Journal of Electronic Imaging

JElectronicImaging.org

Video quality assessment based on motion structure partition similarity of spatiotemporal slice images

Peng Yan
Xuanqin Mou

Peng Yan, Xuanqin Mou, "Video quality assessment based on motion structure partition similarity of spatiotemporal slice images," *J. Electron. Imaging* **27**(3), 033019 (2018), doi: 10.1117/1.JEI.27.3.033019.

Video quality assessment based on motion structure partition similarity of spatiotemporal slice images

Peng Yan and Xuanqin Mou*
Xi'an Jiaotong University, Institute of Image Processing and Pattern Recognition, Xi'an, China

Abstract. Video quality assessment (VQA) is becoming increasingly important as a comprehensive measure of video quality. This paper proposes a full-reference VQA (FR-VQA) algorithm based on the motion structure partition similarity of spatiotemporal slice (STS) images. To achieve this objective, a number of FR-image quality assessment algorithms were applied slice by slice to video STS images to compare their performance of detecting structure similarity of STS images. The algorithm that performed the best was selected to detect the similarity between motion-partitioning STS images. Next, as motion objects in the video sequence were found to have different influences on the prediction performance in terms of moving speed and track, the STS images were divided into simple and complex motion regions, and their contributions to the VQA task determined. Consequently, a promising effective and efficient VQA model, called STS-MSPS, is also proposed. Experimental evaluations conducted based on various annotated VQA databases indicate that the proposed STS-MSPS achieves state-of-the-art prediction performances in terms of correlations with subjective evaluation and statistical significance tests. This paper also shows that STS images by themselves provide sufficient information for VQA tasks and that the proposed complex motion region of an STS image is predominantly responsible for yielding a high-precision model. *© 2018 SPIE and IS&T* [DOI: 10.1117/1.JEI.27.3.033019]

Keywords: full reference; motion structure partition similarity; video quality assessment; video spatiotemporal slice images.
Paper 170659 received Aug. 25, 2017; accepted for publication May 7, 2018; published online May 23, 2018.

1 Introduction

As videos become ubiquitous, consumers are becoming considerably more particular about video quality. Consequently, video quality assessment (VQA) is becoming increasingly important as a comprehensive measure of the video quality being provided to consumers. In general, VQA involves two methodologies: subjective VQA and objective VQA. Subjective VQA requires human participants, which makes it time consuming, laborious, and inefficient, but precise in representing the perceptual quality of the human vision system (HVS). Hence, it is usually employed as the benchmark to evaluate objective metrics. Objective VQA evaluates the perceptual video quality automatically and accurately using algorithms. Based on the extent of the reference video information required, the type of objective VQA used to evaluate the perceptual quality of the distorted counterpart of the reference video is typically divided into three categories: full-reference (FR), reduced-reference (RR), and no-reference (NR) VQA. FR-VQA accesses the entire reference video, RR-VQA uses partial reference video feature information, and NR-VQA evaluates the perceptual quality without any reference video information. Because FR-VQA obtains the reference video, more fruitful FR research achievements have emerged compared with RR-VQA and NR-VQA.

The simplest and most widely used FR evaluation metrics are mean square error (MSE) and peak signal-to-noise ratio (PSNR). MSE does not consider the structural information hidden in the neighboring pixels of an image or a video; hence, it cannot satisfactorily reflect perceptual quality when used for evaluation applications. With the development of image quality assessment (IQA), numerous advanced evaluation metrics that consider various perceptual properties of HVS or structural features that account for natural scene statistics have been proposed. Because HVS has evolved to understand natural scenes, both HVS models and visual structure features have become efficient in the design of evaluation models for perceptual quality.[1–12] For example, visual signal-to-noise ratio,[1] most apparent distortion (MAD),[2] visual information fidelity (VIF),[3] video quality model,[4] and motion-based video integrity evaluation (MOVIE)[5] are HVS-based IQA or VQA methods, whereas structural similarity (SSIM),[6] edge-based SSIM (ESSIM),[7] feature similarity (FSIM),[8] multiscale SSIM (MS-SSIM),[9] Riesz-transform-based feature similarity (RFSIM),[10] gradient magnitude similarity deviation (GMSD),[11] and three-component SSIM (3-SSIM)[12] are structure-based IQA or VQA methods. However, owing to the considerable developments in the field of advanced visual signal processing in recent decades, more approaches based on visual structure features than those based on HVS models have been proposed in evaluation model design, considering that at present, it is still difficult to exactly understand how HVS perceives the visual quality of a natural scene.[6–12]

The perceptual quality evaluation model design based on visual structure features typically requires three steps to create the evaluation model: visual structure feature extraction, point-based error measurement based on the extracted

*Address all correspondence to: Xuanqin Mou, E-mail: xqmou@mail.xjtu.edu.cn

features, and error pooling over the entire visual scene.[6–12] For example, the SSIM algorithm,[6] which assumes that HVS tends to perceive local structures in natural scenes, has achieved considerably higher perceived relevance than PSNR and MSE. First, SSIM extracts the luminance, contrast, and structural features from images. Then, it obtains the feature similarity map via point-by-point comparison of the extracted feature similarities between the reference and tested images. Finally, it applies the mean pooling method over the similarity maps to obtain an integrated measurement of perceptual quality. SSIM was widely successful in IQA; subsequently, Wang et al.[13] were the first to extend its use to VQA by applying perceptual weights to local regions and whole frames. The three-SSIM algorithm,[12] which further considers the impact of image content, was subsequently proposed. The algorithm divides the content of video frames into edges, textures, and smooth areas and assigns different weights to these three areas according to the different perception sensitivities of human vision to different contents.

The algorithms outlined above show that in the development of an evaluation model based on visual structure features, feature design and extraction are key problems. Recently, features extracted from spatiotemporal slice (STS) images have been exploited in video parsing applications. In the VQA model design, spatiotemporal MAD (ST-MAD)[14] and its improved and extended version ViS3[15] adopted a combination of distortion measures of image frames and similarity measures of STS images between the reference and distortion videos. In our study, we found that STS images alone contain sufficient information for advanced VQA model design. Based on this finding, we propose an FR-VQA metric called STS-MSPS that produces promising quality prediction accuracy with a simple computation framework.

The rest of the paper is organized as follows: Sec. 2 provides a brief review of related work. Section 3 describes the proposed method. Section 4 discusses the experimental evaluations conducted and the results obtained. Finally, Sec. 5 presents concluding remarks.

2 Related Work and Our Contribution

STS images, which are generated by the division of three-dimensional (3-D) video data along the time axis, contain rich temporal information on a large temporal scale. Consequently, they are used for a wide range of applications, such as video partitioning,[16] video motion analysis and segmentation,[17] as well as video shot clustering and retrieval.[18] In addition, STS images provide a method for extracting visual structure features for evaluation of model design. For example, ViS3, which was proposed by Vu and Chandler[15] mainly computes the spatiotemporal distortion by detecting the similarity of STS images. More specifically, ViS3 first computes the spatial distortion using the improved MAD[2] algorithm to process video frames. Then, it computes the spatiotemporal distortion by detecting the similarity of STS maps and checking their differences using an HVS model. Finally, the two types of distortions are combined into one video distortion index. Although ViS3 achieves high evaluation relevance because of the use of MAD[2] and the HVS models, it has relatively high computational complexity.

Video STS images include trajectory information about the objects in a video over a large timescale. In the past, VQA algorithms failed to introduce a better temporal distortion calculation method than simple temporal distortion calculation methods, such as frame quality average methods,[12,19] motion-level weighted methods,[13,20] and frame quality pooling methods in small timescales.[21] The ViS3 algorithm provides video quality evaluation on a large timescale using STS images. In addition, it captures distortion in STS images using the HVS model. Because MAD and the HVS models are used in the model design, ViS3 is more complicated. Therefore, in our earlier research, we proposed a simple and efficient VQA framework, called SSTS-GMSD/STS-GMSD,[22] to compute video STS image similarity. The starting point of this framework is detection of the motion structural similarity of original and test video STS images using an excellent structural similarity algorithm. As the GMSD[11] algorithm exhibits superior performance in structural distortion detection, we used the GMSD similarity of STS images for video quality evaluation. Both the SSTS-GMSD and STS-GMSD algorithms are better than other algorithms when tested on the laboratory for image and video engineering (LIVE) VQA database; further, the calculation is simple and fast.

