$. Equation (2.24) can be further reduced to Kullback-Leibler divergence (KLD) method when the number of query feature N is large.

$$g(\boldsymbol{x}) \xrightarrow{N \to \infty} \arg\min_{i} E_q\left[\log P(\boldsymbol{x}_j \mid y = i)\right] = \arg\min_{i} KL(Q \parallel P_i) \qquad (2.25)$$

where the $KL(Q \parallel P_i)$ is the KLD between the query density and that associated with the *i*th database image class. Both ML and KLD metrics perform equally well for global queries based on entire images (Vasconcelos & Lippman, 2000). However, the ML method allows the above idea to be applied to the subset of the retrieved image. On the other hand, the KLD has a closed-form expression and it has smaller computational complexity compared with ML.

The KLD can be approximated by using a first order Taylor series approximation for the logarithm function at $x = 1$, $\log(x) \approx x - 1$. This yields the χ^2 statistic proposed by (Schiele & Crowley, 1996; Comanicui et al., 1999)

$$KL(Q \parallel P_i) \approx \int \frac{\left[P(\boldsymbol{x} \mid q) - P(\boldsymbol{x} \mid y = i)\right]^2}{P(\boldsymbol{x} \mid y = i)} dx \qquad (2.26)$$

It is noted that the χ^2 statistic may not be able to perform as good as ML.

Alternatively, if the KLD has a Gaussian probability density and the orthonormal covariance matrices, then it becomes

$$g(\boldsymbol{x}) = \arg\max_i \log|\Sigma_i| + L_i' \tag{2.27}$$

Equation (2.27) is called the Quadratic distance (QD) if $L_i' = \frac{1}{N}\sum_n (\boldsymbol{x}_n - \mu_i)^T \Sigma_i^{-1} (\boldsymbol{x}_n - \mu_i)$ and Mahalanobis distance (MD) (Santani & Jain, 1999) if $L_i' = trace\left[\Sigma_i^{-1}\hat{\Sigma}_x\right] + M_i$ where $\hat{\Sigma}_x$ is the sample covariance matrix of $\boldsymbol{x}_n$ and $M_i = (\hat{\boldsymbol{x}}_i - \mu_i)^T \Sigma_i (\hat{\boldsymbol{x}}_i - \mu_i)$. Both QD and MD only make sense if the image features for all classes are Gaussian distributed. Furthermore, the MD is not robust to certain transformations which involve rotation and scaling.

In 2003, Goldberger et al. (2003) extended the KLD to the image retrieval problem with mixture of two Gaussians. The unscented approximation of the KL-divergence for the case of mixture of Gaussians is as follow

$$KL_{match}(f \,||\, g) = \sum_{i=1}^{n} \alpha_i \left(KL\left(f_i \,||\, g_{\pi(i)}\right) + \log\frac{\alpha_i}{\beta_{\pi(i)}} \right) \tag{2.28}$$

where $KL\left(f_i \,||\, g_{\pi(i)}\right) = \frac{1}{2}\left(\log\frac{|\Sigma_2|}{|\Sigma_1|} + trace\left(\Sigma_2^{-1}\Sigma_1\right) + (\mu_1 - \mu_2)^T \Sigma_2^{-1} (\mu_1 - \mu_2) \right)$ is the KL-divergence between the two mixture of Gaussians $f = \sum_{i=1}^{n} \alpha_i N\left(\mu_{1,i}, \Sigma_{1,i}\right)$ and $g = \sum_{j=1}^{n} \beta_j N\left(\mu_{2,j}, \Sigma_{2,j}\right)$. The experiments show that this similarity measure produces results that are very close to large sample Monte-Carlo based ground truth.

(viii) Mutual Information (MI)

The MI measure proposed by Girod (1981) in 1981 is one of the earliest statistical based ISM. The MI measure is motivated by the mutual information rate between two jointly Gaussian processes in the form

$$MI = -\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \ln\left(1 - \frac{\left|\Phi_{vu}\left(\omega_1,\omega_2\right)\right|^2}{\Phi_v\left(\omega_1,\omega_2\right)\Phi_u\left(\omega_1,\omega_2\right)}\right) d\omega_1 d\omega_2 \tag{2.29}$$

where Φ_{vu} is the cross spectrum between *u* and *v*. Where as Φ_v and Φ_u are the power spectra of *v* and *u*, respectively for the two dimensions ω_1 and ω_2 of digital frequency.