Considering the good performance of STS-based VQA algorithms, we further studied STS images and found that they mainly contain two different movement structure regions, i.e., simple and complex motion regions, in association with quality prediction. We subsequently conducted a series of experiments to determine the influence of these regions on quality assessment. Then, we proposed a promising FR-VQA called STS-MSPS based on the CMSD measures of the two STS image regions between the reference and distortion videos. In general, the contributions of this paper are twofold: (1) it shows that STS images alone can effectively predict the perceptual quality of video sequences, and accordingly leads an efficient VQA metric by introducing the GMS metric to measure the difference between the reference and distortion STS images. The proposed metric performs as one of the best state-of-the-art VQA metrics in terms of prediction accuracy. (2) By exploring the influence of movements of objects in image sequence on video quality evaluation, we divide the STS images into simple and complex motion areas based on the speed and track factors of motion objects, and show that the complex motion areas contain most of the perceptual quality information associated with a video sequence.

3 Methods

3.1 *Motion Structure Areas of STS Images*

Video slice images can be of various types. Different slice images can be generated using different cutting methods. In this paper, only two slice types are considered: slices perpendicular to the time axis (video frames) and slices parallel to the time axis (STS images). With regard to STS images, we study only two types: vertical and horizontal STS images. Vertical STS images are constructed using the same columns of a video, whereas horizontal STS images are composed of the same rows of a video. From a video frame, scene content can be observed, whereas from an STS image, motion structure images constructed by pixels at the same row/column can be observed at different

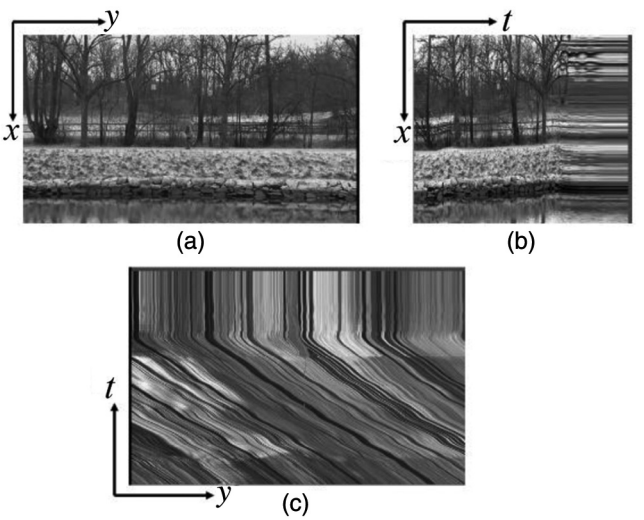

Fig. 1 Example of slice images of video from the LIVE VQA database (a) frame, (b) vertical STS image, and (c) horizontal STS image.

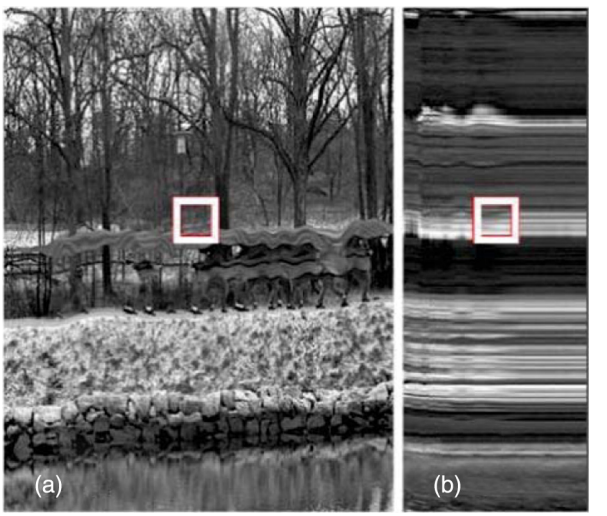

Fig. 3 Segmentation of STS region. (a) The complex motion area, whereas (b) the simple motion area. This image is a vertical STS image from the video called "pr1_50fps.yuv" in the LIVE VQA database.

times. Figure 1 shows an example of a frame, a vertical STS image, and a horizontal STS image.

From the STS images, structure information (edges and textures), motion information (track tilt), etc. can be observed. Different moving objects have different motion tracks in the STS images, and edges are present between the tracks of different objects. Using the edges of objects in STS images for segmentation, various targets can be obtained. Because of these characteristics of STS images, they have been used in video parsing applications, such as video partitioning,[16] video motion segmentation,[17] and video shot clustering and retrieval.[18]

STS images contain motion information over a large time scale, and they have structured content characteristics. Therefore, we theorized that the structural similarity of STS images is sufficient for VQA. Accordingly, a VQA method based on the structural similarity of STS images was proposed in our earlier work.[22] The detailed framework of this method is shown in Fig. 2. In further research and experiments, we found that STS images mainly have two motion structure areas. Therefore, in the present study, we developed an FR-VQA method based on the motion structure partitions of STS images and attempted to determine the influences of these motion structure areas on VQA. The proposed method considers not only the impact of motion structure information on visual perception but also the fact that the human eye is more sensitive to edges and distortion.

The human eye's sensitivity to the distortion in videos is related to the movement intensity of objects. Considering the influences of the movement speed of objects for video quality evaluation, video STS images can be divided into three parts: static areas, uniform speed areas, and complex motion areas. In STS images, static areas are characterized by horizontal or vertical stripes, whereas uniform speed areas are characterized by slanting stripes. Complex motion areas contain various motion speeds, and the tracks of objects in these areas show structures in the form of a pseudonatural image. Because the static and uniform speed areas have only simple stripe-like motion structures in STS images, we call these two areas as simple motion areas. Figure 3 shows an example of a simple motion area and a complex motion area.

For simple motion areas, the human eye can predict the movement of objects. In STS images, these areas have simple structures: only horizontal, vertical, or slanted stripes. However, the complex motion areas contain more motion structure information. Typically, they are formed by cutting objects that are moving at different speeds in many directions. The number of edges and textures is greater in this region.

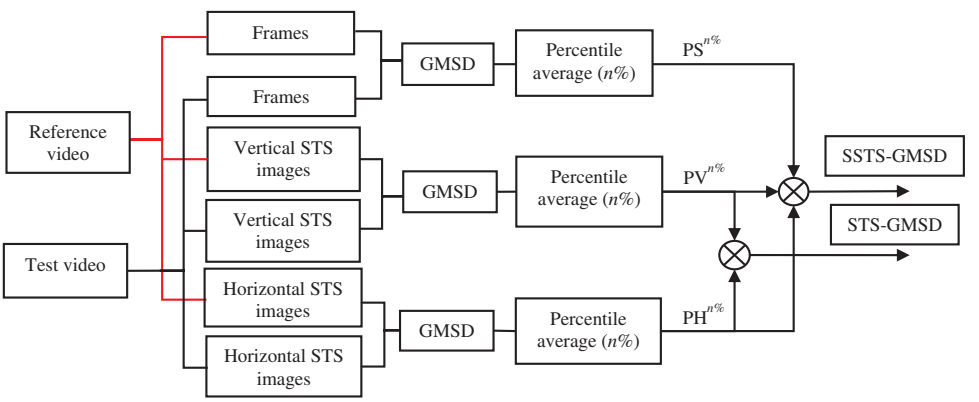

Fig. 2 Flowchart of the STS-GMSD and SSTS-GMSD algorithms.

Because the simple and complex motion areas are all structural, distortion in the two regions will destroy the motion structure. Then, the information that comes from the structure's changes in these motion areas can be used to measure the quality of the video. To quantize the motion structure distortion, an algorithm that can detect the structure's changes in these areas is needed.

Among the available structural similarity algorithms, SSIM is a milestone in quality assessment. However, when blur and geometric distortion exists, the performance of SSIM deteriorates. To overcome the defects of SSIM, many algorithms, such as ESSIM,[7] FSIM,[8] MS-SSIM,[9] gradient-based structural similarity,[23] and GMSD,[11] have been proposed. Among these algorithms, GMSD, which uses only the image gradient magnitude, is an excellent perceptual FR-IQA method. Specifically, GMSD uses the pixel-wise GMS between the reference and distorted images combined with a pooling strategy—the standard deviation of the GMS map—to accurately predict perceptual image quality.[11] GMSD improves the evaluation accuracy for blurred images and achieves very good quality evaluation relevance for a number of mainstream IQA databases.

On the basis of the above analysis, the GMSD algorithm was viewed as an alternative for us to extract structural distortion information from the simple and complex motion areas. However, before formally selecting it, we compared its performance with that of other mainstream FR-IQA methods for STS images. Thus, we discuss the experiment to select an algorithm for detecting motion structure partition similarity of STS images on the LIVE VQA database,[24,25] which has long been a VQA ground-truth database.

We first replaced the GMSD algorithm with mainstream FR-IQA algorithms and performed the objective VQA with the framework shown in Fig. 2. Then, the objective values were calculated separately for each frame and the STS image of the video matrix from three dimensions (video frames, vertical STS images, and horizontal STS images). After the objective scores of each dimension were obtained, we selected all the scores and averaged them to obtain a quality index. Then, we selected the worst 20% of the scores and averaged them to obtain another quality index. Finally, 16 evaluation performances of the mainstream FR-IQA algorithms in terms of the structural similarity of video STS images were obtained. The specific results are shown in Table 1. The results show that the GMSD algorithm is the best structural similarity-detecting algorithm for STS images among the tested mainstream FR-IQA algorithms on the LIVE VQA database. Therefore, the GMSD algorithm was used to compute the motion structure partition similarity of STS images.

In Table 1, the symbol "·" represents the multiplication operation; for example, $PS_{n\%} \cdot PH_{n\%} \cdot PV_{n\%}$ is the product of the $PS_{n\%}$, $PH_{n\%}$, and $PV_{n\%}$ indices. $PS_{n\%}$ is found by applying the FR-IQA algorithm frame by frame to the reference and test videos and then calculating the average quality of the worst $n\%$ of all the results as the quality index. $PH_{n\%}$ and $PV_{n\%}$ are found by applying the FR-IQA algorithms to the horizontal and vertical STS images, respectively, just as in the case of the frames. $V1$, $V2$, $V3$, and $V4$ are four video quality indices, which are computed by the formulas given under them in Table 1.

Table 1 Performance comparison of FR-IQA algorithms extended to video STS images in terms of Spearman rank order correlation coefficient (SROCC). The table is arranged in the descending order of the last column.

	$V1$	$V2$	$V3$	$V4$
Algorithm	$PS_{100\%} \cdot PV_{100\%} \cdot PH_{100\%}$	$PS_{20\%} \cdot PV_{20\%} \cdot PH_{20\%}$	$PV_{100\%} \cdot PH_{100\%}$	$PV_{20\%} \cdot PH_{20\%}$
GMSD[11]	0.816	0.838	0.822	0.831
NLOG-MSE[26]	0.760	0.795	0.782	0.791
NLOG-COR[26]	0.750	0.783	0.769	0.771
MAD[2]	0.789	0.764	0.788	0.765
PAMSE[27]	0.712	0.721	0.726	0.728
NQM[28]	0.668	0.688	0.656	0.679
SRSIM[29]	0.638	0.712	0.589	0.664
FSIM[8]	0.636	0.701	0.572	0.607
VIF[30]	0.612	0.644	0.625	0.603
GSM[31]	0.597	0.642	0.547	0.580
PSNR	0.524	0.574	0.526	0.567
MSSIM[9]	0.546	0.556	0.557	0.541
SSIM[6]	0.587	0.62	0.526	0.525
RFSIM[10]	0.536	0.585	0.468	0.500
IFC[32]	0.540	0.534	0.569	0.480
UQI[33]	0.465	0.415	0.461	0.366

The comparison results of the $V1$ and $V2$ indices as well as the $V3$ and $V4$ indices in Table 1 indicate that taking the average of the worst 20% of all results improves relevance with subjective evaluation. The comparison results of $V1$ and $V3$ as well as $V2$ and $V4$ in Table 1 show that the Spearman rank order correlation coefficient (SROCC) values fluctuate when the effects of spatial information and the STS images are jointly considered; that is, the evaluation performance in this case does not improve compared with that in the case where only the STS images are used. Thus, it is clear that the STS information by itself is sufficient for evaluation of the quality of videos.

According to the $V4$ quality index, the performance ranking of the tested FR-VQA algorithms is as follows: GMSD, normalized Laplacian of Gaussian mean square error (NLOG-MSE), NLOG-COR, MAD, perceptual fidelity aware mean squared error (PAMSE), noise quality measure (NQM), spectral residual based similarity (SRSIM), FSIM, VIF, GSM, PSNR, multiscale SSIM index (MSSIM), SSIM, RFSIM, information fidelity criterion (IFC), and universal quality index (UQI). This result shows that the GMSD

algorithm is the most suitable algorithm for detecting the motion structure partition similarities of STS images on the LIVE VQA database. The performance of the SSIM-related indicators is poor, especially that of UQI, which is the prototype of SSIM. FSIM is the only SSIM-related indicator with acceptable performance for the STS images.

3.2 *FR-VQA Algorithm Based on Motion Structure Partition Similarity of STS Images*

After choosing GMSD to detect the motion structure distortion in the STS images, we further studied the impact of the statistical characteristics of the two different regions (simple and complex motion structure areas) of the video STS images on VQA. Consequently, we propose an FR-VQA algorithm based on the motion structure partition similarity of STS images.

As shown in area B of Fig. 3, the simple motion area represents the image areas where the tracks of objects appear like stripes (when the object is static or along a particular direction of uniform movement) in the STS images. As shown in area A of Fig. 3, the complex motion area represents the image areas, where the tracks of objects appear like natural images (many different movement directions are present in this region).

The proposed algorithm is called the STS-based motion structure partition similarity (STS-MSPS) algorithm. Figure 4 shows the framework of STS-MSPS. In the framework, only the luminance component of videos is considered, and GMSD is used to detect the motion structure partition similarity of the STS images of the reference and distorted videos. First, all the STS images, including the vertical and horizontal STS images, are extracted. Next, each STS image is divided into two motion areas: simple areas and complex motion areas. Then, the motion structure partition similarities of the two regions are measured using the GMSD algorithm. The GMSD scores of the two motion areas of each slice are weighted to integrate one GMSD score. To observe the influence of these two regions on VQA, we adopted distinct weights in the experiment. Finally, we apply percentile processing, which has been proven to be an effective importance pooling method in previous studies,[34,35] to all the GMSD values of the video STS sequences to obtain the video quality index. The percentile average with a percentile pc% is denoted by $\mathrm{PA}^{\mathrm{pc}\%}(\cdot)$, and the video quality score is expressed as follows:

$$\text{H-MSPS} = \mathrm{PA}^{\mathrm{pc}\%}(\mathrm{GMSD}_H^{\mathrm{nh}}), \qquad 0 \le \mathrm{pc} \le 100, \tag{1}$$

$$\text{V-MSPS} = \mathrm{PA}^{\mathrm{pc}\%}(\mathrm{GMSD}_V^{\mathrm{nv}}), \qquad 0 \le \mathrm{pc} \le 100, \tag{2}$$

$$\text{STS-MSPS} = \text{H-MSPS} \times \text{V-MSPS}, \tag{3}$$

where $\mathrm{GMSD}_H^{\mathrm{nh}}$ and $\mathrm{GMSD}_V^{\mathrm{nv}}$ are the GMSD values of the horizontal and vertical STS sequences, respectively. nh and nv are the number of horizontal and vertical STS images, respectively.

The most important part of the proposed algorithm is the segmentation of the two motion areas. When the regional segmentation is accurate, we can further validate the impact of the two regions on VQA. In the following subsection, the method for dividing the motion structure areas of the STS images is described in detail.