Instead, Girod (1981) has shown that the ratio $\frac{\ln\left(MI\right)}{\ln\left(MSE\right)}$ performs about as well as a single human observer with a very good robustness.

Another quality measure originated from mutual information concept is the Mutual Information Similarity (MIS) measure. An application of the MIS is given in Chen et al. (2003) in a remote sensing problem. They defined MIS of two random variables *A* and *B* as

$$I\left(A,B\right) = H\left(A\right) + H\left(B\right) - H\left(A,B\right) \tag{2.30}$$

where $H\left(A\right) = \sum_a -P_A\left(a\right)\log P_A\left(a\right)$, $H\left(B\right) = \sum_b -P_B\left(b\right)\log P_B\left(b\right)$ are the entropies of *A* and *B*, and $H\left(A,B\right) = \sum_{a,b} -P_{A,B}\left(a,b\right)\log P_{A,B}\left(a,b\right)$ is their joint entropy. The MIS measure provides consistent result on multimodal registration problems. It is relatively simple to understand and compute with only information on image histogram required.

(ix) Information Theory Divergence

Rubner et al. (2001) investigated two special cases of information theory divergence. The first divergence is the Kullback-Leibler divergence (Vidal et al., 1998) discussed in Section 2.5.7 but now it is defined as

$$D(X,Y)=\sum_i f(i:X)\log\frac{f(i:X)}{f(i:Y)} \tag{2.31}$$

where $f(i:X)$ and $f(i:Y)$ are the histogram entry of images *X* and *Y*, which correspond to the number of image pixels in bin *i*. This KL divergence measures how inefficient an average it would be to code on histogram using the other as the true distribution for coding. Its drawback is that the KL divergence becomes infinite if $f(i:Y)$ is smaller than $f(i:X)$.

The second divergence is called the Jeffrey divergence (JD) or Jensen-Shannon divergence. It is a modified version from KL divergence and is defined by

$$D(X,Y)=\sum_i f(i;X)\log\frac{f(i;X)}{\hat{f}(i)}+f(i;Y)\log\frac{f(i;Y)}{\hat{f}(i)}$$

where $\hat{f}(i)=\left[f(i:X)+f(i:Y)\right]/2$ is the mean histogram. In contrast to the KL divergence, JD is symmetric and numerically stable when comparing two empirical distributions.

(x) Kolmogorov Simirnov distance (KS)

The KS distance was originally introduced in Geman (1990) for image segmentation. The empirical cumulative distribution for the image *X* and image *Y* is obtained from their corresponding image histogram. Then the maxima discrepancy between these cumulative distributions along channel (colour band) *r* is defined as

$$D^r(X,Y)=\max_i\left|F^r(i;X)-F^r(i;Y)\right| \tag{2.32}$$

Another similar nonparametric distance measure is the statistic of the Cramer von Mises (CvM). Instead of the maxima discrepancy, CvM is defined as the squared Euclidean distance between the two cumulative distributions

$$D^r(X,Y)=\sum_i\left[F^r(i;X)-F^r(i;Y)\right]^2 \tag{2.33}$$

Equation (2.32) and Equation (2.33) give a symmetric and invariant measure to arbitrary monotonic feature transformation in one dimension (Rubner et al., 2001). Both of them have low computation cost based only on the number of histogram bins. However, the CvM is only suitable for image that has high contrast, but not consistent for the low contrast image.

(xi) Chi-squared statistic (Chi2)

Puzicha (1997) investigates the correspondence between two images in image segmentation and retrieval problems. The empirical distributions of two images *X* and *Y* are obtained directly from their image histogram. Then the Chi-squared statistic computed by Puzicha is

$$D(X,Y)=\sum_i \frac{\left[f(i;X)-\hat{f}(i)\right]^2}{\hat{f}(i)} \qquad (2.34)$$

The advantage of Chi-squared statistic is that it is applicable to multidimensional histograms.