3.3 *Division of STS Motion Structure Areas*

The simple and complex areas of STS images are both structures. We analyze their structural differences below:

1. Because the simple motion area is focused toward a particular direction, for example, the tracks of static objects are horizontal or vertical in STS images, we can determine the movement direction of an object from the different direction projections of STS images. The projection variance will be large only along the stripe direction. In other directions, the projection variance will be small. Because the complex motion areas have no specific orientation, the variances of all

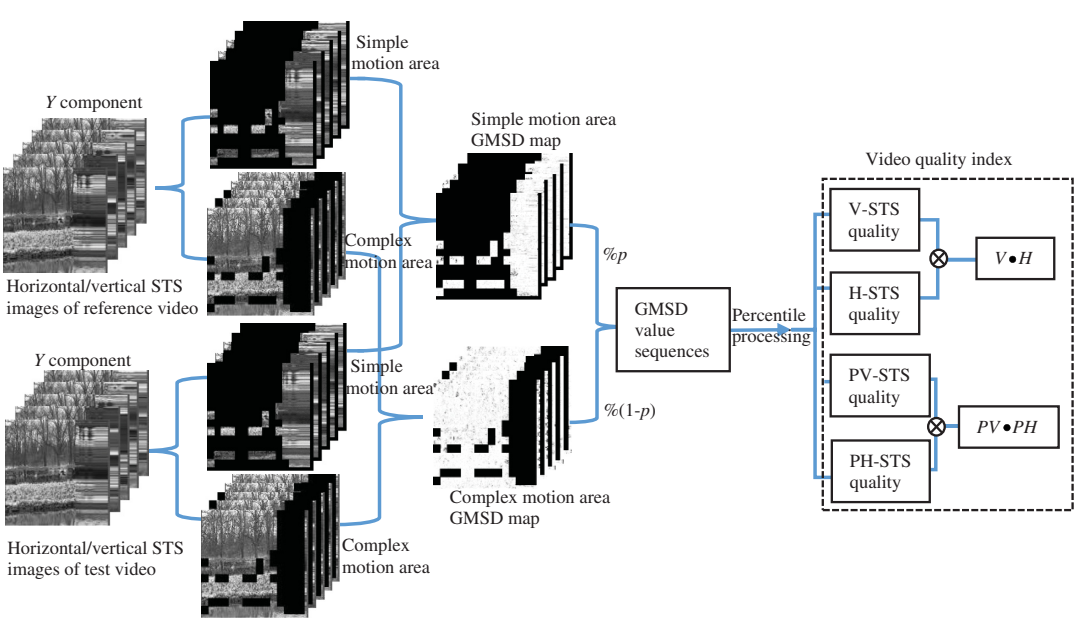

Fig. 4 Framework of the proposed method (STS-MSPS).

the projection values in different directions are virtually the same. Therefore, the variances of different projections can be used to distinguish between simple and complex motion areas.

2. Because the gradient orientation of the STS image is perpendicular to the stripe direction, the gradient orientation can also reflect the local motion feature of the STS image.

Considering the above two points, we directly use the projections of the gradient orientation map of the STS image to distinguish the motion structure areas. The value at any point in the gradient orientation map will always be in the range $[-3.14, 3.14]$, and the value at every point in the gradient orientation map directly represents the direction of the gradient at that point. The projection transformation of the gradient orientation map can directly show the fluctuation of that orientation. Typically, the complex motion area is focused in all directions; therefore, the projections of the gradient orientation map tend to be similar in all directions. However, the simple motion area has a strong direction; therefore, only the projection fluctuations along the gradient orientation are large. This rule can be used to not only classify the simple and complex motion areas of STS images but also design algorithms to distinguish normal pictures with a specific orientation from those with no specific orientation. The following equations can be used to calculate the gradient orientation map of vertical STS images:

$$dx = \frac{\partial I(x,t)}{\partial x}, \tag{4}$$

$$dt = \frac{\partial I(x,t)}{\partial t}, \tag{5}$$

$$\mathrm{mag}(x,t) = \sqrt{dx^2 + dy^2}, \tag{6}$$

$$\mathrm{ang}(x,t) = a\tan\left(\frac{dt}{dx}\right), \tag{7}$$

where $I(x,t)$ is the vertical STS image, dt, dx are the first-order derivation of $I(x,t)$, $\mathrm{mag}(x,t)$ is the gradient map of $I(x,t)$, and $\mathrm{ang}(x,t)$ is the gradient orientation map of $I(x,t)$.

The specific regional segmentation methods are as follows:

1. The gradient orientation maps of the STS images are calculated using Eqs. (4)–(6). The gradient orientation map (Fig. 5) is generated by calculating the angle of the gradient map.
2. The gradient orientation maps of the STS images are divided into blocks of size $n \times n$. If n is too small, the blocks will contain too few stripes and the segmentation accuracy will be low. Conversely, if n is too large the blocks will contain too many stripes in all directions. This would result in the overall fluctuation of variances calculated from different projections becoming small, making the fluctuation unusable for distinguishing the simple motion area. Figures 6(a) and 6(b), respectively, show the normal and zoomed displays of the block, whose size is 32×32, marked by the small white square in area B of Fig. 3. Figures 6(c) and 6(d), respectively, show the normal and zoomed displays of the gradient orientation map of the image block shown in Fig. 6(a).
3. The m variance values of the Radon transform of the gradient orientation map along $\theta_1, \theta_2, \ldots, \theta_m$ degrees are computed according to Eq. (9). Figure 7 shows the fluctuation of the projection image shown in Fig. 6(a) in four directions (0 deg, 45 deg, 90 deg, and 135 deg). The Radon integrals along the 90-deg direction are the most volatile, whereas the fluctuations for the other three degrees are small

$$\theta_i = \frac{180}{m}(i-1), \qquad i = 1,2,\ldots,m \tag{8}$$

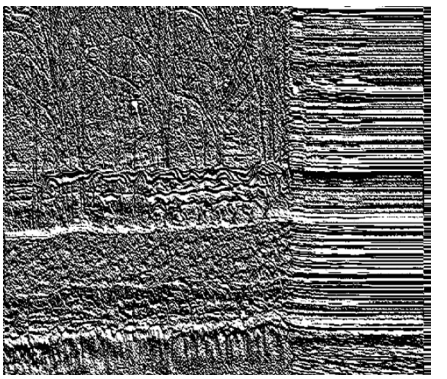

Fig. 5 Gradient orientation map of the image shown in Fig. 3.

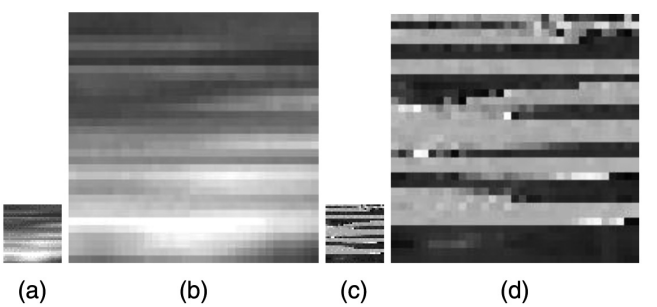

Fig. 6 (a) Normal display of a simple motion block, (b) zoomed display of the simple motion block shown in (a), (c) display of gradient orientation of the simple motion block, and (d) zoomed display of the block shown in (c).

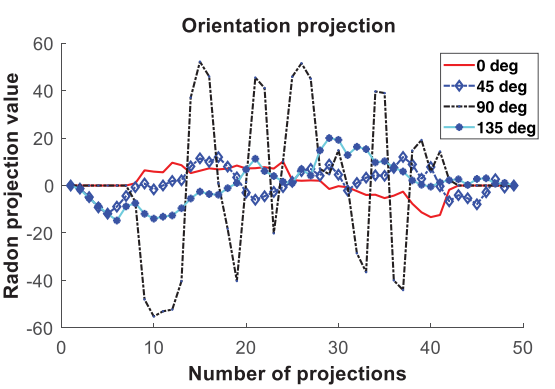

Fig. 7 Projection of the block shown in Fig. 6(a).

where m is the projection number and θ_i is the i'th projection angle.

4. We obtain the variances of the projection values in all directions and select the maximum variance stdmax according to Eq. (10). The mean value of the remaining $m-1$ variances is calculated using Eq. (11), and then the variance ratio stdratio is calculated using Eq. (12). In Fig. 7, stdratio is 4.3981.

$$\mathrm{std}\,\theta_i = \mathrm{std}\{\mathrm{Radon}[\mathrm{ang}(x,t)_{n\times n}, \theta_i]\}, \quad i = 1,2,\ldots,m, \tag{9}$$

$$\mathrm{stdmax} = \max(\mathrm{std}\theta_1, \mathrm{std}\theta_2, \ldots, \mathrm{std}\theta_m), \tag{10}$$

$$\mathrm{stdmean} = \mathrm{mean}[\min_{m-1}(\mathrm{std}\theta_1, \mathrm{std}\theta_2, \ldots, \mathrm{std}\theta_m)], \tag{11}$$

$$\mathrm{stdratio} = \frac{\mathrm{stdmax}}{\mathrm{stdmean}}, \tag{12}$$

$$\mathrm{mask}_{n\times n}(x,t) = \begin{cases} 1_{n\times n}, & \mathrm{stdratio} > T \\ 0_{n\times n}, & \mathrm{stdratio} < T \end{cases}, \tag{13}$$

where $\mathrm{std}\theta_1, \mathrm{std}\theta_2, \ldots, \mathrm{std}\theta_m$ represent the corresponding projection variances in the projection directions $\theta_1, \theta_2, \ldots, \theta_m$ of the gradient orientation images, $\mathrm{Radon}[X,\theta]$ is the Radon transform of matrix X along θ degrees, $\max(\cdot)$ is the max function, and $\min_{m-1}(\cdot)$ is the function used to obtain the $m-1$ smallest values out of the total m values. mean$[\cdot]$ computes the mean value. $\mathrm{mask}_{n\times n}(x,t)$ is the block mask; if its value is $1_{n\times n}$, the block is in the simple motion area, and if its value is $0_{n\times n}$, the block is in the complex motion area. T is the threshold value. From experience, we know that more accurate results can be obtained when T is set to two.

The block in the simple motion area is to be compared with the one in the complex motion area. Accordingly, Figs. 8 and 9, respectively, show a complex motion block, which is marked by the small white square in area A of Fig. 3, and the projection of the block. Figure 9 shows that the projection fluctuations in all four directions are virtually the same. For this block, stdratio is 1.7073.

5. The simple motion areas $\mathrm{mask}(x,t)$ are generated using every block mask computed by Eq. (13). Then, the complex motion template is $1-\mathrm{mask}(x,t)$. Subsequently, using the two templates, we can divide the GMSmap of each STS image into complex motion GMSmap and simple motion GMSmap. The variances of the two regional GMS maps are considered as separate GMSD values, as expressed in Eqs. (17) and (18).

$$\mathrm{GMSmap}_V = \frac{2\mathrm{mag}_r(x,t)\mathrm{mag}_d(x,t)+c}{\mathrm{mag}_r^2(x,t)+\mathrm{mag}_d^2(x,t)+c}, \tag{14}$$

where the subscript V denotes that the GMS map is computed from vertical STS images, subscript r indicates the reference video, and subscript d represents the test video. c is a constant value

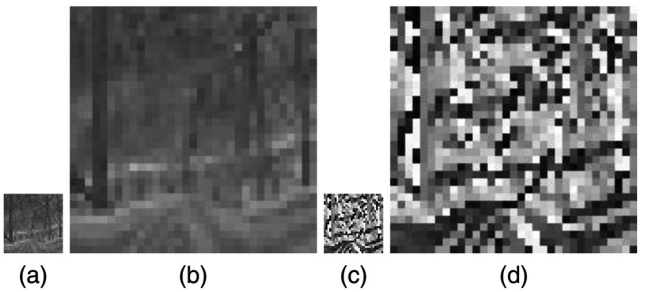

Fig. 8 (a) Block of complex motion area, (b) amplification of the block shown in (a), (c) gradient orientation of the block shown in (a), and (d) zoomed map of the block shown in (c).

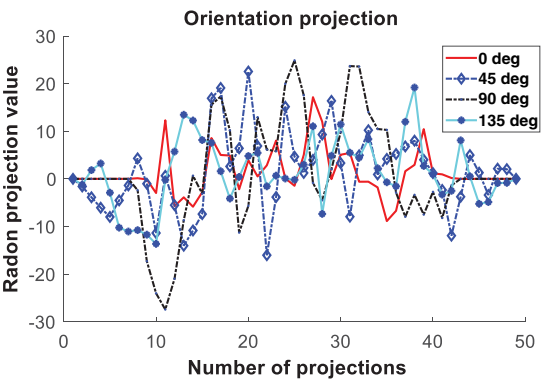

Fig. 9 Projection of the block shown in Fig. 8(a).

$$\mathrm{GMSmap}_V^{\mathrm{simple}} = \mathrm{GMSmap}_V \cdot \mathrm{mask}(x,t), \tag{15}$$

$$\mathrm{GMSmap}_V^{\mathrm{complex}} = \mathrm{GMSmap} \cdot [1-\mathrm{mask}(x,t)], \tag{16}$$

$$\mathrm{GMSD}_V^{\mathrm{simple}} = \mathrm{std}(\mathrm{GMSDmap}_V^{\mathrm{simple}}), \tag{17}$$

$$\mathrm{GMSD}_V^{\mathrm{complex}} = \mathrm{std}(\mathrm{GMSDmap}_V^{\mathrm{complex}}). \tag{18}$$

As an example, Figs. 10(a) and 10(b) show the STS images of reference and test videos, respectively. Figure 10(c) represents the gradient orientation of the image shown in Fig. 10(a).

The image shown in Fig. 10(a) is segmented into blocks; for each block, we obtain the projection along four orientations (0 deg, 45 deg, 90 deg, and 135 deg). The variances of the projection values are used to calculate the variance fluctuation ratio. If the ratio is greater than the threshold T, the small block is considered to have a strong direction and is classified as a simple motion area. If it is less than T, the block is divided into complex motion regions. Subsequently, we generate the appropriate regional template. Figure 11(a) shows the simple motion area template, and Fig. 11(b) shows the complex motion area template. Then, we compute the GMS map [Fig. 12(a)] of the original STS image [Fig. 10(a)] and the distortion STS image

Fig. 10 (a) Vertical STS image of reference video, (b) vertical STS image of test video, and (c) gradient orientation of image shown in (a).

[Fig. 10(b)] using Eq. (14). On obtaining the regional template and GMS map, the regional GMS maps [Figs. 12(b) and 12(c)] are obtained by multiplying the corresponding template and GMS map. Thus, we calculate the standard deviation between the two partial GMSD maps to obtain the regional GMSD values.

6. The two regional GMSD values of each STS image are integrated using different weights according to Eq. (19) to study the different impacts of the two regions on VQA. We set different values for p and calculate the corresponding results

$$\mathrm{GMSD}_V = \mathrm{GMSD}_V^{\mathrm{simple}} \times p\% + \mathrm{GMSD}_V^{\mathrm{complex}} \times (1-p)\%, \tag{19}$$

where $p = 0{,}20{,}50{,}80{,}100$.

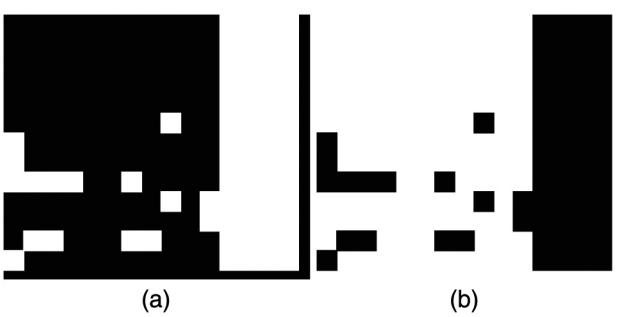

Fig. 11 (a) Generated simple motion area template and (b) generated complex motion area template of the image shown in Fig. 10(a).

The same motion area segmentation methods as those for $I(x,t)$ are used to obtain the integrated GMSD values for the horizontal STS images $I(y,t)$.

4 Experimental Evaluation and Results

4.1 *Video Quality Performance Criteria and Evaluation Procedures*

In addition to the LIVE VQA database discussed above, STS-MSPS was also evaluated on the computational and subjective image quality lab (CSIQ)[15,36] and the image and video processing laboratory (IVP)[37] VQA databases. A general summary of the CSIQ, IVP, and LIVE VQA databases is given in Table 2. All the videos in these databases are in the YUV420 format. However, only the Y component was used in our experiment.