2.5.2 Summary of the Selected SISMs

Properties for some selected SISMs are discussed and compared. These selected SISMs are MSSIM, WSSIM, Q_{color}, WMV, SS, PBSIM, NC, R_S^2, C_{avg}, PE, PID, MIS, JD, KS, Chi2 and R_F^2 (or Rf2, see Chapter 3). These SISMs were chosen to represent different statistical approaches such as moments-based (MSSIM, WSSIM, Q_{color}, WMV, SS and PBSIM), correlation-based (NC, R_S^2, C_{avg} and R_F^2), probability-based (PE), information theory (PID, MIS and JD), and finally nonparametric ideas (KS and Chi2).

The success and limitation of these selected SISMs to cope with the three issues mentioned in Section 2.4.1 to Section 2.4.3 are summarized in Table 2.4. It is shown that none of the SISMs in this survey is capable of handling all of the three image problems simultaneously. From the selected SISMs, only Q_{color} and R_F^2 considered the reference image as random or subject to errors. All other SISMs assume the reference image has a perfect quality. Generally, the moment-based SISMs use two image attributes and which usually are image luminance and image contrast values. The only two measures that use more than two image attributes are PBSIM and PID. However, correlation-based, nonparametric-based, probability-based and measure based on information theory generally use a single image attribute. On the other hand, most of the SISMs measure the image similarity either globally or use local information. Only two measures in the literature review, Weighted SSIM (WSSIM) and Q_{color} combine both local and global information into a single measure.

Table 2.4: Properties of the selected SISMs. Y=yes, N=no. S=single attribute is used, B=bivariate attributes are used, M=multiple attributes are used. L=local measure, G=global measure.

SISM	Perfect Reference	# of image attributes	Local or Global	Dynamic Range	Symmetry	Pre-defined value
MSSIM	Y	B	L	$(-\infty,1]$	Y	Y
WSSIM	Y	B	L/G	$(-\infty,1]$	N	Y
Q_{color}	Y	B	L/G	$[0,\infty)$	N	Y
WMV	Y	B	G	$[0,\infty)$	Y	N
SS	Y	B	G	$[0,1]$	Y	Y
PBSIM	Y	M	G	$[0,\infty)$	Y	N
NC	Y	S	G	$[0,1]$	Y	Y
Rs2	Y	S	G	$[0,1]$	Y	N
C_{avg}	Y	S	L	$[0,1]$	Y	N
Rf2	N	S	G	$[0,1]$	N	N
PE	Y	S	G	$[0,1]$	Y	Y
PID	Y	M	G	$[0,\infty)$	N	N
MIS	N	S	G	$[0,\infty)$	Y	N
JD	Y	S	G	$[0,\infty)$	Y	N
KS	Y	S	G	$[0,\infty)$	Y	N
Chi2	Y	S	G	$[0,\infty)$	N	N

Other properties of SISMs were also discussed. It showed that most of the selected SISMs defined on the ranges $\left[0,1\right]$ and $\left[0,\infty\right)$ except MSSIM on $\left(-\infty,1\right]$. The range $\left[0,1\right]$ is preferable because it is easier for interpretation purposes. There are four non-symmetrical SISMs, namely Q_{color} , R_F^2 , PID and Chi2. This non-symmetrical property does not affect their performance but it may cause inconvenience only when the reference image and the distorted image are not easily differentiable, which is rare in practice. Most SISMs do not required pre-defined values to calculate the quality index except MSSIM. The need for pre-defined values will not obstruct the accuracy of the measure, but it will degrade the usefulness of the measure.

In summary, the survey carried out and the discussions that followed provide strong evident for the use of R_F^2 and R_P^2 as a measure of similarity (or quality) between two images. The survey showed that R_P^2 has not been used for this purpose before. An important observation from the survey was that none of the SISMs considered were capable of handling the three important image issues stated, and the later chapters will show that R_P^2 can handle all this problems. Finally, properties of R_P^2 suggest that it is potentially convenience measure to be used.