To evaluate the performance of STS-MSPS, the SROCC, Pearson linear correlation coefficient (PLCC), and root mean squared error (RMSE) of the objective and subjective values were calculated in the experiment. Following the recommendations[39] proposed by the Video Quality Experts Group, a four-parameter logistic transform was applied to the raw predicted scores before computing the PLCC and RMSE values. The transformation formula is

$$Q_j' = \frac{(\beta_1 - \beta_2)}{1 + \exp\left(\frac{Q_j - \beta_3}{|\beta_4|}\right)} + \beta_2, \tag{20}$$

where Q_j' is the predicted subjective score for a test video j and Q_j denotes the original objective score obtained using a VQA algorithm. β_1, β_2, β_3, and β_4 are free parameters selected to provide the best fit of the predicted subjective score to the subjective rating scores. The initial values of

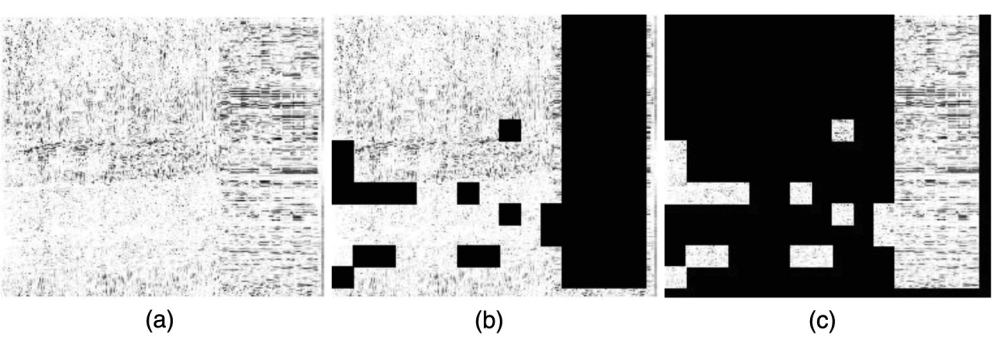

Fig. 12 GMS maps of (a) images shown in Figs. 10(a) and 10(b), (b) complex motion area, and (c) simple motion area.

Table 2 General summary of the evaluated VQA databases.

Database	Number of distorted videos	Video resolution	Frame rates (fps)	Number of distortion types	Number of subjects	Subjective methods	Subjective score type
CSIQ	216	832×480	24, 25, 30, 60	6	35	SAMVIQ[36]	DMOS
IVP	128	1920×1088	25	4	42	ACR-HR[37]	DMOS
LIVE	150	768×462	25, 50	4	38	Single stimulus[38]	DMOS

the four β free parameters were chosen based on the recommendation in Ref. 39. In addition to the three criteria outlined above, two statistical significance tests were performed on the VQA databases to evaluate the performance of STS-MSPS in Sec. 4.3.

In the experiment, the vertical and horizontal STS images of each video were first extracted. Then, each STS image was divided into two areas (complex and simple motion areas) using the methods proposed in Sec. 3.3. Subsequently, the GMSD values of the two image areas of a video sequence were computed, and different weights were assigned to the two zones to obtain one GMSD value. In the experiment, the weights of the simple motion area were 100%, 80%, 50%, 20%, and 0%, whereas those of the complex motion area were 0%, 20%, 50%, 80%, and 100%, respectively. Finally, the degree of correlation between the objective and subjective scores was calculated. The number of projection orientations was four, and the block size n for these orientations was set to 8, 16, 32, and 64. In our experiment, the default value used in a previous study[11] was adopted for the constant c in Eq. (14).

4.2 *Performance of the Proposed VQA Algorithms Based on Motion Structure Partition Similarity of STS Images*

To determine the parameters of STS-MSPS, we conducted a more intensive experiment on the LIVE VQA database and then validated the effectiveness of the STS-MSPS with fixed parameters on other VQA databases. Table 3 summarizes the evaluation results with different block sizes.

The H-STS and V-STS indices listed in Table 3 were obtained by first applying the proposed algorithm to every STS image (horizontal or vertical) of the reference and test videos and then calculating the average results of all the scores to obtain the video quality index. The PH-STS and PV-STS indices represent the percentile average processing of the H-STS and V-STS indices. For example, the PH-STS index indicates that we took the worst 20% of all the results by applying the proposed algorithm to every horizontal STS image and then averaged them to obtain the video quality index. The $V \cdot H$ index is the product of the V-STS and H-STS indices. Similarly, the $PV \cdot PH$ index is the product of the PH-STS and PV-STS indices.

As can be seen from Table 3, the SROCC scores increase as the weights of the complex motion areas increase. When the proportion of the complex motion areas is 80% or 100%, the relevance is virtually the same. For example, in terms of the three columns V-STS, PV-STS, and $PV \cdot PH$, the SROCC score in the complex motion region with a 100% proportion is slightly higher than that with an 80% proportion. In terms of the three columns, H-STS, PH-STS, and $V \cdot H$, the SROCC score in the complex motion region with a 100% proportion is slightly lower than that with an 80% proportion. Overall, the proportion of the complex motion area should be between 80% and 100%.

By separately comparing the V-STS index with the PV-STS index, the H-STS index with the PH-STS index, and the $V \cdot H$ index with the $PV \cdot PH$ index (Table 3), we can conclude that percentile average processing can provide better relevance in most cases. We found that the relevance is the best when the block is 32×32. This may be because the algorithm can segment the two kinds of regions more accurately with this block size. Therefore, when testing the STS-MSPS on the CSIQ and IVP VQA databases, we used a block size of 32×32. The specific results obtained are shown in Table 4.

The results for the STS-based VQA[15,22] algorithms indicate that video quality evaluation can be achieved using the information from STS images. In this study, an algorithm for partitioning video STS images was designed and the impact of different areas of the STS images on VQA investigated. The experimental results show that the different areas of the video STS images did not have equal status in the evaluation. In practice, the complex motion area should take a large proportion, nearly 80% to 100%, and in this manner, a better correlation with subjective evaluation can be achieved. In the following performance comparison, we set the complex motion proportion as 100% and block size as 32 because these values typically provide good performance on all the three databases. Table 5 presents the performances of some mainstream FR-VQA algorithms in terms of SROCC, PLCC, and RMSE. STS-MSPS is better than any other FR-VQA algorithm listed in Table 5 on the LIVE VQA database, ranks second on the CSIQ VQA database, and third on the IVP VQA database. The scatter plots marked with the coefficient of determination R-square of STS-MSPS on different VQA databases are shown in Fig. 13.

The results of this study show that the complex motion areas have a considerably greater impact than the simple motion areas on VQA. Because all the complex motion areas in the video can construct motion tubes, in the future, we intend to extract visual distortion from these motion tubes as that may help to improve the prediction accuracy of VQA.

4.3 *Statistical Evaluation*

In this section, we outline the F-test performed on the ratio of the variance of the residual error from one objective model to that of another objective model to assess the statistical significance differences[15,25,40] in the performance of STS-MSPS and other algorithms. The algorithms PSNR, SSIM, VQM, ST-MAD, ViS3, and STS-MSPS were compared. SSIM[41] was performed on a frame-by-frame basis using default

Table 3 SROCC of the proposed methods with different block sizes on the LIVE VQA database.

Block size	Simple motion area (%)	Complex motion area (%)	*V*-STS	*H*-STS	*PV*-STS	*PH*-STS	$V \cdot H$	$PV \cdot PH$
8×8	0	100	0.852	0.777	0.843	0.800	0.832	0.839
	20	80	0.840	0.781	0.837	0.805	0.826	0.838
	50	50	0.822	0.780	0.823	0.794	0.811	0.820
	80	20	0.797	0.769	0.799	0.783	0.792	0.800
	100	0	0.772	0.758	0.783	0.765	0.775	0.784
16×16	0	100	0.856	0.778	0.846	0.804	0.835	0.844
	20	80	0.846	0.785	0.839	0.804	0.835	0.838
	50	50	0.826	0.784	0.821	0.782	0.816	0.813
	80	20	0.792	0.769	0.793	0.755	0.793	0.783
	100	0	0.770	0.753	0.772	0.732	0.772	0.764
32×32	0	100	0.858	0.782	0.852	0.804	0.837	0.847
	20	80	0.854	0.805	0.847	0.806	0.847	0.843
	50	50	0.804	0.784	0.826	0.782	0.803	0.817
	80	20	0.722	0.722	0.795	0.744	0.731	0.785
	100	0	0.675	0.660	0.776	0.714	0.672	0.764
64×64	0	100	0.848	0.775	0.845	0.796	0.826	0.838
	20	80	0.833	0.797	0.83	0.802	0.832	0.832
	50	50	0.747	0.737	0.776	0.772	0.755	0.788
	80	20	0.621	0.601	0.716	0.725	0.620	0.724
	100	0	0.544	0.516	0.691	0.688	0.531	0.684

parameters. Before conducting the *F*-test, a Jarque–Bera[42] test was performed to determine whether the residuals of competing algorithms have Gaussian distributions on the three above-mentioned VQA databases with DMOS. All algorithms passed the Jarque–Bera test on the LIVE VQA database. STS-MSPS did not pass the Jarque–Bera test on the CSIQ and IVP databases, but the variance of residuals of STS-MSPS was the smallest on the CSIQ database, and it ranked second overall on the CSIQ database. Table 6 shows the variance of the residuals between the VQA algorithms' predictions and DMOS values. Table 7 shows the statistical evaluation results obtained on the LIVE VQA database.

In Table 7, a symbol value of "1" indicates that the statistical performance of the VQA model in the row is superior to that of the model in the column, a symbol value of "0" indicates that the statistical performance of the model in the row is inferior to that of the model in the column, and "–" indicates that the statistical performance of the model in the row is equivalent to that of the model in the column. Table 7 shows that the variance of residuals yielded by STS-MSPS is not significantly different from those of ST-MAD and ViS3, whereas it is significantly better than PSNR, SSIM, and VQM on the LIVE database.

To eliminate the normality requirement for residuals of *F*-test, we further conducted a Wilcoxon rank-sum test,[43] which is based solely on the rank order of the observed samples, to test the significance of the compared methods. Wilcoxon rank-sum tests the equality of the median values of two distributions. In an NR-IQA,[44] Wilcoxon rank-sum test is used to compute the significance median of Spearman coefficient. For FR-VQA, it can also be used. Note that the results of the Wilcoxon rank-sum test were obtained using the medians of the absolute residual of the compared FR-VQA's predictions. The smaller the median of absolute residual is, the better the algorithm's performance will be. Table 8 shows the results of the compared FR-VQA algorithms with a significance level of 0.05. Similar conclusions to the *F*-test were obtained on the LIVE VQA database. In addition, STS-MSPS has

Table 4 SROCC of the proposed algorithms on the CSIQ and IVP databases. The data marked in bold are the proposed algorithm's performance. The following parameters values were used with the algorithms: block size $= 32 \times 32$; threshold $= 2$.

Database	Simple motion area (%)	Complex motion area (%)	*V*-STS	*H*-STS	*PV*-STS	*PH*-STS	$V \cdot H$	$PV \cdot PH$
CSIQ	0	100	0.804	0.798	0.783	0.844	0.805	**0.821**
	20	80	0.804	0.788	0.788	0.839	0.797	0.821
	50	50	0.787	0.760	0.794	0.829	0.777	0.816
	80	20	0.762	0.730	0.787	0.812	0.751	0.808
	100	0	0.742	0.706	0.779	0.797	0.729	0.795
IVP	0	100	0.868	0.816	0.873	0.845	0.853	**0.876**
	20	80	0.841	0.795	0.853	0.836	0.828	0.856
	50	50	0.789	0.770	0.819	0.818	0.779	0.823
	80	20	0.739	0.729	0.786	0.799	0.742	0.801
	100	0	0.704	0.699	0.767	0.788	0.708	0.786

Table 5 Comparison of the performance of VQA metrics on the three VQA databases.

Performance criteria	Database name	PSNR	VQM	MOVIE	TQV	ST-MAD	ViS3	STS-MSPS
SROCC	LIVE	0.523	0.756	0.789	0.802	*0.825*	0.816	**0.847**
	IVP	0.728	0.845	*0.880*	0.701	0.721	**0.896**	0.876
	CSIQ	0.579	0.789	0.806	0.814	0.736	**0.841**	*0.821*
PLCC	LIVE	0.549	0.770	0.812	0.815	*0.830*	0.829	**0.856**
	IVP	0.723	0.847	*0.879*	0.722	0.725	**0.896**	0.867
	CSIQ	0.565	0.769	0.788	0.795	0.724	**0.830**	*0.801*
RMSE	LIVE	9.175	7.010	6.425	6.357	*6.118*	6.146	**5.684**
	IVP	0.730	0.561	*0.504*	0.731	0.728	**0.470**	0.528
	CSIQ	13.724	10.63	10.231	10.09	11.475	**9.273**	*9.955*

Note: Best is in bold and second is in italic.

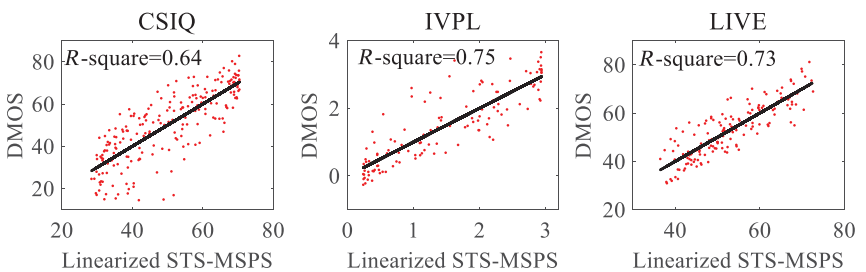

Fig. 13 Scatter plots of the linearized STS-MSPS on the three VQA databases.

Table 6 Variance of the residuals between VQA algorithms' predictions and DMOS values.

Database	PSNR	SSIM	VQM	ST-MAD	ViS3	STS-MSPS
LIVE	106.89	90.56	62.81	37.68	37.02	32.52
CSIQ	202.68	277.78	118.27	132.30	90.00	99.56
IVP	216.16	312.88	85.56	148.47	93.54	77.95

Table 7 Results of the *F*-test performed on the residuals between model predictions and DMOS values on the LIVE VQA database.

	PSNR	SSIM	VQM	ST-MAD	ViS3	STS-MSPS
PSNR	—	—	0	0	0	0
SSIM	—	—	0	0	0	0
VQM	1	1	—	0	0	0
ST-MAD	1	1	1	—	—	—
ViS3	1	1	1	—	—	—
STS-MSPS	1	1	1	—	—	—

equal median values to ViS3, ST-MAD, and VQM on the CSIQ and IVP databases.

4.4 *Runtime Comparison*

Furthermore, we examined the runtime of the STS-based VQA algorithms. All algorithms were executed on a computer with an Intel(R) Core(TM) i7-7700K 4.2 GHz CPU, 16 GB RAM, Microsoft Windows 10 64-bit, and MATLAB 2017b. We selected the BQMALL reference video and its first distorted video from the CSIQ database to process. The test videos contained 600 frames of size 832×480. The MATLAB source codes of all the algorithms were obtained from the respective authors. The runtime results are shown in Table 9.

Table 9 shows that our algorithm has an advantage in terms of runtime (i.e., faster) over STS-MAD and VIS3.

5 Conclusions

In this paper, we proposed an FR-VQA algorithm based on the motion structure partition similarity of STS images that effectively integrates the spatial and temporal information of videos. First, the STS images are divided into simple and complex motion areas according to the statistical characteristics of motion structures. Then, the distortions are measured using the GMS of the partitions of STS images and assigned different weights. Our experimental results indicate that when the proportion of the complex motion areas is >80%, the evaluation correlation is excellent, which shows that the HVS may be more sensitive to distortion in these complex motion areas. The performance of the proposed algorithm was found to be best on the LIVE VQA database, and on par with that of other state-of-the-art FR-VQA algorithms on the CSIQ and IVP benchmark databases. In addition, the proposed algorithm is faster than the other STS-based algorithms evaluated.

Table 8 Results of rank-sum test on the three VQA databases.

		PSNR	SSIM	VQM	ST-MAD	ViS3	STS-MSPS
LIVE	PSNR	—	—	0	0	0	0
	SSIM	—	—	0	0	0	0
	VQM	1	1	—	0	0	0
	ST-MAD	1	1	1	—	—	—
	ViS3	1	1	1	—	—	—
	STS-MSPS	1	1	1	—	—	—
CSIQ	PSNR	—	—	0	0	0	0
	SSIM	—	—	0	0	0	0
	VQM	1	1	—	—	0	—
	ST-MAD	1	1	—	—	0	0
	ViS3	1	1	1	1	—	—
	STS-MSPS	1	1	—	1	—	—
IVP	PSNR	—	—	0	0	0	0
	SSIM	—	—	0	0	0	0
	VQM	1	1	—	1	—	—
	ST-MAD	1	1	0	—	0	0
	ViS3	1	1	—	1	—	—
	STS-MSPS	1	1	—	1	—	—

Table 9 Runtime comparison of the STS-based VQA algorithms.

	Runtime (s)	Ratio to STS-MSPS
ST-MAD	1110.02	4.10
ViS3	373.22	1.38
STS-MSPS	270.60	1

Acknowledgments

This work was supported in part by the National Key Research and Development Program of China (No. 2016YFA0202003) and the National Natural Science Foundation of China (No. 61571359).

References

1. D. M. Chandler and S. S. Hemami, "VSNR: a wavelet-based visual signal-to-noise ratio for natural images," *IEEE Trans. Image Process.* **16**(9), 2284–2298 (2007).
2. D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *J. Electron. Imaging* **19**(1), 011006 (2010).
3. H. R. Sheikh and A. C. Bovik, "A visual information fidelity approach to video quality assessment," in *First Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics*, pp. 23–25 (2005).
4. M. H. Pinson and S. Wolf, "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcast.* **50**(3), 312–322 (2004).
5. K. Seshadrinathan and A. C. Bovik, "Motion tuned spatio-temporal quality assessment of natural videos," *IEEE Trans. Image Process.* **19**(2), 335–350 (2010).
6. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
7. G.-H. Chen et al., "Edge-based structural similarity for image quality assessment," in *IEEE Int. Conf. on Acoustics Speech and Signal Processing Proc.*, Vol. 2, pp. II-933–II-936, IEEE (2006).
8. L. Zhang et al., "FSIM: a feature similarity index for image quality assessment," *IEEE Trans. Image Process.* **20**(8), 2378–2386 (2011).
9. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Thrity-Seventh Asilomar Conf. on Signals, Systems and Computers*, Vol. 2, pp. 1398–1402 (2003).
10. R. Transforms et al., "RFSIM: a feature based image quality assessment metric using Riesz transforms," in *IEEE Int. Conf. on Image Processing*, pp. 321–324 (2010).
11. W. Xue et al., "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," *IEEE Trans. Image Process.* **23**(2), 684–695 (2014).
12. C. Li and A. C. Bovik, "Content-weighted video quality assessment using a three-component image model," *J. Electron. Imaging* **19**(1), 011003 (2010).
13. Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Process. Image Commun.* **19**(2), 121–132 (2004).
14. P. V. Vu, C. T. Vu, and D. M. Chandler, "A spatiotemporal most-apparent-distortion model for video quality assessment," in *18th IEEE Int. Conf. on Image Processing (ICIP)*, pp. 2505–2508 (2011).
15. P. V. Vu and D. M. Chandler, "ViS3: an algorithm for video quality assessment via analysis of spatial and spatiotemporal slices," *J. Electron. Imaging* **23**(1), 013016 (2014).
16. C. W. Ngo, T. C. Pong, and R. T. Chin, "Video partitioning by temporal slice coherency," *IEEE Trans. Circuits Syst. Video Technol.* **11**(8), 941–953 (2001).
17. C. W. Ngo, T. C. Pong, and H. J. Zhang, "Motion analysis and segmentation through spatio-temporal slices processing," *IEEE Trans. Image Process.* **12**(3), 341–355 (2003).
18. C. W. Ngo, T. C. Pong, and H. J. Zhang, "On clustering and retrieval of video shots through temporal slices analysis," *IEEE Trans. Multimedia* **4**(4), 446–458 (2002).
19. K. T. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Process.* **70**(3), 279–294 (1998).
20. A. K. Moorthy and A. C. Bovik, "Efficient motion weighted spatio-temporal video SSIM index," *Proc. SPIE* **7527**, 75271I (2010).
21. S. Wolf and M. H. Pinson, "Spatial-temporal distortion metric for in-service quality monitoring of any digital video system," *Proc. SPIE* **3845**, 266–277 (1999).
22. P. Yan, X. Mou, and W. Xue, "Video quality assessment via gradient magnitude similarity deviation of spatial and spatiotemporal slices," *Proc. SPIE* **9411**, 94110M (2015).
23. G. Chen, C. Yang, and S. Xie, "Gradient-based structural similarity for image quality assessment," in *Int. Conf. on Image Processing*, pp. 2929–2932 (2006).
24. K. Seshadrinathan et al., "A subjective study to evaluate video quality assessment algorithms," *Proc. SPIE* **7527**, 75270H (2010).
25. K. Seshadrinathan et al., "Study of subjective and objective quality assessment of video," *IEEE Trans. Image Process.* **19**(6), 1427–1441 (2010).
26. W. Xue and X. Mou, "Image quality assessment with mean squared error in a log based perceptual response domain," in *IEEE China Summit and Int. Conf. on Signal and Information Processing (ChinaSIP)*, pp. 315–319 (2014).
27. W. Xue et al., "Perceptual fidelity aware mean squared error," in *IEEE Int. Conf. on Computer Vision (ICCV)*, pp. 705–712, IEEE (2013).
28. N. Damera-Venkata et al., "Image quality assessment based on a degradation model," *IEEE Trans. Image Process.* **9**(4), 636–650 (2000).
29. L. Zhang and H. Li, "SR-SIM: a fast and high performance IQA index based on spectral residual," in *19th IEEE Int. Conf. on Image Processing (ICIP)*, pp. 1473–1476, IEEE (2012).
30. H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.* **15**(2), 430–444 (2006).
31. A. Liu, W. Lin, and M. Narwaria, "Image quality assessment based on gradient similarity," *IEEE Trans. Image Process.* **21**(4), 1500–1512 (2012).
32. H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.* **14**(12), 2117–2128 (2005).
33. Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Process. Lett.* **9**(3), 81–84 (2002).
34. A. K. Moorthy and A. C. Bovik, "Perceptually significant spatial pooling techniques for image quality assessment," *Proc. SPIE* **7240**, 724012 (2009).
35. K. Lee et al., "Temporal pooling of video quality estimates using perceptual motion models," in *Proc. of Int. Conf. on Image Processing (ICIP)*, pp. 2493–2496 (2010).
36. Laboratory of Computational Perception and Image Quality Oklahoma State University, "CSIQ video database," 2013, http://vision.eng.shizuoka.ac.jp/mod/page/view.php?id=24.
37. Image and Video Processing Lab, The Chinese University of Hong Kong, "IVP subjective quality video database," 2011, http://ivp.ee.cuhk.edu.hk/research/database/subjective/index.shtml.
38. International Telecommunication Union, "Methodology for the subjective assessment of the quality of television pictures," Technical Report (2002).
39. Video Quality Expert Group, "Final report from the video quality experts group on the validation of objective models of video quality assessment," Group 28, 2008, https://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-i/frtv-phase-i.aspx (February 2005).
40. H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.* **15**(11), 3440–3451 (2006).
41. Z. Wang, "SSIM image quality index," https://ece.uwaterloo.ca/~z70wang/research/ssim/ (February 2003).
42. C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Econ. Lett.* **6**(3), 255–259 (1980).
43. K. Kafadar and D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Press, Boca Raton, Florida (1997).
44. W. Xue et al., "Blind image quality assessment using joint statistics of gradient magnitude and laplacian features," *IEEE Trans. Image Process.* **23**(11), 4850–4862 (2014).

Peng Yan received his BS degree in electronic information science and technology from Taishan University, Taian, China, in 2010 and his MS degree in signal and information processing from Chang'an University, Xi'an, Shaanxi, China, in 2013. Currently, he is pursuing the PhD in information and communication engineering at Xi'an Jiaotong University. His research interests include the image quality assessment and video quality assessment.

Xuanqin Mou has been with the Institute of Image Processing and Pattern Recognition (IPPR), Electronic and Information Engineering School, Xi'an Jiaotong University, since 1987. He has been an associate professor since 1997, and a professor since 2002. Currently, he is the director of IPPR. Dr. Mou has authored or coauthored more than 200 peer-reviewed journal or conference papers. He has been granted as the Technology Academy Award for invention by the Ministry of Education of China, and the Technology Academy Awards from the Government of Shaanxi Province, China.